Tese apresentada à Pró-Reitoria de Pós-Graduação do Instituto Tecnológico de Aeronáutica, como parte dos requisitos para obtenção do título de Doutor em Ciências no Programa de Pós-Graduação em Física, Área de Física Nuclear.

Bruno da Silva Gonçalves

REDES NEURAIS BAYESIANAS NA INFERÊNCIA DE EQUAÇÕES DE ESTADO POLITRÓPICAS PARA ESTRELAS DE NÊUTRONS

Tese aprovada em sua versão final pelos abaixo assinados:

Prof. Dr. César Henrique Lenzi

Orientador

Profª. Drª. Mariana Dutra da Rosa Lourenço

Coorientadora

Prof. Dr. Sérgio José Barbosa Duarte

Coorientador

Campo Montenegro
São José dos Campos, SP - Brasil
2025

Dados Internacionais de Catalogação-na-Publicação (CIP)
Divisão de Informação e Documentação

Gonçalves, Bruno da Silva
Redes neurais bayesianas na inferência de equações de estado politrópicas para estrelas de nêutrons / Bruno da Silva Gonçalves.
São José dos Campos, 2025.
110f.

Tese de Doutorado – Programa de Pós-Graduação em Física. Área de Física Nuclear – Instituto Tecnológico de Aeronáutica, 2025. Orientador: Prof. Dr. César Henrique Lenzi. Coorientadora: Profª. Drª. Mariana Dutra da Rosa Lourenço. Coorientador: Prof. Dr. Sérgio José Barbosa Duarte.

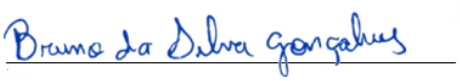
1. Equações de Estado. 2. Estrelas de Nêutrons. 3. Redes Neurais. I. Instituto Tecnológico de Aeronáutica. II. Título.

REFERÊNCIA BIBLIOGRÁFICA

GONÇALVES, Bruno da Silva. **Redes neurais bayesianas na inferência de equações de estado politrópicas para estrelas de nêutrons**. 2025. 110f. Doutorado em Física – Instituto Tecnológico de Aeronáutica, São José dos Campos, 2025.

CESSÃO DE DIREITOS

NOME DO AUTOR: Bruno da Silva Gonçalves
TITULO DO TRABALHO: Redes neurais bayesianas na inferência de equações de estado politrópicas para estrelas de nêutrons.
TIPO DO TRABALHO/ANO: Tese / 2025



Bruno da Silva Gonçalves
Av. Cidade Jardim, 679
12.233-066 – São José dos Campos–SP

REDES NEURAIS BAYESIANAS NA INFERÊNCIA DE EQUAÇÕES DE ESTADO POLITRÓPICAS PARA ESTRELAS DE NÊUTRONS

Bruno da Silva Gonçalves

Composição da Banca Examinadora:

Prof. Dr.	Brett Vern Carlson	Presidente	-	ITA
Prof. Dr.	César Henrique Lenzi	Orientador	-	ITA
Profª. Drª.	Mariana Dutra da Rosa Lourenço	Coorientadora	-	ITA
Prof. Dr.	Sérgio José Barbosa Duarte	Coorientador	-	CBPF
Prof. Dr.	Rene Felipe Keidel Spada	Membro Interno	-	ITA
Dr.	André da Silva Schneider	Membro Externo	-	UFSC
Dr.	Rodolfo Valentim da Costa Lima	Membro Externo	-	UNIFESP

ITA

Em memória ao Prof. Dr.Manuel Máximo Bastos Malheiro de Oliveira, gravo suas seguintes palavras.

"A vida é dinâmica!"

Agradecimentos

Primeiramente agradeço a Deus, pois sem ele nada seria possível. Em seguida, deixo meu agradecimento a pessoa que me apoia em todos os aspectos há 17 anos, minha esposa Joeni Luiza Goulart Gonçalves.

Também gostaria de agradecer às seguintes pessoas; Ao meu orientador, Prof. Dr. César Henrique Lenzi, que além de fundamental para a conclusão deste trabalho, tornou-se um grande amigo. Aos meus coorientadores, Profª. Drª. Mariana Dutra da Rosa Lourenço e Prof. Dr Sérgio José Barbosa Duarte por me direcionarem durante esses importantes passos rumo a astrofísica.

Ao corpo docente e equipe do Programa de Pós-Graduação em Física, por todo conhecimento construído; a empresa Power Of Data que em parceria com a pesquisa forneceu suporte técnico e recursos computacionais indispensáveis; ao apoio do amigo Bruno Jardim, CEO fundador da mesma.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de estudos que tornou factível a concretização deste trabalho.

"I do not know what I may appear to the world; but to myself I seem to have been only like a boy playing on the seashore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me."

— Isaac Newton

Resumo

A análise das equações de estado que descrevem a matéria no interior das estrelas de nêutrons contribui para a compreensão de dois pilares fundamentais da física, a matéria nuclear e a gravitação. Observações recentes, como o evento GW170817, marco inicial da era multimensageira com contrapartida gravitacional, e as medições de massa e raio realizadas pelo NICER para os pulsares PSR J0030+0451 e PSR J0740+6620, geralmente trazem evidências que desafiam explicações imediatas, desencadeando uma série de estudos dedicados à conexão entre teoria e observação. Somados ao avanço computacional, esses resultados motivaram a adoção de métodos de aprendizado de máquina aplicados à física da matéria densa. Neste contexto, o presente estudo aplicou redes neurais bayesianas à inferência de equações de estado politrópicas por partes. Para viabilizar o treinamento, foi gerado um banco de dados de quase cinco milhões de pares massa - raio, obtidos a partir do formalismo politrópico generalizado, que serviu como base algorítmica para a geração em larga escala dos dados de interesse. Sobre essa base, projetou-se uma arquitetura enxuta, com ~ 1164 parâmetros treináveis, *priors* físicos, perda de Huber e regularização KL, capaz de realizar inferência totalmente probabilística. Os testes mostraram alta precisão em cenários controlados (erros absolutos $< 10\%$), sobreposição quase completa com equações de estado de referência em cenários de validação, e consistência com observações reais como as dos estudos citados anteriormente. Em conjunto, os resultados consolidam um método modular e extensível, capaz de quantificar incertezas epistêmicas e acomodar restrições, fortalecendo a aplicação de redes neurais bayesianas na exploração da física e matéria das estrelas de nêutrons.

Abstract

The analysis of the equations of state that describe matter in the interiors of neutron stars advances our understanding of two fundamental pillars of physics, nuclear matter and gravitation. Recent observations-such as the GW170817 event, the inaugural milestone of the multimessenger era with a gravitational-wave counterpart, and NICER's mass and radius measurements for the pulsars PSR J0030+0451 and PSR J0740+6620, often yield evidence that resists immediate explanation, prompting a series of studies devoted to bridging theory and observation. Together with advances in computing, these results have motivated the adoption of machine-learning methods applied to the physics of dense matter. In this context, the present study employs Bayesian neural networks to infer piecewise polytropic equations of state. To enable training, we generated a dataset of nearly five million mass-radius pairs, obtained from the generalized polytropic formalism, which served as the algorithmic basis for large-scale data generation. Building on this foundation, we designed a compact architecture with ~ 1164 trainable parameters, physically informed priors, Huber loss, and KL regularization, capable of fully probabilistic inference. Tests show high accuracy in controlled settings (absolute errors $< 10\%$), near-complete overlap with reference equations of state in validation scenarios, and consistency with real observations such as those cited above. Taken together, the results establish a modular and extensible method that quantifies epistemic uncertainties and accommodates constraints, strengthening the application of Bayesian neural networks to the exploration of neutron-star physics and dense matter.

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas e Siglas

AdamW	Adaptive Moment Estimation with decoupled Weight Decay
AM	Aprendizado de Máquina
CV	Camada Variacional
DC	Decomposição de Cholesky
DF	Dataframe
DI	Densidade de Interface
EAM	Erro Absoluto Médio
EdE	Equação de Estado
EN	Estrela de Nêutrons
EQM	Erro Quadrático Médio
EsC	Equações de Campo
GW	Gravitational Wave
IA	Inteligência Artificial
IC	Intervalo de Confiança
ISS	International Space Station
LIGO	Laser Interferometer Gravitational-wave Observatory
MCMC	Markov Chain Monte Carlo
NICER	Neutron Star Interior Composition Explorer
PH	Perda Huber
PPG	Politrópicas por Partes Generalizadas
REQM	Raiz do Erro Quadrático Médio
RG	Relatividade Geral
RN	Rede Neural
RNB	Rede Neural Bayesiana
SNIa	Supernovas do tipo Ia
TFP	TensorFlow Probability
TOV	Tolman Oppenheimer Volkoff
WD	Weight Decay

Lista de Símbolos

c	Velocidade da luz
c_s	Velocidade do som
c_v	Calor específico a volume constante
G	Constante gravitacional universal
h	Entalpia específica
k	Constante politrópica
k_B	Constante de Boltzmann
M	Massa da estrela
$M_\odot$	Massa solar
$\mathcal{M}$	Massa molar
n_B	Densidade de número bariônico
p	Pressão
$\mathcal{R}$	Constante universal dos gases
R	Raio da estrela
ϵ	Densidade de energia total
ε	Energia interna específica
γ	Índice adiabático
Γ	Expoente politrópico
$\hat{\Gamma}$	Índice politrópico efetivo
Λ	Termo de pressão estática
ρ	Densidade de massa de repouso
T	Temperatura absoluta

Sumário

1 Introdução

As estrelas de nêutrons figuram entre os objetos mais densos, enigmáticos e intrigantes de todo o universo. Formadas a partir do colapso gravitacional de estrelas massivas em explosões chamadas de supernovas, esses corpos compactos podem conter uma massa maior que a do Sol comprimida em uma esfera de raio de ~ 12 km. Com esse nível de densidade, tais estrelas se tornam laboratórios naturais para o estudo de regimes extremos da matéria.

Nas regiões centrais desses objetos, a alta compressão gerada pela força gravitacional faz com que a densidade de números de bárions seja maior que a densidade de saturação de matéria nuclear (Lattimer; Prakash, 2001). Esse efeito é contrabalançado pelo gradiente de pressão, que reflete as propriedades da matéria fria e densa submetida à interação forte.

Embora não permita acessar diretamente os detalhes da estrutura interna, o estudo da equação de estado estabelece conexão entre as propriedades macroscópicas observáveis das estrelas de nêutrons e a física da matéria densa. Esse vínculo torna a equação de estado uma ferramenta indispensável para explorar cenários relevantes na física nuclear e astrofísica, incluindo a possível ocorrência de uma transição de fase hádron–quark em seu núcleo (Alford *et al.*, 2008; Rodrigues; Duarte; de Oliveira, 2011), hipótese que sustenta a existência de estrelas híbridas (Soares *et al.*, 2023; Lattimer, 2012; Oertel *et al.*, 2017).

A descrição detalhada da equação de estado de estrelas de nêutrons ainda se apoia em diversas construções fenomenológicas, que incluem a degenerescência de férmions (decorrente do Princípio de Exclusão de Pauli), e as interações fortes que governam as partículas no núcleo.

Nesse contexto, as contínuas observações astrofísicas têm revelado novas perspectivas sobre as propriedades dessas estrelas. Um exemplo marcante, fruto da colaboração entre os observatórios LIGO (*Laser Interferometer Gravitational-Wave Observatory*) e Virgo, é o evento GW170817 que marcou o início da era das observações multimensageiras que incluem informação de ondas gravitacionais (Abbott *et al.*, 2017).

Paralelamente, medições realizadas pelo NICER (*Neutron Star Interior Composition Explorer*) acoplado a ISS (*International Space Station*) permitiram determinar o raio e a massa de objetos compactos como, PSR J0030 + 0451 e PSR J0740 + 6620, contribuindo

para um entendimento mais aprofundado de suas estruturas internas (Miller *et al.*, 2019; Riley *et al.*, 2019; Miller *et al.*, 2021; Riley *et al.*, 2021).

Esses avanços têm estimulado discussões acerca de potenciais discrepâncias (Cooney; Dedeo; Psaltis, 2010; Moraes *et al.*, 2018) entre os resultados observacionais (Demorest *et al.*, 2010; Antoniadis *et al.*, 2013; Raaijmakers *et al.*, 2021) e aspectos previstos a partir de modelos gravitacionais alternativos, destacando a importância de investigar ainda mais os limites dessas teorias.

Diante desse cenário empolgante, abordagens estatísticas vêm ganhando cada vez mais relevância no refinamento de equações de estado. Observações como as citadas há pouco oferecem restrições que podem ser incorporadas a inferência bayesiana, permitindo-se obter distribuições de probabilidade para parâmetros que descrevem a equação de estado. Essa estratégia possibilita ir além de resultados determinísticos ao estimar de forma mais robusta a região de parâmetros compatíveis com os dados observacionais.

Como alternativa ou complemento aos métodos estatísticos, técnicas de aprendizado de máquina, em particular as redes neurais profundas, têm sido cada vez mais empregadas (Traversi; Char, 2020). Essas redes têm demonstrado grande capacidade de lidar com dados complexos e de alta dimensionalidade em diversas áreas da física. Para o caso das estrelas de nêutrons, esses modelos podem ser treinados por meio de dados gerados computacionalmente, visando prever cenários que envolvam diferentes EdEs e distintas configurações de massa e raio (Fujimoto; Fukushima; Murase, 2021).

O potencial de tais técnicas se deve à sua habilidade de capturar relações complexas entre variáveis físicas, oferecendo maior precisão e eficiência em comparação com métodos tradicionais. Além disso, o crescimento no volume de dados observacionais e o aumento da disponibilidade de poder computacional têm acelerado o desenvolvimento e a adoção dessas abordagens, tornando-as ferramentas cada vez mais atrativas.

Dentre vários exemplos de adesão, em estudos recentes, as redes neurais profundas começaram a ser aplicadas à análise de ondas gravitacionais oriundas da fusão de buracos negros (George; Huerta, 2018), contribuindo para refinar as inferências sobre as propriedades internas desses objetos compactos.

Com esse cenário, à medida que mais dados são coletados e as técnicas de modelagem e inferência avançam, a expectativa é de que surjam respostas mais claras, embora certamente acompanhadas de novas questões, sobre a natureza fundamental das estrelas de nêutrons.

Nesse sentido, a aplicação de técnicas que agrupam conceitos de aprendizado profundo e de estatística Bayesiana podem ampliar as possibilidades de investigação e trazer novas perspectivas sobre o estado da matéria em regimes de alta densidade. Haja vista que esse tipo de abordagem pode não somente buscar reproduzir adequadamente as propriedades

macroscópicas conhecidas desses objetos, como também fornecer estimativas probabilísticas para as quantidades-chave da equação de estado, utilizando conceitos de probabilidade Bayesiana.

O presente trabalho, dedicado à aplicação de redes neurais bayesianas na estimativa de equações de estado para estrelas de nêutrons, está organizado da seguinte forma: o capítulo 1 apresenta o contexto geral, motivação e objetivos da problemática; o capítulo 2 reúne a fundamentação teórica sobre a estrutura estelar, introduz o tema de equações de estado e detalha o formalismo de politrópico por partes utilizado na geração em larga escala de dados para os seguinte passos do estudo; no capítulo 3, revisamos a evolução histórica das redes neurais artificiais, descrevemos em detalhe a arquitetura da rede neural bayesiana proposta, a lógica computacional sobre a inferência variacional e o comportamento das métricas de treinamento obtidas; ainda no mesmo capítulo, é conduzida uma análise sobre o desempenho do modelo aplicado em dois cenários distintos - um deles dedicado a validação inicial para sua utilização, e outro estruturado idealmente sobre um cenário plausível atrelado aos dados do NICER, avaliando sua capacidade de generalização e nível de incerteza; por fim, o capítulo 4 sintetiza as principais conclusões, discute o impacto dos resultados e apresenta perspectivas para trabalhos futuros.

2 Estrelas de Nêutrons

Sabemos que, de modo geral, para se obter resultados razoáveis ao utilizar métodos de aprendizado de máquina (AM), tanto a qualidade quanto a quantidade de dados fornecidos ao modelo são de extrema importância[1].

Nesse sentido, no atual capítulo abordaremos de forma breve aspectos históricos, conceituais e teóricos para a formação do conhecimento necessário sobre as estrelas de nêutrons, passando por dois pontos essenciais, as equações de estado e as soluções de massa-raio (M- R) obtidas a partir das Equações de Campo (EsC) da Relatividade Geral (RG). Em sequência, veremos o formalismo paramétrico responsável pela ampla geração de equações de estado que, por fim, nos permite modelar os dados que serão a base do treinamento do modelo de rede neural (RN).

2.1 Breve História das Remanescentes de Super Novas

Visível em plena luz do dia em um intervalo de tempo entre julho de 1054 a abril de 1056, astrônomos árabes, chineses e japoneses acompanharam e tentaram compreender as implicações astrológicas de um novo astro que surgiu no céu. Tal intrigante fenômeno, hoje conhecido como a Nebulosa do Caranguejo, permaneceu sem a devida compreensão por quase um milênio, mais precisamente até 1931 quando o astrônomo Fritz Zwicky cunhou o termo *supernova* referindo-se a eventos cataclísmicos de explosões estelares responsáveis por liberar uma quantidade colossal de energia.

Um pouco mais tarde, cerca de dois anos após James Chadwick ter descoberto a existência dos nêutrons, Zwicky junto do astrônomo Walter Baade, estimando valores para a perda de massa na explosão de supernovas, propuseram que esse tipo de evento era parte de um processo onde estrelas massivas colapsam sob sua própria gravidade formando um objeto denso compostos predominantemente por nêutrons (Baade; Zwicky, 1934a; Baade; Zwicky, 1934b). Em outras palavras, ambos sugeriram a existência das estrelas de nêutrons (ENs) como um mecanismo unificador por trás dos fenômenos de

[1] Vale ressaltar que cada problema e método de AM utilizado tem suas particularidades.

supernovas e raios cósmicos, uma vez que a energia liberada por algumas supernovas era muito maior do que as que poderiam ser geradas por estrelas comuns.

Em ordem cronológica, partindo dos cálculos realizados por Oppenheimer e Volkoff, onde obteve-se o intervalo, expresso em unidade de massa solar ($M_\odot$), de $0{,}3M_\odot < M < 0{,}75M_\odot$ (Oppenheimer; Volkoff, 1939), os primeiros resultados sobre as famigeradas ENs evidencia que a existência desses objetos singulares já era amplamente discutida mesmo antes da descoberta dos pulsares e da compreensão de sua conexão com as ENs.

Em 1967 a astrofísica Jocelyn Bell Burnell (Hewish *et al.*, 1968), na época estudante de doutorado na Universidade de Cambridge, analisando a fonte de um sinal de rádio incomum que sistematicamente se repetia a cada 1,337 segundos definiu que o mesmo tratava-se de um pulsar.

No decorrer da década, seguido dos respectivos avanços sobre a astronomia de raio - X e a publicação de Bell e Anthony Hewish, intensos debates na comunidade sobre a natureza do sinal eventualmente consolidaram os pulsares como sendo um tipo de EN altamente magnetizada (Pacini, 1967; Gold, 1968). Vale-se destacar que apenas esses objetos compactos em rotação poderiam configurar pulsos de rádio com o período detectado, pondo fim ao que até então havia sido apenas teorizado por Baade e Zwicky.

2.2 Formação das Estrelas de Nêutrons

De forma sucinta, conforme nosso objetivo, as estrelas formam-se a partir de nuvens compostas, basicamente, de hidrogênio e poeira cósmica, constituída por grãos de silicatos, óxidos, grafite e gelo que, sob a influência de fatores como ondas de choque, rotação e pressão, tornam-se gravitacionalmente instáveis.

Regida pela força da gravidade, a região central da nuvem inicia um processo de colapso gravitacional, no qual o gás e a poeira ao redor do núcleo são progressivamente atraídos para o centro. As camadas externas, com momento angular mais significativo, permanecem em órbita ou se dissipam, resultando na formação de uma protoestrela no núcleo denso da estrutura.

Na sequência evolutiva, o aumento da temperatura, decorrente da elevação da densidade durante o colapso gravitacional, conduz ao instante de ignição da fusão termonuclear. A partir desse ponto, as estrelas atingem um estado de equilíbrio hidrostático, no qual a força atrativa da gravidade é contrabalançada pelas pressões térmica e radiativa geradas em seus interiores (de Souza; de Fátima, 2014).

Em uma estrela como o Sol, as pressões internas são sustentadas por meio das reações

de fusão nuclear como, por exemplo, a que envolve átomos de hidrogênio ^{1}H formando um núcleo de hélio ^{4}He, com a diferença de massa/energia de ligação entre os elementos individuais e o constituinte final sendo a energia liberada na reação.

Com relação ao balanço entre a pressão de radiação e a gravidade, após o esgotamento do hidrogênio no núcleo, a estrela passa a fundir elementos progressivamente mais pesados em estágios cada vez mais breves. Entretanto, a energia de ligação por núcleon, que determina o aquecimento do núcleo ao sustentar a fusão e apresenta um aumento expressivo nas fases iniciais, não cresce de forma linear, nem as reações nucleares permanecem exotérmicas até o seu término (Clayton, 1983).

Esse limite, conhecido como ponto do ferro, ocorre porque a formação de núcleos mais pesados que ^{56}Fe passa a demandar mais energia do que libera, marcando o início das reações endotérmicas. Como consequência, a fusão deixa de fornecer sustentação térmica suficiente, permitindo que a gravidade vença as pressões internas e desencadeie o colapso gravitacional do núcleo, processo que só ocorre em estrelas suficientemente massivas, não nas de baixa massa.

Dependendo da massa inicial, diferentes caminhos evolutivos são possíveis. No caso das estrelas com até $\sim 8\,M_{\odot}$, o núcleo nunca atinge temperaturas e pressões suficientes para fundir elementos além do carbono e oxigênio. Após o esgotamento do hélio, a estrela torna-se uma *gigante vermelha*, expandindo suas camadas externas. Durante essa fase, pulsações e instabilidades ejetam o envelope externo, formando uma *nebulosa planetária*. O remanescente é uma *anã branca*, composta tipicamente de hélio, carbono e oxigênio, cuja sustentação é fornecida pela pressão de elétrons degenerados ($p_e \propto \rho_e^{5/3}$ no regime não relativístico, onde ρ_e é a densidade de elétrons). Como anãs brancas típicas já se encontram em temperaturas centrais muito inferiores às necessárias para a fusão nuclear, sua estrutura de equilíbrio hidrostático é dominada por efeitos quânticos de degenerescência, e a relação massa-raio depende pouco da temperatura. A massa máxima suportada por este mecanismo é $M \approx 1{,}4\,M_{\odot}$, conhecida como limite de Chandrasekhar (Chandrasekhar, 1931).

Para estrelas com massas iniciais na faixa[2] de $8\,M_{\odot} \lesssim M \lesssim 30\,M_{\odot}$, a fusão progride até a formação de um núcleo de ferro. Quando a massa inercial do núcleo excede $\approx 1,4\,M_{\odot}$, a pressão de degenerescência dos elétrons não consegue mais sustentar o equilíbrio, e o colapso hidrodinâmico se inicia. À medida que a densidade ultrapassa $\rho \sim 10^{9}\,\mathrm{g\,cm^{-3}}$, os elétrons tornam-se relativísticos e são capturados por prótons:

$$e^{-} + p \rightarrow n + \nu_e, \tag{2.1}$$

[2] A separação entre quais estrelas formam estrelas de nêutrons ou buracos negros não é estritamente definida. Simulações em (O'Connor; Ott, 2011) mostram que outros fatores, como metalicidade, rotação e estrutura interna, influenciam fortemente o desfecho.

processo de neutronização que reduz Y_e e remove energia via emissão de neutrinos. Simultaneamente, a fotodesintegração dos núcleos de ferro

$$\gamma + {}^{56}\mathrm{Fe} \rightarrow 13\,{}^{4}\mathrm{He} + 4n - 124\,\mathrm{MeV} \tag{2.2}$$

consome energia, acelerando a queda de pressão. Esses mecanismos conjugados levam ao colapso gravitacional do núcleo.

Ao atingir temperaturas de $\sim 10^{11}\,\mathrm{K}$, a energia térmica aumenta devido à compressão, e os neutrinos ficam temporariamente aprisionados no núcleo da protoestrela de nêutrons (PNS). O colapso é interrompido pelo aumento abrupto da rigidez da equação de estado quando a densidade nuclear é alcançada, e não pela pressão dos neutrinos. Nesse ponto, forma-se uma onda de choque que inicialmente perde energia, mas que pode ser revitalizada pelo aquecimento por neutrinos, transferindo impulso e energia cinética às ondas de choque, permitindo que estas expulsem as camadas externas da estrela. O evento resultante é uma supernova do tipo II[3].

Após a explosão, a PNS sofre um intenso processo de *cooling* (resfriamento) por emissão de neutrinos, reduzindo sua temperatura de $\sim 10^{11}\,\mathrm{K}$ para $\sim 10^{9}\,\mathrm{K}$ em dezenas de segundos, e contraindo-se até se tornar uma estrela de nêutrons estável (Shapiro; Teukolsky, 1983).

Embora constituídas principalmente de nêutrons, essas estrelas também contêm prótons, léptons e, em regimes mais internos e energéticos, podem conter o octeto bariônico completo, mésons ou até plasma de quarks e glúons, confinados em um raio de $\sim 12\,\mathrm{km}$ e densidade central de $\sim 10^{15}\,\mathrm{g/cm^3}$.

Em estrelas com massas iniciais $M \gtrsim 30\,M_\odot$, nem mesmo a pressão de degenerescência dos nêutrons é suficiente para sustentar o equilíbrio. O colapso prossegue até formar um buraco negro, caracterizado por um horizonte de eventos do qual nem a luz escapa. Para casos de simetria esférica e estática, a solução é a métrica de Schwarzschild (Schwarzschild, 1916).

[3]Existem as supernovas do tipo Ia, que ocorrem em anãs brancas localizadas em sistemas binários fechados. Nessas configurações, a anã branca pode acumular massa a partir de sua estrela companheira por meio de acreção, processo em que o material flui de uma estrela para outra devido à transferência de momento angular, seja via transbordamento do lóbulo de Roche ou através de ventos estelares capturados gravitacionalmente, que também resultam em uma estrela de nêutrons.

2.3 Equações de Tolman-Oppenheimer-Volkoff

Notando-se as propriedades macroscópicas das ENs como sua rápida rotação com períodos da ordem de segundos ou milissegundos, magnitude dos campos magnéticos na ordem de 10^{11} a 10^{13} G, densidade $\sim 10^{15} g/cm^3$, compacticidade

$$C = \frac{GM}{Rc^2}, \tag{2.3}$$

a razão entre sua M e R, multiplicada pela constante da gravitação universal G e a velocidade da luz c, com valores típicos de $C_{EN} \sim 0{,}1$. Seus extremos potenciais gravitacionais inviabilizam uma descrição estrutural por meio de conceitos puramente newtonianos. Sendo assim, para a adequada compreensão desses objetos, veremos a seguir o necessário formalismo teórico o qual é embasado pela teoria da RG proposta por Albert Einstein (Einstein, 1915).

Apontando agora para o formalismo da RG, recordemos que a gravidade manifesta-se na própria geometria do espaço-tempo por meio da métrica

$$ds^2 = g_{\mu\nu}\, dx^\mu dx^\nu, \tag{2.4}$$

onde $g_{\mu\nu}$ é o tensor métrico, seguido pelo diferencial das coordenadas x^μ e x^ν.

A partir dela constrói-se a conexão sem torção de Levi-Civita,

$$\Gamma^\alpha_{\mu\nu} = \tfrac{1}{2}\, g^{\alpha\rho}(\partial_\mu g_{\rho\nu} + \partial_\nu g_{\rho\mu} - \partial_\rho g_{\mu\nu})\,, \tag{2.5}$$

que nos permite definir a derivada covariante ∇_α, e obter o tensor de Riemann a partir do comutador de duas derivadas covariantes,

$$R^\rho{}_{\mu\alpha\nu} = \partial_\alpha \Gamma^\rho_{\mu\nu} - \partial_\nu \Gamma^\rho_{\mu\alpha} + \Gamma^\rho_{\alpha\beta}\Gamma^\beta_{\mu\nu} - \Gamma^\rho_{\beta\nu}\Gamma^\beta_{\mu\alpha}. \tag{2.6}$$

Em seguida, por contração obtemos o tensor de Ricci

$$R_{\mu\nu} = R^\alpha{}_{\mu\alpha\nu} = \partial_\alpha \Gamma^\alpha_{\mu\nu} - \partial_\nu \Gamma^\alpha_{\mu\alpha} + \Gamma^\alpha_{\alpha\beta}\Gamma^\beta_{\mu\nu} - \Gamma^\alpha_{\beta\nu}\Gamma^\beta_{\mu\alpha}, \tag{2.7}$$

o qual novamente contraído, chega-se ao escalar,

$$R = R^\mu_\mu = g^{\mu\nu} R_{\mu\nu}. \tag{2.8}$$

Indicado que a estrutura métrica é o objeto fundamental que contém a dinâmica gravitacional, é correto afirmar que todas as informações sobre aceleração gravitacional

estão contidas em $g_{\mu\nu}$. Em conformidade com a ação de Einstein-Hilbert,

$$S = \int d^4x\sqrt{-g}\left(\frac{R}{2\chi} + L\right), \tag{2.9}$$

onde g refere-se ao determinante do tensor métrico ($\det|g_{\mu\nu}|$), R o escalar de curvatura, L a densidade lagrangeana de matéria e χ uma constante[4]. Para obtermos a dinâmica de $g_{\mu\nu}$ devemos aplicar o princípio da mínima ação sobre a mesma.

O procedimento para a RG é análogo ao da mecânica clássica, notando-se que o "campo" dinâmico não é uma coordenada, e sim a própria métrica que trás informações sobre a curvatura do espaço-tempo.

A variação $\delta S = 0$ segue, de forma condensada, os passos:

i) Variação da medida de volume:

$$\delta\big(\sqrt{-g}\big) = -\tfrac{1}{2}\sqrt{-g}\, g_{\mu\nu}\, \delta g^{\mu\nu}. \tag{2.10}$$

ii) Variação do termo de curvatura:

$$\sqrt{-g}\, g^{\mu\nu}\delta R_{\mu\nu} = \sqrt{-g}\, \nabla_\alpha V^\alpha, \tag{2.11}$$

iii) Reunião dos termos gravitacional e material, usando a definição do tensor energia-momento:

$$T_{\mu\nu} = -\tfrac{2}{\sqrt{-g}}\, \delta\big(\sqrt{-g}\, L\big)/\delta g^{\mu\nu}. \tag{2.12}$$

iv) Como $\delta g^{\mu\nu}$ é arbitrário[5], obtém-se do princípio da mínima ação sobre a Eq.(2.9) as equações de campo:

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}. \tag{2.13}$$

Compondo um conjunto não-linear em $g_{\mu\nu}$ de 10 equações diferenciais parciais, com o termo a esquerda, chamado de tensor de Einstein ($G_{\mu\nu}$), responsável por descrever a gravidade como sendo a curvatura da métrica no espaço-tempo, enquanto no lado direito $T_{\mu\nu}$ é associado à distribuição de matéria-energia contida nos limites estabelecidos ao campo.

[4]Note-se que a constante $\chi = 8\pi G/c4$, é obtida por meio da aproximação no limite de campo fraco.

[5]Devido a $\delta g^{\mu\nu}$ poder ser escolhido livremente em cada ponto, o lema fundamental do cálculo variacional exige que o fator que o acompanha em δS seja nulo.

Tendo em mente que nosso objetivo é solucionar as equações de equilíbrio da estrutura estelar (Tolman, 1939; Oppenheimer; Volkoff, 1939), a priori, assumimos que nossa estrela é simetricamente esférica e estática. Definições que implicam na invariância das propriedades físicas sobre rotações e mudanças temporais em sua geometria, permitindo-nos soluções independentes do tempo.

Através das condições acima, a métrica mais geral em tais condições é escrita[6]

$$ds^2 = e^{2\nu(r)}dt^2 - e^{2\lambda(r)}dr^2 - r^2\left(d\theta^2 + \operatorname{sen}^2\theta d\Phi^2\right), \tag{2.14}$$

onde, nesse caso, o tensor métrico fica definido como

$$g_{\mu\nu} = \operatorname{diag}\left(e^{2\nu(r)}, -e^{2\lambda(r)}, -r^2, -r^2\operatorname{sen}^2\theta\right) \tag{2.15}$$

Além das exigências anteriores, consideramos que a estrela é composta por um fluido perfeito, cuja evolução local é descrita por transformações adiabáticas. Nessas condições, o tensor energia-momento assume a forma

$$T^{\mu\nu} = (\epsilon + p)\, u_\mu u_\nu - p g_{\mu\nu}, \tag{2.16}$$

com p, ϵ e $u_\mu = dx_\mu/d\tau$, sendo a pressão, densidade de energia e quadri-velocidade de um elemento do fluido, respectivamente.

Trazendo significado a sigla TOV, formulada por Richard Chase Tolman, Julius Robert Oppenheimer e George Michael Volkoff, para obtermos a equação que descreve a variação da pressão em função do R da estrela partiremos da seguinte relação

$$dM = \epsilon(r)\, dV, \tag{2.17}$$

onde dM é o elemento infinitesimal de massa-energia e $\epsilon(r)$ é a densidade de energia em função do raio pelo elemento de volume dV.

Por simplicidade, adotamos a notação em que $c = 1$. Considerando ainda que a estrela possui simetria esférica e fazendo uso da equivalência entre massa e energia estabelecida pela Relatividade Restrita, podemos reescrever a Eq.(2.17) como

$$\frac{dM}{dr} = 4\pi r^2 \epsilon(r). \tag{2.18}$$

[6]Existem quatro principais razões para escolhas dos fatores exponenciais. 1° Sua positividade mantém a assinatura $(+,---)$, sem restrições extras. 2° Simplificação algébrica no $R_{\mu\nu}$. 3° A interpretação de $\nu(r)$ no limite fraco, favorece a ligação da RG à mecânica clássica. 4° Com $e^{-2\lambda} = 1 - 2GM(r)/r$, a função de massa contida dentro do raio $M(r)$, que será introduzida mais adiante, surge naturalmente, facilitando a resolução das equações de TOV.

Estabelecido as condições junto a métrica e introduzido a relação massa-energia do fluido pela Eq.(2.18), restam os dois potenciais métricos desconhecidos, $\nu(r)$ e $\lambda(r)$, que devem ser obtidos dinamicamente.

O caminho natural consiste em inseri-los nas EsC da RG Eq.(2.13), cujos termos de fonte são dados pelo tensor energia-momento de fluido perfeito definido na Eq.(2.16).

O procedimento desenvolve-se nos seguintes passos:

- **Cálculo das conexões de Levi-Civita:** Substitui-se a métrica definida na Eq.(2.14) na Eq.(2.5). As únicas derivadas envolvidas são de $\nu(r)$ e $\lambda(r)$, pois $g_{\theta\theta}$ e $g_{\phi\phi}$ dependem apenas de r.

- **Construção do tensor de Ricci:** As conexões obtidas no passo anterior são inseridas em $R^{\rho}{}_{\mu\alpha\nu}$. A contração $R_{\mu\nu} = R^{\alpha}{}_{\mu\alpha\nu}$ gera as seguintes quatro componentes independentes, todas escritas em termos de $\nu(r)$, $\lambda(r)$ e suas derivadas

$$R_{00} = \left[-\nu''(r) - \nu'^2(r) + \nu'(r)\lambda'(r) - 2\frac{\nu'(r)}{r}\right] e^{2[\nu(r)-\lambda(r)]}, \tag{2.19}$$

$$R_{11} = \nu''(r) + \nu'^2(r) - \nu'(r)\lambda'(r) - 2\frac{\nu'(r)}{r}, \tag{2.20}$$

$$R_{22} = [r\nu'(r) - r\lambda'(r) + 1]\, e^{-2\lambda(r)} - 1, \tag{2.21}$$

$$R_{33} = \left[(r\nu'(r) - r\lambda'(r) + 1)\, e^{-2\lambda(r)} - 1\right] \sin^2\theta, \tag{2.22}$$

e por último

$$R = 2e^{-2\lambda(r)}\left[-\nu''(r) - \nu'^2(r) + \nu'(r)\lambda' - \frac{2}{r}\left(\nu'(r) - \lambda'(r)\right) - \frac{1}{r^2}\right] + \frac{2}{r^2}. \tag{2.23}$$

- **Igualdade $\mathbf{G}_{\mu\nu} = \mathbf{8\pi}\, \mathbf{T}_{\mu\nu}$:** Levando em conta a diagonalidade presente em nossa métrica ($\mu \neq \nu = 0$), com os resultados encontrados para o tensor de Ricci, obtemos os seguintes componentes do tensor de Einstein

$$G_0^0 = \left[\frac{1}{r^2} - 2\frac{\lambda'(r)}{r}\right] e^{-2\lambda(r)} - \frac{1}{r^2}, \tag{2.24}$$

$$G_1^1 = \left[\frac{1}{r^2} - 2\frac{\nu'(r)}{r}\right] e^{-2\lambda(r)} - \frac{1}{r^2}, \tag{2.25}$$

$$G_2^2 = G_3^3 = \left[\nu''(r) + \nu'^2(r) - \nu'(r)\lambda'(r) + \frac{\nu'(r) - \lambda'(r)}{r}\right] e^{-2\lambda(r)}. \tag{2.26}$$

Após calcularmos os termos referentes a curvatura nas EsC, devemos buscar seus correspondentes atrelados a fonte de matéria/energia. Sendo assim, visto que no referencial do fluido $u^\mu = (1,0,0,0)$, as componentes não nulas do tensor energia-momento, em termos da pressão e densidade de energia, são dadas por:

$$T^\mu_{\ \nu} = \mathrm{diag}(\epsilon, -p, -p, -p) \tag{2.27}$$

Substituindo cada elemento presente em $T^\mu_{\ \nu}$ por seu correspondente em $G^\mu_{\ \nu}$, encontra-se o seguinte sistema de equações

$$-8\pi G\epsilon = \left[\frac{1}{r^2} - 2\frac{\lambda'(r)}{r}\right] e^{-2\lambda(r)} - \frac{1}{r^2}, \tag{2.28}$$

$$8\pi Gp = \left[\frac{1}{r^2} - 2\frac{\nu'(r)}{r}\right] e^{-2\lambda(r)} - \frac{1}{r^2}, \tag{2.29}$$

$$8\pi Gp = \left[\nu''(r) + \nu'^2(r) - \nu'(r)\lambda'(r) + \frac{\nu'(r) - \lambda'(r)}{r}\right] e^{-2\lambda(r)}. \tag{2.30}$$

Notando-se que a relação

$$\begin{aligned}\frac{d}{dr}\left[r\left(1 - e^{-2\lambda(r)}\right)\right] &= 1 - e^{-2\lambda(r)} + r\left[2\lambda'(r)e^{-2\lambda(r)}\right] \\ &= \left[-1 + 2r\lambda'(r)\right] e^{-2\lambda(r)} + 1,\end{aligned} \tag{2.31}$$

multiplicada por r^{-2} é

$$-\frac{1}{r^2}\frac{d}{dr}\left[r\left(1 - e^{-2\lambda(r)}\right)\right] = \left[\frac{1}{r^2} - 2\frac{\lambda'(r)}{r}\right] e^{-2\lambda(r)} - \frac{1}{r^2}. \tag{2.32}$$

Utilizando a Eq.(2.28), podemos substituir o lado direito desse resultado e obter que

$$\frac{d}{dr}\left[r\left(1 - e^{-2\lambda(r)}\right)\right] = 8\pi G r^2 \epsilon. \tag{2.33}$$

Integrando ambos os lados em r,

$$r\left(1 - e^{-2\lambda(r)}\right) = \int_0^r 8\pi G r^2 \epsilon dr. \tag{2.34}$$

Definindo a massa gravitacional interna ao raio r através da Eq.(2.18) como

$$M(r) = \int_0^r 4\pi r'^2 \epsilon(r') dr', \tag{2.35}$$

podemos isolar o termo exponencial na Eq.(2.34) e obter

$$e^{-2\lambda(r)} = 1 - \frac{2GM(r)}{r}. \tag{2.36}$$

Em seguida, isolando respectivamente os termos $\lambda'(r)$ e $\nu'(r)$ por meio das Eqs.(2.28 - 2.29), podemos calcular $\nu''(r)$, $\nu'^2(r)$ e $\nu'(r)\lambda'(r)$, e substituí-los na Eq.(2.30) para obter

$$[\,p(r) + \epsilon(r)\,]\left[e^{2\lambda(r)}\left(8\pi G\, r^2\, p(r) + 1\right) - 1\right] + 2r\, p'(r) = 0. \tag{2.37}$$

Se definirmos um fator de curvatura:

$$C(r) = e^{2\lambda(r)}\left[8\pi G r^2 p(r) + 1\right] - 1, \tag{2.38}$$

rearranjando alguns termos a Eq.(2.37) pode ser reescrita como

$$p'(r) \;=\; -\frac{1}{2r}\,[p(r) + \epsilon(r)]\,C(r), \tag{2.39}$$

destacando que a variação da pressão depende diretamente da densidade total $p + \epsilon$, do fator de curvatura $C(r)$, e da razão inversa do raio.

Mantendo nosso objetivo em vista, fazendo mão da Eq.(2.36), com um pouco de álgebra sobre a relação vista na Eq.(2.37), podemos encontrar a equação TOV que define a variação da pressão em função do raio

$$\frac{dp}{dr} \;=\; -\frac{G\,\epsilon(r)\,M(r)}{r^2}\left[1 + \frac{p(r)}{\epsilon(r)}\right]\left[1 + \frac{4\pi r^3\, p(r)}{M(r)}\right]\left[1 - \frac{2GM(r)}{r}\right]^{-1}, \tag{2.40}$$

ou ainda, de forma mais compacta

$$\frac{dp}{dr} \;=\; -\frac{G\,[\epsilon(r) + p(r)]\,[M(r) + 4\pi r^3 p(r)]}{r^2\left[1 - \frac{2GM(r)}{r}\right]}. \tag{2.41}$$

Por fim, conduzindo nossa métrica ao caso do vácuo ($T_{\mu\nu} = 0$), ao somarmos as relações obtidas para o cálculo da Eq.(2.37) de λ' e ν', ao integrarmos ambos os termos em relação a r, temos que

$$\nu(r) + \lambda(r) = \mathcal{C}, \tag{2.42}$$

onde $\mathcal{C}$ trata-se da constante de integração.

Em contrapartida, a exigência de que a métrica seja assintoticamente plana implica que

$$\lim_{r\to\infty}[\nu(r)+\lambda(r)]=0, \tag{2.43}$$

pois em regiões distantes da estrela, onde o campo gravitacional se torna desprezível, espera-se que a métrica converja para a métrica de Minkowski, ou seja, $g_{\mu\nu}\to\eta_{\mu\nu}$. Contudo, levando-se em conta o resultado da Eq.(2.42), conclui-se que para qualquer valor de r podemos escrever

$$e^{2\nu(r)} \;=\; e^{-2\lambda(r)}, \tag{2.44}$$

que, a partir da Eq.(2.36), nos permite definir

$$e^{2\nu(r)} \;=\; 1-\frac{2GM(r)}{r}. \tag{2.45}$$

Até então, tratado as equações que descrevem a massa, o raio e a relação entre essas grandezas com a pressão gerada no interior das ENs, cabe-nos agora destacar que tais configurações correspondem a uma estrutura em equilíbrio hidrostático e, sendo assim, relacionadas a um mínimo ou máximo em termos de energia.

Em razão disso, evidenciamos que tal conformação não assegura a imperturbabilidade estelar desejada, o que nos leva ao princípio de Le Chatelier (Shapiro; Teukolsky, 1983), o qual afirma que, ao longo das configurações de equilíbrio correspondentes à equação de estado que satisfaz à estabilidade da matéria que compõe a estrela, vale-se que

$$\frac{dp}{d\epsilon} \;\geq\; 0. \tag{2.46}$$

Tal condição possui uma interpretação física direta. Ela assegura que a velocidade do som no meio, definida como $c_s^2=dp/d\epsilon$, seja real e não negativa. Isso garante que perturbações locais se propaguem em vez de se amplificar exponencialmente, sendo, portanto, um requisito mínimo de estabilidade causal e mecânica do fluido estelar.

Como ilustrado na Fig.(2.1), a estabilidade frente a modos de oscilação radial não é contínua ao longo de toda a curva $M(\epsilon_c)$, onde ϵ_c representa a densidade de energia central da estrela. Essa curva traça a sequência de soluções em equilíbrio hidrostático obtidas para uma dada equação de estado.

A primeira região de estabilidade estende-se desde o início da curva até o ponto A, que marca a massa máxima suportada por uma anã branca. Nesse domínio, todas as configurações satisfazem $\partial M/\partial\epsilon_c>0$, garantindo a estabilidade frente a perturbações

radiais. Após o ponto A, a derivada torna-se negativa e a curva decresce até o ponto B, definindo uma zona de instabilidade. Nessa faixa, pequenas perturbações na densidade central não podem ser compensadas por reequilíbrio de pressão, resultando em colapso gravitacional.

A segunda zona de estabilidade emerge entre os pontos B e C, delimitando as configurações estáveis associadas às estrelas de nêutrons. Nesse intervalo, o aumento da densidade central leva novamente a um aumento da massa, preservando a positividade do modo fundamental de oscilação radial. O ponto C corresponde à massa máxima de uma estrela de nêutrons estável, além da qual novas instabilidades se instalam.

Em densidades superiores, representando objetos mais compactos e com menor raio, surgem novos pontos críticos (como D e E), associados a mudanças de estabilidade de modos radiais superiores. Esses domínios conduzem a regimes onde as soluções de equilíbrio podem se tornar instáveis ou já não correspondem a estrelas, mas sim a objetos colapsados, como buracos negros.

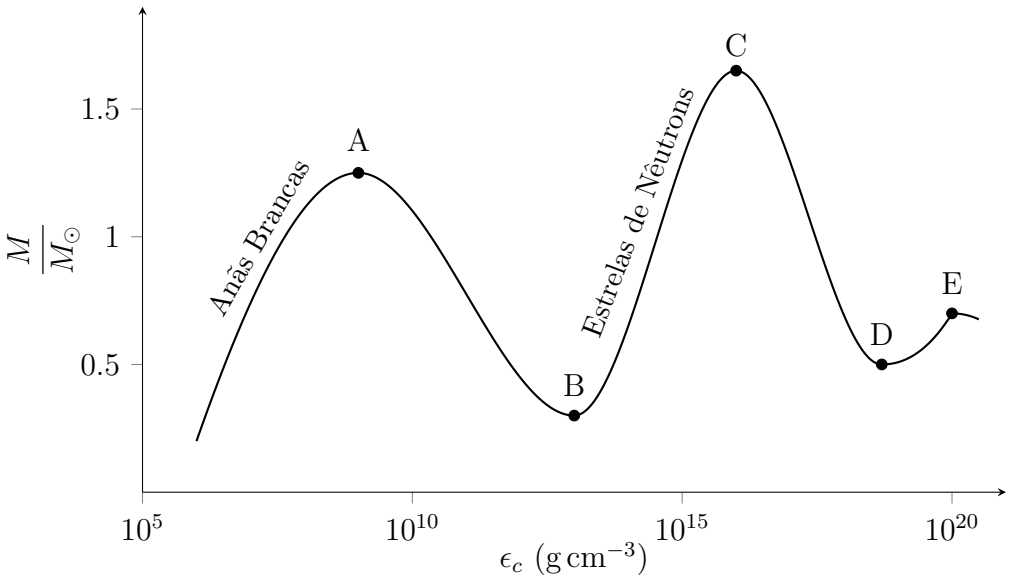


FIGURA 2.1 – Diagrama esquemático da massa em função da densidade central para configurações de equilíbrio estelar, indicando os pontos de inflexão associados a transições de estabilidade (Shapiro; Teukolsky, 1983).

Uma condição necessária, embora não suficiente, para a estabilidade estelar frente a perturbações radiais é dada por:

$$\frac{\partial M(\epsilon_c)}{\partial \epsilon_c} > 0. \tag{2.47}$$

Essa relação implica que, se a estrela for levemente perturbada, um novo estado de equilíbrio exigiria uma massa maior. A gravidade, nesse contexto, atua como uma força restauradora, estabilizando o sistema.

Por outro lado, quando a derivada se torna negativa,

$$\frac{\partial M(\epsilon_c)}{\partial \epsilon_c} < 0, \tag{2.48}$$

a configuração encontra-se em uma região instável. Perturbações levam a um desequilíbrio em que a gravidade domina de forma irreversível, provocando o colapso da estrela.

Abordados os aspectos de equilíbrio e estabilidade de estrelas compactas, a seguir aprofundaremos nossa visão em busca do entendimento teórico do principal componente na modelagem destes objetos, as equações de estado.

2.4 Equações de Estado

O que caracteriza um sistema físico? Podemos responder essa questão estabelecendo um sistema físico como um conjunto de entidades materiais, sujeito a interações descritas por leis da natureza, para o qual podemos definir:

- Um conjunto de graus de liberdade (partículas, campos, excitações coletivas).
- Uma dinâmica governada por um Hamiltoniano.
- Observáveis mensuráveis que permitem comparação entre teoria e experimento.

Todo sistema físico tem uma equação de estado associada, que estabelece relações entre as funções de estado que nos informam quanto ao estado termodinâmico do sistema. Por exemplo, pressão, entropia e temperatura são funções de estado. Em outras palavras, as equações de estado são vínculos que traduzem, em linguagem macroscópica, não apenas o conteúdo microscópico, isto é, a constituição e interações das partículas, mas também as simetrias fundamentais do sistema (como isotropia, homogeneidade ou simetrias internas), que impõem restrições adicionais à sua descrição.

Assim, um funcional mínimo para especificar o estado de fase da matéria e prever sua evolução frente a perturbações mecânicas, térmicas ou de composição é, como já reconhecemos, uma equação de estado. Elas aparecem como, restrições geométricas no espaço de estados, fechamento dinâmico para leis de conservação, condições de consistência que asseguram estabilidade, causalidade e positividade de entropia.

Entender uma equação de estado implica compreender como estatística, simetrias e escalas se combinam para produzir comportamento coletivo, fornecendo a ligação indispensável entre microfísica e fenômenos observáveis em qualquer domínio da ciência física.

Uma das atuais fronteiras do conhecimento, a determinação dos parâmetros que caracterizam uma equação de estado que descreve a dinâmica da microestrutura de uma

estrela de nêutrons, é o cerne deste trabalho.

Esse desafio depende basicamente do estabelecimento de vínculos de natureza distinta, sendo eles formais, conceituais e fenomenológicos. Buscando contribuir para esse entendimento através de uma representação simples e eficaz dentre a gama de modelos complexos existentes na literatura (Tooper, 1964), a seguinte seção será destinada a apresentação do modelo de equações as quais fornecem a estrutura teórica para a geração do conjunto de EdEs que será parte da base de informações para nosso modelo de rede neural.

2.4.1 Equações de Estado Politrópicas

Para estabelecer o formalismo de equações de estado que adotaremos no capítulo seguinte, partimos de uma convenção comum, consideraremos a velocidade da luz como unidade ($c = 1$). Tal escolha simplifica expressões e revela simetrias importantes, mas para mantermos o rigor físico necessário e evitar ambiguidades, inicialmente, faremos algumas observações diante as implicações dessa notação sobre as grandezas envolvidas no contexto.

Posto isso, a densidade de massa de repouso, doravante chamada simplesmente de **densidade de massa**, dada como a razão entre a massa e o volume $\rho = M/V$, compreendida em unidade de g/cm^3.

Por sua vez, a energia interna específica, definida como a razão entre a energia interna e a massa, $\varepsilon = U/M$, é uma grandeza adimensional nesse sistema de unidades.

Com essas definições, podemos agora introduzir a densidade de energia total do sistema, composta pela contribuição de repouso e energia interna:

$$\epsilon \; = \; \rho(1 + \varepsilon), \tag{2.49}$$

a qual será referida simplesmente por **densidade de energia**. Observe que em nossa notação, a dimensão de energia de repouso ($\rho c^2 \equiv \rho$), passa a ser descrita em unidade de g/cm^3, em concordância com a adimensionalidade mencionada a respeito do termo ε.

Vale-se ainda destacar que, embora nessa notação a massa e energia tenham as mesmas unidades numéricas, conceitualmente, a massa de repouso continua sendo quantidade conservada (bariônica), já a energia interna depende da termodinâmica local, podendo variar com temperatura e densidade.

Estabelecida essa base, passamos a considerar os fundamentos termodinâmicos que permitem relacionar essas grandezas de forma dinâmica. A forma geral da primeira lei

para sistemas fechados pode ser escrita como:

$$dU = T\,dS - p\,dV + \sum_i \mu_i\,dN_i, \tag{2.50}$$

onde T é a temperatura, S a entropia total, p a pressão, μ_i o potencial químico e N_i o número de partículas da i-ésima espécie.

Para o nosso propósito, assumindo um fluido isolado, em regime isentrópico ($dS = 0$), composto por uma única espécie (ou por um conjunto de partículas em equilíbrio) com número de partículas constante ($dN_i = 0$), a equação reduz-se a:

$$dU = -p\,dV. \tag{2.51}$$

Dividindo ambos os lados por M, lembrando que a massa é fixa ($V/M = 1/\rho$),

$$d\left(\frac{\epsilon}{\rho}\right) = -p\,d\left(\frac{1}{\rho}\right). \tag{2.52}$$

Considerando a definição da Eq.(2.49), verificamos que o termo a esquerda dessa equação é igualmente descrito por

$$d\left(\frac{\epsilon}{\rho}\right) \;=\; d(\varepsilon + 1) \;=\; d\varepsilon. \tag{2.53}$$

Por outro lado, do diferencial total do termo a direita, temos a identidade

$$d\left(\frac{1}{\rho}\right) \;=\; \frac{d}{d\rho}\left(\frac{1}{\rho}\right)d\rho \;=\; -\frac{d\rho}{\rho^2}. \tag{2.54}$$

Igualando esses dois últimos resultados,

$$d\varepsilon = \frac{p}{\rho^2}\,d\rho, \tag{2.55}$$

relação chave, será utilizada mais a frente, para obtermos a forma final pretendida de nossa EdE politrópica e respectivas relações de densidade de energia.

Para motivar naturalmente o surgimento da equação de estado politrópica, explanaremos agora um gás ideal em regime isocórico. Essa abordagem oferece uma ponte física clara entre temperatura, energia interna e pressão, elementos centrais para compreender o comportamento termodinâmico do fluido.

Nessas condições, todo o calor fornecido ao sistema contribui exclusivamente para a

variação da energia interna. Recordemos que, como não existe mudança no volume,

$$\varepsilon = c_v\, T, \tag{2.56}$$

onde c_v trata-se do calor específico a volume constante (C., 1985).

A seguir, tendo a forma clássica da EdE para gases ideais monoatômicos:

$$p \;=\; \frac{N}{V}\, k_B\, T, \tag{2.57}$$

com N o número de partículas, V o volume e k_B a constante de Boltzmann. Dado que cada partícula possui massa m (logo, $\rho = mN/V$), substituindo T pela relação fornecida na Eq.(2.56), obtém-se

$$p \;=\; \left(\frac{k_B}{m}\right)\rho\, T \;=\; \left(\frac{k_B}{m\, c_v}\right)\rho\, \varepsilon. \tag{2.58}$$

Como passo de encerramento do intervalo motivador, associando o índice adiabático[7] $\gamma = c_p/c_v$, a relação de Mayer para gases ideais (Mayer; Goeppert, 1940; Landau; Lifshitz, 1987). Verifica-se que

$$\gamma - 1 = \frac{c_p}{c_v} - 1 = \frac{R_s}{c_v} = \frac{k_B}{m\, c_v}, \tag{2.59}$$

onde $R_s = \mathcal{R}/\mathcal{M}$ é a constante específica do gás, dada como a razão entre a constante universal dos gases $\mathcal{R}$ e a massa molar $\mathcal{M}$.

Resultado que permite, através da Eq.(2.58), obter a EdE para um gás ideal:

$$p \;=\; (\gamma - 1)\, \rho\, \varepsilon, \tag{2.60}$$

reescrevendo a forma da pressão (Rezzolla; Zanotti, 2013), que servirá como base para continuarmos.

Partimos agora da Eq.(2.60) para obter a forma funcional da pressão em termos de ρ, tomando sua derivada total:

$$\frac{dp}{d\rho} \;=\; (\gamma - 1)\, \frac{d}{d\rho}\,(\rho\, \varepsilon) \;=\; (\gamma - 1)\left(\varepsilon + \rho\, \frac{d\varepsilon}{d\rho}\right). \tag{2.61}$$

Substituindo o termo $\rho\, d\varepsilon/d\rho$ pela resultado da Eq.(2.55), e expressando ε novamente através da Eq.(2.60), obtemos:

$$\frac{dp}{d\rho} = \gamma\, \frac{p}{\rho}. \tag{2.62}$$

[7] Definido como a razão entre os calores específicos a pressão e volume constante.

Separando as variáveis de pressão e densidade de massa, sua integração nos leva a

$$\int \frac{dp}{p} = \gamma \int \frac{d\rho}{\rho} \quad \Rightarrow \quad \ln p = \gamma \, \ln \rho + \ln k,$$

que, utilizando as propriedades de logaritmo natural, especificando a constante politrópica $k = e^C$, definimos

$$p = k \, \rho^{\Gamma}, \tag{2.63}$$

com $\Gamma = \gamma$ o índice[8]. Assim, pudemos ver a equação de estado politrópica surgir como consequência de um modelo de gás ideal.

Retomando a Eq.(2.55) com a forma final politrópica, Eq.(2.63), temos

$$d\varepsilon = k \, \rho^{\Gamma - 2} \, d\rho, \tag{2.64}$$

que pode ser integrada, para $\Gamma \neq 1$, fornecendo

$$\varepsilon = \alpha + \frac{k}{\Gamma - 1} \, \rho^{\Gamma - 1}, \tag{2.65}$$

sendo essa a equação que descreve a energia interna por unidade de massa no formalismo. A constante[9] de integração α será posteriormente determinada através de condições de contorno específicas, essencial para estabelecermos a estrutura final da teoria que iremos utilizar como base para geração dos dados em larga escala.

Substituindo ε na Eq.(2.49), a densidade de energia torna-se:

$$\epsilon = (1 + \alpha) \, \rho + \frac{k}{\Gamma - 1} \, \rho^{\Gamma}. \tag{2.66}$$

Após termos encontrado as equações que descrevem a pressão e a densidade de energia nesse formalismo, podemos calcular a velocidade de propagação de perturbações adiabáticas. Em meios barotrópicos, define-se:

$$c_s^2 \; = \; \left(\frac{\partial p}{\partial \epsilon} \right)_s \; = \; \frac{dp}{d\rho} \, \frac{d\rho}{d\epsilon}, \tag{2.67}$$

uma vez que p e ϵ são grandezas termodinâmicas que dependem somente da densidade de massa ($p = p(\rho) \rightarrow \epsilon = \epsilon(\rho)$).

[8] Simulações numéricas que empregam uma EdE politrópica partem da suposição implícita da igualdade entre os índices politrópico e adiabático, de modo que ϵ pode ser calculada integrando a Eq.(2.66). Nesta seção, assumiremos que $\Gamma = \gamma$ (Rezzolla; Zanotti, 2013).

[9] Adotando a convenção usual $\varepsilon(T = 0) = 0$, pode-se fixar a constante de integração como nula. Contudo, em decorrência de sua futura aplicação, a mesma será mantida nas equações.

Para calcular $d\rho/d\epsilon$, é conveniente estabelecermos uma relação que envolve a entalpia especifica e as densidades de massa e energia. Nesse sentido, expandindo ambos os diferenciais da Eq.(2.52), temos

$$\frac{d\epsilon}{\rho} - \frac{\epsilon}{\rho^2}\, d\rho \;=\; p\,\frac{d\rho}{\rho^2}, \tag{2.68}$$

que multiplicado por ρ^2, isolando $d\epsilon$ obtém-se,

$$d\epsilon \;=\; \left(\frac{\epsilon + p}{\rho}\right) d\rho. \tag{2.69}$$

Definida, por construção, como a soma da energia interna de um sistema com a de sua capacidade em realizar trabalho mecânico,

$$h \;=\; \frac{\epsilon + p}{\rho}, \tag{2.70}$$

a entalpia específica, pode ser utilizada como substituta do termo entre parênteses na equação anterior e produzir,

$$d\epsilon \;=\; h\, d\rho. \tag{2.71}$$

Com essa relação, derivando em ρ a Eq.(2.63), podemos obtendo a forma final

$$c_s^2 \;=\; \frac{1}{h}\,\frac{dp}{d\rho} \;=\; \frac{\Gamma p}{\epsilon + p}, \tag{2.72}$$

da equação que fornece a velocidade do som para o formalismo politrópico.

2.4.2 Politrópicas por Partes Generalizadas

Utilizando a Eq.(2.63) como base da parametrização na região de alta densidade das EdEs para as ENs, notemos que as equações politrópicas podem ser formuladas como uma união de termos:

$$p_i(\rho) \;=\; k_i \rho^{\Gamma_i} + \Lambda_i, \tag{2.73}$$

essa sendo a equação de estado politrópica por partes generalizada (PPG).

O índice politrópico,

$$\Gamma_i = \frac{\ln[(p_i - \Lambda_i)/(p_{\,i-1} - \Lambda_i)]}{\ln(\rho_i/\rho_{\,i-1})}, \tag{2.74}$$

obtido ao avaliar dois pontos de pressão (p_{i-1}, ρ_{i-1}) e (p_i, ρ_i) em um mesmo intervalo i.

O termo Λ_i, trata-se de uma constante que integra um conjunto de condições, o qual tem por objetivo garantir a continuidade da pressão, densidade de energia e diferenciabilidade, nas densidades de interface (DI), onde existe a segmentação dessas equações que discutiremos mais a frente.

Vale destacar que, embora a equação de estado não precise, em geral, satisfazer a condição de continuidade, o presente trabalho restringe-se à análise de equações de estado que a obedecem.

Impondo a condição de diferenciabilidade nas DIs, somos conduzidos as relações de ajuste

$$\begin{aligned}
\left.\frac{dp_i}{d\rho}\right|_{\rho_i} &= \left.\frac{dp_{i+1}}{d\rho}\right|_{\rho_i}, \\
\Gamma_i\, k_i\, \rho_i^{\Gamma_i-1} &= \Gamma_{i+1}\, k_{i+1}\, \rho_i^{\Gamma_{i+1}-1}, \\
k_{i+1} &= k_i\, \frac{\Gamma_i}{\Gamma_{i+1}} \rho_i^{\Gamma_i-\Gamma_{i+1}},
\end{aligned} \tag{2.75}$$

para a constante politrópica.

Avaliando a condição de continuidade da pressão na DI $\rho = \rho_i$, produz-se a relação

$$\begin{aligned}
p_i(\rho_i) &= p_{i+1}(\rho_i), \\
k_i\, \rho_i^{\Gamma_i} + \Lambda_i &= k_{i+1}\, \rho_i^{\Gamma_{i+1}} + \Lambda_{i+1}.
\end{aligned} \tag{2.76}$$

Recolocando o primeiro termo a direita dessa expressão pelo resultado obtido da condição de diferenciabilidade, Eq.(2.75), isola-se o termo Λ do intervalo $i+1$ como,

$$\Lambda_{i+1} = \Lambda_i + \left(1 - \frac{\Gamma_i}{\Gamma_{i+1}}\right) k_i\, \rho_i^{\Gamma_i}. \tag{2.77}$$

Considerando a Eq.(2.55), substituindo p agora pela Eq.(2.73) para um intervalo i:

$$d\varepsilon_i = \left(k_i\, \rho^{\Gamma_i-2} + \Lambda_i\, \rho^{-2}\right) d\rho. \tag{2.78}$$

Integrando termo a termo (com $\Gamma_i \neq 1$):

$$\begin{aligned}
\int d\varepsilon_i &= k_i \int \rho^{\Gamma_i-2}\, d\rho \;+\; \Lambda_i \int \rho^{-2}\, d\rho, \\
\varepsilon_i &= \frac{k_i}{\Gamma_i - 1}\, \rho^{\Gamma_i-1} \;-\; \frac{\Lambda_i}{\rho} + \alpha_i.
\end{aligned} \tag{2.79}$$

Substituindo esse resultado na definição de ϵ, obtemos

$$\epsilon_i = (1+\alpha_i)\,\rho + \frac{k_i}{\Gamma_i - 1}\,\rho^{\Gamma_i} - \Lambda_i, \tag{2.80}$$

sendo essa a expressão que define a densidade de energia do i-ésimo intervalo.

Conforme mencionado, agora, exigindo a continuidade na Eq.(2.80) sob densidade de transição, $\epsilon_i(\rho) = \epsilon_{i+1}(\rho)$, obtém-se a relação

$$\alpha_{i+1} \;=\; \alpha_i \;+\; \Gamma_i\,\frac{\Gamma_{i+1}-\Gamma_i}{(\Gamma_{i+1}-1)\;(\Gamma_i-1)}\,k_i\,\rho_i^{\Gamma_i-1}, \tag{2.81}$$

que fixa o valor da constante α para o intervalo de seguimento $i+1$.

Veremos agora a forma da equação que fornece a velocidade na qual ondas sonoras se propagam no fluido descrito por nossa EdE. Inicialmente, tomadas as mesmas condições vistas na seção anterior junto da relação para entalpia específica, a partir da definição da velocidade do som,

$$c_{s_i}^2 \;=\; \frac{1}{h_i}\left(\frac{dp_i}{d\rho}\right), \tag{2.82}$$

devemos encontrar h_i e $dp_i/d\rho$, levando em conta as características do formalismo de PPG em questão.

Nesse sentido, observando que o excerto puramente barotrópico de nossa equação de estado é $p_i - \Lambda_i = k_i\rho^{\Gamma_i}$. A derivada da pressão definida na Eq.(2.73), com relação a densidade de massa,

$$\frac{dp_i}{d\rho} \;=\; \Gamma_i\,k_i\,\rho^{\Gamma_i-1} \;=\; \frac{\Gamma_i\,k_i\,\rho^{\Gamma_i}}{\rho}, \tag{2.83}$$

pode ser escrita como,

$$\frac{dp_i}{d\rho} \;=\; \frac{\Gamma_i}{\rho}\,(p_i - \Lambda_i)\,. \tag{2.84}$$

Dado a forma encontrada para ϵ_i, junto das definições de p_i e h_i, podemos calcular

$$\begin{aligned}
h_i &= \frac{\epsilon_i + p_i}{\rho},\\
h_i &= (1+\alpha_i) \;+\; \frac{k_i}{\Gamma_i - 1}\,\rho^{\Gamma_i-1} \;-\; \Lambda_i \;+\; k_i\rho^{\Gamma_i-1} \;+\; \Lambda_i,\\
h_i &= (1+\alpha_i) \;+\; \frac{\Gamma_i}{\Gamma_i - 1}\,k_i\rho^{\Gamma_i-1},
\end{aligned} \tag{2.85}$$

evidenciando que o termo Λ_i não contribui sobre a entalpia específica, mantendo-se com a forma barotrópica padrão.

Diante desse fato, por meio do resultado visto na Eq.(2.84), temos que

$$c_{s_i}^2 \;=\; \frac{\Gamma_i\,(p_i-\Lambda_i)}{\epsilon_i+p_i}, \tag{2.86}$$

que fornece a velocidade do som dentre o i-ésimo seguimento particionado.

Como parte final dessa seção, iremos denotar alguns pontos relevantes a respeito da equação que fornece a velocidade do som, atrelados a sua futura aplicação sobre a modelagem do fluído estelar e respectiva obtenção das curvas de M-R por meio da solução das equações de TOV.

Primeiramente, podemos observar que no limite de $\Lambda_i \to 0$, a equação para $c_{s_i}^2$ recupera a forma derivada da equação politrópica padrão. Ponto positivo, tendo em vista que, em dado limite, recuperar-se o comportamento descrito pela teoria precursora é requisito básico, o qual deve se mostrar consistente frente as derivações consequentes da mesma.

Além disso, a exigência de causalidade

$$c_{s_i}^2 \;\leq\; 1, \tag{2.87}$$

implica a condição

$$\Gamma_i\,(p_i-\Lambda_i) \;\leq\; \epsilon_i+p_i, \tag{2.88}$$

que limita simultaneamente os valores admissíveis de Γ_i e a magnitude do deslocamento Λ_i em cada intervalo politrópico.

No contexto considerado, convém destacarmos que o fator presente no numerador da Eq. (2.86), responsável por expressar a parte termicamente ativa[10], representa a rigidez efetiva[11] do fluido

$$K \;=\; 9\rho\,\frac{dp}{d\rho} \;=\; 9\Gamma_i\,(p_i-\Lambda_i)\,, \tag{2.89}$$

a qual está desvinculado da contribuição puramente estática de Λ_i.

Por sua vez, no denominador, o termo que corresponde à inércia relativística total[12] $\epsilon_i + p_i$, é o responsável por resistir à propagação da perturbação acústica. Note que, apesar de deslocar individualmente ϵ_i, esse par é independente do termo Λ_i, bem como pode ser visto ao rebuscar o comentário feito na obtenção da Eq.(2.85). Nessa passagem

[10]Entende-se a fração da pressão que responde a variações de densidade e, portanto, participa na propagação de perturbações acústicas. Como Λ_i é constante com respeito a ρ ($\partial\Lambda_i/\partial\rho = 0$), ele não contribui para essa resposta e é subtraído no cálculo de $dp/d\rho$.

[11]Também conhecido como módulo de incompressibilidade, definido em matéria nuclear como $K \equiv 9\,\rho\,(\partial p/\partial\rho)$, representando a resistência volumétrica do fluido à compressão.

[12]Podemos traduzi-la como o "peso" que deve ser acelerado.

apontamos que o mesmo é eliminado no calculo de h_i e, portanto, não contribui a resposta de uma compressão infinitesimal do fluido.

Posto isso, ao analisarmos as implicações desse formalismo sobre a estrutura estelar através da Eq.(2.41), vemos que, como Λ_i desloca ϵ_i e p_i em igual magnitude, ele altera[13] a massa-energia integrada, afetando indiretamente a estrutura final sem influenciar a rigidez local. Em ultima análise, o conjunto formado pelas condições impostas de continuidade, diferenciabilidade e causalidade define um subespaço viável no domínio dos parâmetros $(k_i, \Gamma_i, \Lambda_i)$. A determinação desse subespaço possui relevância prática, onde apenas configurações nele contidas fornecerão, por meio das equações de TOV juntamente com a Eq. (2.86), sequências M-R para nosso estudo.

Sendo assim, a seção seguinte implementará as PPG na integração numérica das equações de TOV, explorando como variações em Γ_i, k_i e Λ_i moldam famílias de soluções de M-R.

2.5 Modelagem

Para prosseguirmos com efetividade em direção ao nosso objetivo, é essencial manter claras, ao longo de todo o desenvolvimento, duas ópticas sobre a aplicação do AM. Primeiramente, é preciso reconhecer a direta influência que cada fase individual do processo possui sobre a qualidade dos resultados. Em segundo, a adoção de uma visão sistêmica que integre essas etapas de forma consistente torna-se um requisito para otimizar o processo de estruturação do modelo como um todo e garantir resultados robustos e coerentes.

Nesse sentido, em consonância com a óptica sistêmica mencionada, os princípios que serão discutidos à frente foram elencados por promover um processo de treinamento congruente com as particularidades do problema em análise, contribuindo positivamente para a qualidade dos futuros resultados.

Após explorarmos o cunho teórico do modelo de EdEs, abordaremos os principais pilares necessários para estabelecermos uma base sólida, desde a etapa de geração e composição dos dados que formarão nosso *dataframe* (DF) até a validação final do modelo. Para tanto, indicamos como elementos centrais no desenvolvimento da modelagem e pré-processamento de nosso DF a preparação de dados, analise do tipo de modelo × finalidade, configuração de treinamento, e estratégia de validação, os quais serão abordados em sequência.

Iniciando pelo elemento mais comumente chamado de pré-processamento, essa etapa abrange coleta criteriosa, limpeza, balanceamento, tratamento de valores ausentes, nor-

[13] O campo de densidade de energia, que integra a massa $M(r)$ na TOV é alterado por Λ_i.

malização e verificação de distribuição estatística dos dados. A qualidade dessa etapa determina o teto de desempenho de qualquer modelo subsequente.

Como segundo de nossa lista, é fundamental compreender se a tarefa é de regressão, classificação, clusterização ou outra categoria. O alinhamento claro entre o objetivo da aplicação, variáveis disponíveis e resultado esperado (relação DF $\rightleftarrows$ saída) orienta a escolha da técnica mais adequada.

A parte de configuração do treinamento, inclui seleção do algoritmo específico, ajuste de hiperparâmetros (por busca em grade, random search ou métodos bayesianos) e definição de políticas de regularização. Essas decisões devem refletir as características do conjunto de dados, categoria do modelo e as limitações operacionais do projeto.

Embora tenhamos colocado como último elementos, a estratégia de validação é tão importante quanto qualquer outro destaque anterior. A adoção de k-fold, validação temporal ou divisão treino-validação-teste (conforme o contexto), acompanhada de métricas compatíveis com o tipo de tarefa, permite avaliar de forma coerente a performance e comportamento do modelo durante seu processo de treinamento.

Com base nessas considerações, a seguir apresentamos os passos para a construção e o processamento de nosso dataframe final. Para manter a clareza das definições e resultados que serão discutidos, todas as referências à **densidade** serão expressas em unidades de $\mathbf{MeV\,fm^{-3}}$.

Decorrente do formalismo de PPG, o primeiro passo na construção de nosso algoritmo de geração de EdEs consiste em definir os parâmetros que asseguram as condições de continuidade e diferenciabilidade para cada uma das DIs, as quais conectam dois intervalos em ρ.

Posto isso, nas Tabelas 2.1 - 2.2 são apresentados[14] os parâmetros necessários para ajustar/conectar os segmentos em cada DI presente no intervalo $0{,}0 < \rho \leq \rho_0$, onde $\rho_0 = 1{,}15 \times 10^2\,\mathrm{MeV\,fm^{-3}}$, define o ponto[15] de separação entre as regiões de baixa e alta densidade em nosso modelo. Baseado em PPG, esse conjunto de parâmetros estabelece a equação de estado SLy4 como responsável por descrever[16] a região que antecede o intervalo paramétrico referente ao núcleo estelar (Douchin; Haensel, 2001). Os mesmos podem ser obtidos através das informações dispostas nas tabelas enumeradas como 2 e 3, do apêndice B da literatura utilizada como referência sobre EdE politrópicas (O'Boyle *et al.*, 2020).

[14] A notação de subíndices s_i, onde $i = 1, ..., 5$, é utilizada como forma de referenciar o respectivo número da linha na tabela onde os parâmetros são encontrados na literatura de PPG.

[15] Ainda nesse capitulo, mais a frente abordaremos as razões para a escolha desse valor.

[16] De acordo com os autores de PPG, os parâmetros definidos pelas Tabelas 2.1 - 2.2 também ajustam a equação de estado QHC19, para descrever o intervalo definido como região média da crosta (Baym *et al.*, 2019).

TABELA 2.1 – Valores das DIs para configuração da equação de estado SLy4 que definem os pontos de transição entre os segmentos.

i	ρ_{s_i} $(\mathbf{MeV\,fm^{-3}})$
0	$3{,}52 \times 10^{-7}$
1	$1{,}02 \times 10^{-4}$
2	$1{,}87 \times 10^{-1}$
3	$2{,}98 \times 10^{-1}$
4	$5{,}35 \times 10^{1}$
5	$4{,}15 \times 10^{2}$

TABELA 2.2 – Parâmetros que definem a EdE politrópica segmentada no formalismo PPG. Esse ajuste foi projetado para reproduzir com precisão o índice politrópico, bem como a pressão em baixas densidades para a SLy4. Cada linha corresponde a um dos intervalos de densidade definidos na Tabela 2.1, onde os coeficientes $(k_{s_i}, \Gamma_{s_i}, \Lambda_{s_i}, a_{s_i})$ são aplicáveis, e o último segmento é truncado em $\rho_0 = 1{,}15 \times 10^2\,\mathrm{MeV\,fm^{-3}}$.

i	$\rho_{s_{i-1}} < \rho \leq \rho_{s_i}$	$k_{s_i}\left[(\mathbf{MeV\,fm^{-3}})^{1-\Gamma_{s_i}}\right]$	Γ_{s_i}	$\Lambda_{s_i}(\mathbf{MeV\,fm^{-3}})$	a_{s_i}
0	$0 < \rho \leq \rho_{s_0}$	$1{,}59 \times 10^{-1}$	1,611	0	0
1	$\rho_{s_0} < \rho \leq \rho_{s_1}$	$1{,}40 \times 10^{-2}$	1,440	$-7{,}59 \times 10^{-13}$	$-1{,}861 \times 10^{-5}$
2	$\rho_{s_1} < \rho \leq \rho_{s_2}$	$3{,}28 \times 10^{-3}$	1,269	$-3{,}38 \times 10^{-9}$	$-5{,}278 \times 10^{-4}$
3	$\rho_{s_2} < \rho \leq \rho_{s_3}$	$-1{,}24 \times 10^{-5}$	−1,841	$6{,}69 \times 10^{-4}$	$1{,}035 \times 10^{-2}$
4	$\rho_{s_3} < \rho \leq \rho_{s_4}$	$8{,}35 \times 10^{-4}$	1,382	$3{,}97 \times 10^{-4}$	$8{,}208 \times 10^{-3}$
5	$\rho_{s_4} < \rho \leq \rho_0$	$5{,}05 \times 10^{-7}$	3,045	$\mathbf{1{,}12 \times 10^{-1}}$	$\mathbf{1{,}94 \times 10^{-2}}$

As informações complementares (valores em negrito na Tabela 2.2), que não foram calculados na versão de origem[17] das tabelas, estão relacionadas ao seguinte fato. Observando que a DI definida para iniciar a segmentação de nosso modelo parametrizado ρ_0 é menor que a densidade[18] ρ_{s_5}, utilizamos as equações de recorrência para obter os novos valores (que antes seriam denotados por Λ_{s_5} e α_{s_5}) dos parâmetros que compõem

[17] Na literatura utilizada como referência para o formalismo de PPG, tais valores não são fornecidos.

[18] Essa informação esta dispostas no item (6) da Seção IV.C da referência adotada sobre PPG (O'Boyle *et al.*, 2020).

parte do conjunto responsável pela conexão desejada em ρ_0. Definidos, então, como Λ_0 e α_0, os valores obtidos que estão dispostos em negrito na Tabela 2.2, juntos aos demais[19] vinculados aos outros seguimentos de menor densidade.

Com a Tabela 2.2 em mãos, configuram-se todas as informações necessárias para a descrição de nossa equação de estado da região de baixa densidade. A partir desse valor (ρ_0), inicia-se a região de alta densidade onde o modelo paramétrico com suas segmentações assume o controle de descrição da equação de estado. Observe que, possuindo por completo os parâmetros do formalismo de PPG até ρ_0, todos os demais parâmetros relativos aos seguimentos da parametrização que surgiram, são obtidos de forma recorrente definindo-se as seguintes DIs e respectivos Γ_i, diretamente através do conjunto de expressões formado pelas Eqs.(2.74 - 2.81).

Conforme idealizado, na Tabela 2.3 organizamos uma visão das DIs, regiões da estrutura estelar e correspondente equação de estado que descreve o segmento do modelo.

Em síntese, as regiões de densidade do nosso modelo são definidas de modo que, a SLy4 é válida em descrever a matéria estelar na região de densidade até ρ_0, enquanto nas regiões de densidade superior, emprega-se a parametrização baseada em uma subdivisão logaritmicamente espaçada em 3 partes, dado intervalo ρ_0 e $\rho_3 = 1{,}12 \times 10^3\,\mathrm{MeV\,fm^{-3}}$.

Optamos por posicionar as quebras de densidade definindo 3 segmentos (iniciados por ρ_0) com espaçamento uniforme no $\log\rho$, buscando melhorar a capacidade do formalismo em reproduzir as grandezas macroscópicas calculadas pelas equações de estrutura, notadamente massa e raio (Raithel; Ozel; Psaltis, 2017). Existem duas razões principais para essa escolha. A primeira é que a equação de estado varia por ordens de grandeza, de modo que dividir o domínio em fatores multiplicativos iguais garante uma distribuição da resolução proporcional à própria variação física. A segunda é que as quantidades macroscópicas M, R e o momento de inércia I apresentam maior sensibilidade às densidades próximas da densidade de saturação nuclear, $\rho_{\mathrm{sat}} \approx 1{,}51 \times 10^2\,\mathrm{MeV\,fm^{-3}}$. Dessa forma, o espaçamento adotado concentra, sem aumentar o número total de segmentos, mais pontos de quebra justamente na região em que a resposta de M, R e I é mais pronunciada (Raithel; Ozel; Psaltis, 2016).

Equivalente ao espaçamento uniforme em $\log\rho$, a escolha desses pontos em progressão geométrica aparece tanto nas politrópicas ajustadas em $\log P$-$\log\rho$ de (Read *et al.*, 2009), quanto na reconstrução por pressões em densidades fiduciais[20] (Ozel; Psaltis, 2009).

[19] Visto que até ρ_0 nossa equação de estado trata-se da SLy4, os valores de k_{s_5} e Γ_{s_5} devem ser mantidos os mesmo conforme descritos na literatura sobre PPG.

[20] Densidades pré-selecionadas em progressão geométrica nas quais a pressão é tratada como parâmetro livre

TABELA 2.3 – Segmentos de densidade, equações de estado empregadas e composição dominante, segundo o esquema unificado SLy4 para crosta e núcleo externo. Esse conjunto de informações fornece uma visão estrutural básica da EN, gerada através do formalismo em questão.

	Região - EdE	$\boldsymbol{\rho}\,(\mathbf{MeV\,fm^{-3}})$	Composição esperada
Crosta	Externa – SLy4	$\rho < 2{,}0 \times 10^{-1}$	Rede cristalina de núcleos pesados (Fe, etc.) imersos em gás de elétrons degenerados.
	Interna – SLy4	$2{,}0 \times 10^{-1} \leq \rho < 2{,}0$	Núcleos muito ricos em nêutrons imersos em gás de nêutrons livres e elétrons relativísticos (pós-gotejamento).
	Interna – SLy4	$2{,}0 \leq \rho < 6{,}0 \times 10^{1}$	Possíveis fases pasta (lâminas, tubos, bolhas), coexistindo com gás de nêutrons; meio inhomogêneo em transição rumo ao núcleo.
Núcleo	Externo – SLy4	$6{,}0 \times 10^{1} \leq \rho < 1{,}15 \times 10^{2}$	Matéria uniforme $npe(\mu)$ em equilíbrio β; fração pequena de prótons e elétrons; surgimento de múons quando $\mu_e \geq m_\mu c^2$.
Parametrizada	1° Segmento	$1{,}15{\times}10^{2} \leq \rho < 2{,}45{\times}10^{2}$	Continuação politrópica ajustada a observáveis astrofísicos.
	2° Segmento	$2{,}45{\times}10^{2} \leq \rho < 5{,}24{\times}10^{2}$	$\cdots$
	3° Segmento	$5{,}24{\times}10^{2} \leq \rho < 1{,}12{\times}10^{3}$	$\cdots$

Com efeito sobre o próximo passo da modelagem, traremos luz sobre uma das principais características que diferem o formalismo generalizado do padrão, recordando o desenvolvimento feito na seção dedicada às equações de estado politrópicas. Nela, junto das definições atreladas a um fluido barotrópico e a uma equação de estado de um gás ideal, a partir da 1° lei da termodinâmica pudemos (sem a necessidade de um *ansatz*) derivar a seguinte igualdade para o expoente politrópico:

$$\hat{\Gamma}_i \;=\; \frac{\rho}{p_i}\,\frac{dp_i}{d\rho}, \tag{2.90}$$

ao qual, por conveniência, passaremos a nos referir como **índice politrópico efetivo local**.

Olhando o lado direito dessa equação, conforme já verificado, os termos relacionados com a derivada da pressão (bem como $c_{s_i}^2$ no PPG) são livres[21] do termo Λ_i. Contudo, diferente do formalismo padrão, embora o termo extra não contribua com a resposta dinâmica da pressão ($dp_i/d\rho$), ele ainda altera seu nível absoluto. Fato que deve ser levado em conta ao se expressar o índice efetivo. Para tal, lembrando que $k_i\,\rho^{\Gamma_i} = p_i - \Lambda_i$, substituindo a forma generalizada da pressão na Eq.(2.90), obtém-se diretamente

$$\hat{\Gamma}_i \;=\; \Gamma_i\!\left(1 - \frac{\Lambda_i}{p_i}\right). \tag{2.91}$$

Essa nova relação evidencia como se medir a rigidez efetiva local do formalismo. Facilmente verificável, temos que no limite $\Lambda \to 0$, recupera-se a identidade $\hat{\Gamma} \to \Gamma$, do caso politrópico padrão.

Primeiramente, a Eq.(2.91) explicita que o sinal de Λ_i modula a rigidez efetiva de tal modo que, se $\Lambda_i > 0$, então $\hat{\Gamma}_i < \Gamma_i$ (amolecimento relativo); se $\Lambda_i < 0$, então $\hat{\Gamma}_i > \Gamma_i$ (endurecimento relativo). Contudo, para $\Lambda_i < 0$ é necessário a restrição física adicional de manter $p_i(\rho) > 0$ em todo o domínio de interesse. Próximo ao limite $p_i \to 0^+$, a razão Λ_i/p_i diverge e, por conseguinte, $\hat{\Gamma}_i$ cresce sem limite, indicando uma região não física que deve ser excluída da análise. Assim, além das checagens de monotonicidade ($dp/d\rho > 0$), impõe-se explicitamente a positividade da pressão ao longo do domínio integrado nas equações de TOV.

A expressão encontrada para o índice politrópico efetivo é de nosso interesse, pois estabelece uma métrica que traduz e permite a comparação direta com os valores dos limites impostos sobre cada Γ do formalismo padrão, os quais regulam a regidez da equação de estado. Para ilustrar o impacto de Λ sobre $\hat{\Gamma}$, consideramos as Figs.(2.2, 2.3, 2.4).

Fixando, para a análise ilustrativa, $k = 2$ e $\Gamma = 3$, a Fig.(2.2) mostra que curvas com $\Lambda > 0$ permanecem abaixo do valor Γ nominal[22] (linha tracejada em preto) até que $k\,\rho^{\Gamma} \gg \Lambda$, quando então $\hat{\Gamma} \to \Gamma$. Para $\Lambda < 0$, o comportamento é dual, a curva correspondente permanece acima de Γ, refletindo o endurecimento efetivo, mas somente é traçada na faixa em que $p(\rho) > 0$ (regiões não físicas são omitidas).

Na Fig.(2.3), mantendo ρ fixo, observa-se a dependência de $\hat{\Gamma}$ com Λ. Para $\Lambda > 0$, $\hat{\Gamma}$ decresce monotonicamente e tende a zero quando $\Lambda \to \infty$. Para $\Lambda < 0$, $\hat{\Gamma}$ cresce acima de Γ e diverge ao se aproximar do limite $\Lambda \to -\,k\,\rho^{\Gamma}$ pelo lado direito (fronteira $p \to 0^+$); por isso, a vizinhança proibida é omitida no gráfico.

[21] Voltando à forma como se descrevem a pressão e a densidade de energia nas Eqs. (2.73, 2.80), note que o termo Λ_i desloca ambos com sinais opostos, preservando o termo $\epsilon_i + p_i$ presente na expressão da velocidade do som.

[22] Valor fixo de Γ parametrizado para todo o intervalo de densidade, correspondente a descrição do formalismo padrão.

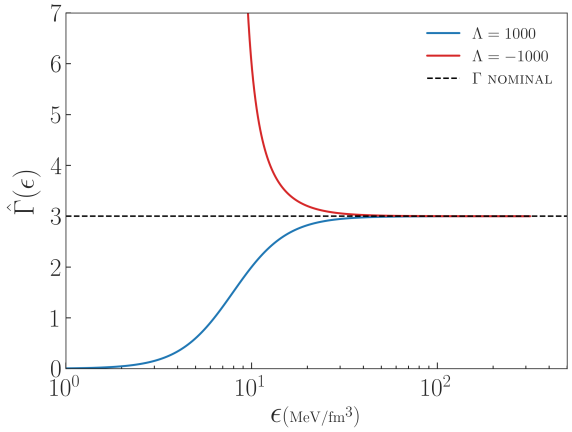


FIGURA 2.2 – Índice efetivo $\hat{\Gamma}(\epsilon)$ para dois sinais de Λ. Curvas com $\Lambda > 0$ ficam abaixo de Γ; com $\Lambda < 0$ ficam acima, traçadas apenas onde $p > 0$.

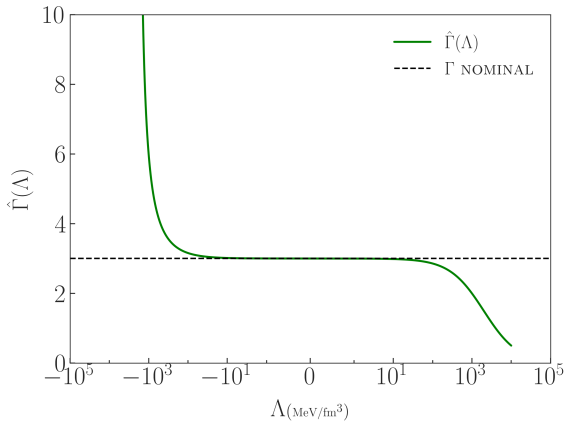


FIGURA 2.3 – Dependência $\hat{\Gamma}(\Lambda)$ em ϵ fixa. Para $\Lambda > 0$, $\hat{\Gamma} \downarrow 0$; para $\Lambda < 0$, $\hat{\Gamma} > \Gamma$ e diverge quando $\Lambda \to -\,k\,\epsilon^{\Gamma}$ $(p \to 0^{+})$.

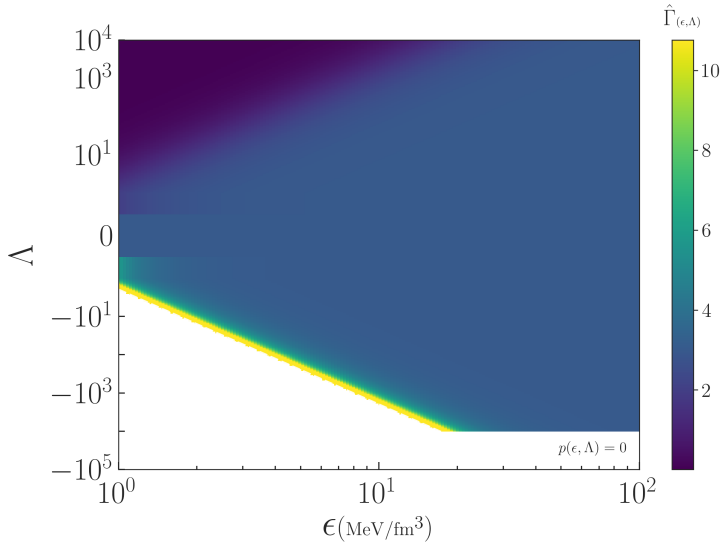


FIGURA 2.4 – Mapa de calor para $\hat{\Gamma}(\epsilon, \Lambda)$. A linha $p = 0$ $(\Lambda = -k\,\epsilon^{\Gamma})$ é indicada; a região não física $(p \leq 0)$ é mascarada.

Finalmente, para visualizar simultaneamente a atuação conjunta de ϵ e Λ, a Fig.(2.4) apresenta um mapa de calor de $\hat{\Gamma}(\epsilon, \Lambda)$ cobrindo $\Lambda > 0$ e $\Lambda < 0$. Observa-se que o impacto de Λ é mais expressivo em baixas densidades. Valores positivos reduzem a rigidez efetiva, enquanto valores negativos a elevam, limitados pela fronteira física $p = 0$. Em regiões de alta densidade, o termo barotrópico $k\,\epsilon^{\Gamma}$ domina e restaura a equivalência com o formalismo padrão, independentemente do sinal de Λ.

Com isso, tendo em vista analisar posteriormente se tais definições foram ou não justificadas, *a priori*, faremos a adoção dos limites inferiores e superiores para os índices Γ_i de cada DI, tendo como objetivo equilibrar: (i) fidelidade a cálculos nucleares em

baixas densidades, (ii) liberdade para rigidez intermediária sem violar a causalidade, e (iii) compatibilidade com massas e raios observados em ENs próximas a $2M_\odot$. Com esse enquadramento, os intervalos são motivados por evidências empíricas e vínculos físicos:

- **Cobertura empírica da crosta e região de saturação:** Read et al. demonstraram que três segmentos com $1 \lesssim \Gamma \lesssim 4$ reproduzem ~ 30 EdEs nucleares com erro inferior a 4% (Read *et al.*, 2009). O refinamento posterior de Greif *et al.* indica que estender o teto de Γ_1 até 4,5 amplia a diversidade de rigidez sem degradar o ajuste (Greif *et al.*, 2019).

- **Flexibilidade sob o limite causal:** Hebeler *et al.* propuseram priors tão amplos quanto $\Gamma \simeq 8$, mas mostraram que valores extremos entram rapidamente em conflito com $c_s^2 \leq 1$ (Hebeler *et al.*, 2013). A faixa $\Gamma_2 \in [1{,}5,\ 5{,}0]$ segue o pico estatístico obtido em amostras bayesianas recentes, que privilegiam $2 \lesssim \Gamma_2 \lesssim 4$ e minimizam soluções superluminais (Greif *et al.*, 2019).

- **Consistência com limites observacionais:** Estudos como os de Raithel & Most e Özel & Freire compilaram massas $\approx 2M_\odot$ e raios $R \approx$ [11 - 13] km, verificaram que $\Gamma \gtrsim 5$ raramente satisfaz simultaneamente observações de massa e raio (Raithel; Most, 2023; Ozel; Freire, 2016).

Para os três segmentos da região de alta densidade, definidos os intervalos

$$\Gamma_1 \in [1{,}3,\ 4{,}5], \quad \Gamma_2 \in [1{,}5,\ 5{,}0], \quad \Gamma_3 \in [2{,}0,\ 5{,}0], \tag{2.92}$$

de modo a cobrir, com parcimônia, desde regimes macios até rígidos para as equações de estado.

Dando sequência, sabemos que o princípio da causalidade é um requisito físico básico, o qual estabelece que a velocidade máxima de propagação de uma onda em qualquer material é a velocidade da luz no vácuo. Em nosso contexto com as ENs, para garantir que as equações de estado respeitem esse princípio, a velocidade do som não pode exceder tal limite em todas as pressões e densidades possíveis. Para tanto, durante a integração das equações de TOV, exigimos localmente através da Eq.(2.86), a seguinte condição normalizada,

$$c_{s_i}^2 \;=\; \frac{\Gamma_i\,(p_i - \Lambda_i)}{\epsilon_i + p_i} \;\leqslant\; 1, \tag{2.93}$$

que define para cada conjunto $(\Gamma_1, \Gamma_2, \Gamma_3)$, o limite superior $\rho_{\max}$ de exploração de densidade central na malha TOV, assegurando que todas as soluções obtidas em nosso modelo o respeitem.

Além das considerações associadas à causalidade e limites de Γ, é fundamental reconhecer o papel decisivo que estudos observacionais exercem na validação de soluções propostas por modelos teóricos. Com base nisso, estabelecemos uma condição adicional que filtra as soluções obtidas a partir de pesquisas astrofísicas recentes de relevância (Abbott *et al.*, 2017; Miller *et al.*, 2019; Riley *et al.*, 2019; Miller *et al.*, 2021; Riley *et al.*, 2021).

Essa escolha impacta diretamente a composição de nosso futuro banco de dados, pois o critério seleciona apenas as EdEs geradas que se mantêm consistentes com as informações fornecidas por tais trabalhos. Em particular, utilizamos as regiões delimitadas por vínculos observacionais provenientes das medições realizadas pelo NICER para os valores de massa e raio dos pulsares PSR J0030+0451 e PSR J0740+6620, bem como as restrições derivadas da análise das ondas gravitacionais associadas ao evento GW170817.

O conjunto dessas informações define uma faixa plausível para a relação de massa e raio das ENs. Impomos, portanto, que qualquer solução das equações de TOV, juntamente com a respectiva EdE presente em nossa base, deva necessariamente intersectar, em pelo menos um ponto, cada uma das regiões estabelecidas no diagrama M-R segundo os estudos citados.

Levando em conta a base teórica e as restrições mencionadas, utilizando a linguagem Python, desenvolvemos um código, o qual encontra-se disponível através do link [Repositório COD], que implementa nosso modelo paramétrico para geração dos dados em larga escala.

Para a execução desse código, é necessário apenas que defina-se o número de conjuntos/trios Γ_i a serem amostrados, no argumento `n` da função principal `main_execution`.

O procedimento completo da execução do código segue quatro passos bem objetivos:

1. **Amostragem dos parâmetros**
   São randomicamente selecionados n-conjuntos de índices politrópicos a partir de distribuições uniformes, com amostragem independente em cada intervalo previamente estabelecido na Eq.(2.92), de modo que,
   $$\Gamma_i \sim \mathcal{U}\big(\Gamma_i^{\min}, \Gamma_i^{\max}\big), \quad i = 1, 2, 3.$$
2. **Construção das constante e monotonicidade**
   Para cada trio amostrado de Γ_i, calcula-se $(k_i,\ \Lambda_i,\ \alpha_i)$ para as DIs assegurando a continuidade de p, ϵ e $dp/d\rho$, seguido pela verificação de positividade em todo o intervalo de variação $(p > 0)$ e $(p_1 < p_2 < p_3)$ que valida o conjunto para seguir.
3. **Causalidade**
   Avalia-se a condição $c_s^2 \leq 1$ em um leque de densidades centrais; apenas os conjuntos que obedecem a esta restrição recebem um limite superior $\rho_{\max}$ e seguem adiante.

4. **Integração das equações de TOV**
   Para cada EdE aceita, resolve-se o sistema dp/dr e dm/dr até a pressão superficial $p(r) = 0$, devolvendo massa M e raio R.

Dessa forma, escolhendo o valor de n, o código gera um conjunto diversificado de EdEs fisicamente aceitáveis e calcula de forma automática as respectivas soluções M-R das estrelas. O resultante é um banco de n-EdEs fisicamente válidas com suas soluções TOV correspondentes.

Realizado testes iniciais para aproximar a taxa que estima a quantidade de equações de estado resultantes pós-filtro observacional, definimos uma quantidade total inicial de $2{,}3 \times 10^5$ EdEs para o processamento, as quais resultaram em $\sim 9{,}7 \times 10^4$ curvas viáveis de acordo com o filtro observacional, as quais são exibidas na Fig.(2.5) seguida por suas respectivas soluções M-R na Fig.(2.6).

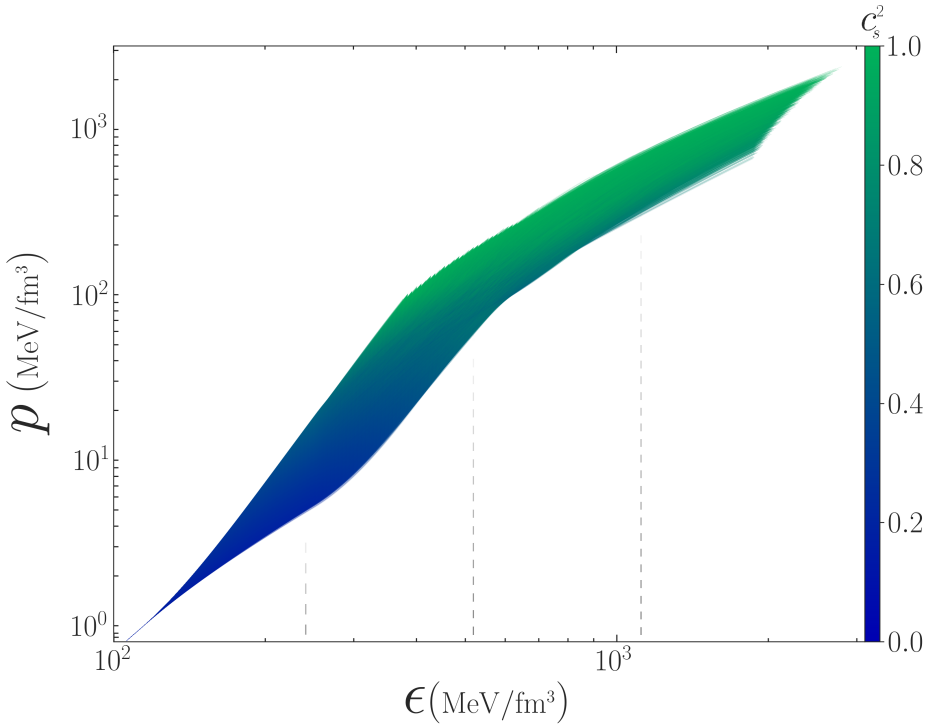


FIGURA 2.5 – Conjunto de EdEs geradas segundo o formalismo PPG, conforme Eq.(2.73), com segmentações contínuas em pressão, energia e derivadas. A região subnuclear é descrita pela SLy4, enquanto a de alta densidade é parametrizada por três segmentos politrópicos com índices Γ_i variando nos intervalos $\Gamma_1 \in [1{,}3, 4{,}5]$, $\Gamma_2 \in [1{,}5, 5{,}0]$ e $\Gamma_3 \in [2{,}0, 5{,}0]$, balanceando rigidez física, causalidade e vínculos observacionais. A coloração presente nas curvas $p(\epsilon)$ correspondem ao valor normalizado de c_s^2, os quais além de indicar o comportamento de rigidez do fluído, evidência a causalidade exigida sob o modelo. As linhas tracejadas verticais indicam as DIs entre segmentos, ao longo das quais são garantidas transições suaves. O conjunto final, composto por $9{,}77 \times 10^4$ EdEs viáveis, é consistente com os dados observacionais de 5 recentes estudo sobre a determinação da M-R de pulsares massivos.

Temos representado as curvas $p(\epsilon)$ como as EdEs no gráfico, onde o gradiente de cores faz alusão aos valores associados à velocidade do som, trazendo uma perspectiva sobre como a rigidez do fluido estelar varia. As linhas tracejadas marcam as DIs que conectam as respectivas partições, também onde evidenciam-se as transições suaves que o formalismo PPG fornece.

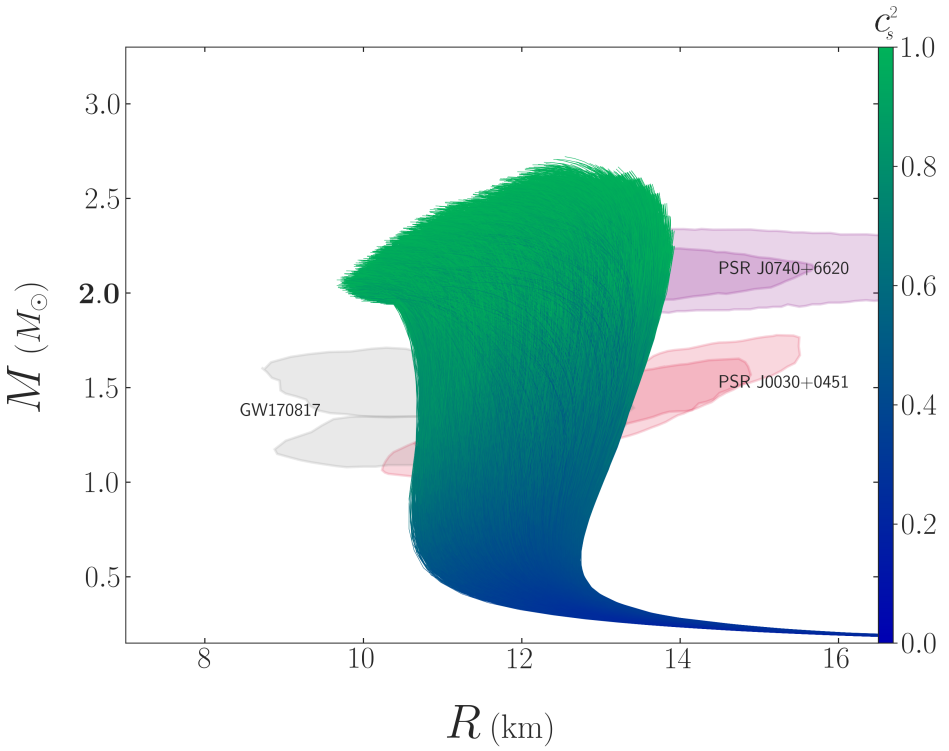


FIGURA 2.6 – Soluções M-R das equações de TOV para o conjunto de EdEs representadas na Fig.(2.5). Cada curva corresponde à solução gerada por uma EdE específica, coloridas de acordo com o valor normalizado de c_s^2, avaliado no centro estelar correspondente, refletindo a rigidez máxima do par M-R no diagrama sob o critério causal. As 6 regiões sombreadas representam os domínios observacionais derivados de estudos recentes: o evento de ondas gravitacionais GW170817 (Abbott *et al.*, 2017) e as medições do NICER para os pulsares PSR J0030+0451 (Riley *et al.*, 2019; Miller *et al.*, 2019) e PSR J0740+6620 (Riley *et al.*, 2021; Miller *et al.*, 2021). O conjunto de soluções mostra intersecções consistentes com todas essas regiões, assegurando que apenas configurações compatíveis com os vínculos observacionais foram consideradas em nossa base final de dados.

Após termos consolidado o conjunto de curvas que será utilizado nas seguintes fases do trabalho, como ultima análise desta seção, avaliamos $\hat{\Gamma}_i$ definidos sobre as DIs, ou seja, o índice exponencial efetivo dos três segmentos parametrizados. Podemos ver os resultados na seguinte Tabela.2.4:

TABELA 2.4 – Valores verificados como limites máximos e mínimos, e estatísticas de média e desvio padrão de $\hat{\Gamma}_i$ avaliadas em cada DI.

	min	max	média	σ
$\hat{\Gamma}_1$	1,869	4,351	3,432	0,554
$\hat{\Gamma}_2$	1,863	4,977	3,654	0,796
$\hat{\Gamma}_3$	2,045	4,995	3,509	0,769

Como verificação complementar, avaliamos $\hat{\Gamma}_1$ em densidades fiduciais $\rho_{\rm sat}$, $2\rho_{\rm sat}$ e $3\rho_{\rm sat}$, agregando todas as equações de estado aprovadas:

TABELA 2.5 – Diagnóstico de Γ_1 em densidades fiduciais (amostra final).

ρ	min	max	média
$\rho_{\rm sat}$	2,170	3,803	3,180
$2\rho_{\rm sat}$	2,643	4,696	3,539
$3\rho_{\rm sat}$	1,967	4,956	3,646

Esses diagnósticos sustentam que os intervalos definidos na Eq.(2.92) são amplos o bastante para abarcar a rigidez efetiva requerida pelas soluções compatíveis com causalidade e observações, sem evidência de truncamento indevido do espaço físico relevante: (i) os valores máximos de $\hat{\Gamma}_i$ alcançam ~ 5 nos segmentos mais densos, e (ii) as médias permanecem na faixa $\sim$ 3,2 - 3,7 em $\rho \in [\rho_{\rm sat},\, 3\rho_{\rm sat}]$, coerentes com massas $\sim 2M_\odot$ e raios $\sim$ 11-13 km obtidos no conjunto filtrado.

2.5.1 Pré-Processamento

Mais comumente conhecido como *Dataprep*, a preparação dos dados deve ser cuidadosamente adaptada às correlações existentes entre os tipos de dados envolvidos, à técnica aplicada, à estrutura e ao propósito final do modelo.

No próximo passo, com as EdEs $\rightleftarrows$ M-R previamente geradas, conduziremos esse fluxo avançando para a etapa de consolidação dessas informações, preparando o conjunto de dados que servirá de base para o treinamento de nosso futuro modelo de RN.

Este trabalho tem como objetivo construir um modelo que assegure sua viabilidade prática e aplicabilidade em contextos observacionais reais (Ozel; Psaltis, 2009). Para isso,

com base em suposições plausíveis, considera-se um cenário no qual são coletados dados de (M, R) referentes a 6 estrelas de nêutrons, os quais servirão como entradas do modelo.

Neste contexto, o problema é formulado como uma tarefa de regressão, na qual, dados os valores de massa e raio desses seis objetos, busca-se estimar os parâmetros da equação de estado que melhor reproduzem essa configuração observacional.

Com o cenário idealizado, definimos que, para cada EdE de nossa base, são amostrados n subconjuntos distintos, cada um contendo 6 pontos extraídos de sua respectiva curva de M-R. A seleção dos pontos ao longo de cada curva de M-R é realizada aleatoriamente, respeitando as regiões observacionais, que delimitam os domínios fisicamente plausíveis de massa e raio.

Esse processo de reamostragem busca através da diversidade, promover maior capacidade de generalização a RN, sem perder a robustez aliada aos limites observacionais dos estudos já citados.

Uma vez definidos os critérios anteriores, com o valor de n-conjuntos = 50, obteve-se um total aproximado de $4{,}88 \times 10^6$ linhas para o DF, que é composto por 15 colunas, sendo as primeiras 12 destinadas aos valores de massa e raio amostrados e as últimas 3 os parâmetros Γ da respectiva EdE.

É preciso ressaltar que, por conta do processo de amostragem aleatória dos pontos de (M, R), foi benéfico estabelecer uma diretriz de ordenação sobre os mesmos. Portanto, adotamos como critério a verificação da permutação dos pares (M, R) que minimiza comprimento de curva seguido da orientação inicial pelo ponto com menor valor de massa.

Após a consolidação inicial do DF, os dados passam por uma padronização de tipo numérico, sendo convertidos para `float32`, garantindo maior eficiência computacional nas etapas posteriores.

Em seguida é feita a divisão dos dados em três subconjuntos mutuamente exclusivos: conjunto de treinamento (60% dos dados), conjunto de validação (20%) e conjunto de teste (20%). Essa divisão permite avaliar o desempenho do modelo durante o ajuste e verificar sua capacidade de generalização em dados não vistos.

Com essa separação, aplica-se uma instancia de ruído aleatório ao conjunto de treinamento, sob cada par ordenado (M, R), com variâncias $\sigma_M^2 = 0{,}01$ e $\sigma_R^2 = 0{,}03$, a fim de simular incertezas de natureza observacional típicas das medições astrofísicas de massa e raio, incorporando assim os erros experimentais inerentes ao processo de aquisição de dados. Por conseguinte, o conjunto de treinamento original e sua versão ruidosa são então unificados, resultando em um aumento da quantidade de amostras de treino, seguido de um embaralhamento global das amostras para evitar ordenações artificiais.

Para viabilizar o treinamento adequado da rede neural, as variáveis de entrada (massa

e raio) e os parâmetros-alvo (índices politrópicos Γ) são normalizados[23] individualmente através de transformações lineares, mapeando seus respectivos domínios para o intervalo $[0, 1]$.

Sendo assim, para cada variável x, o procedimento de normalização é definido por:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{2.94}$$

em que $x_{\min}$ e $x_{\max}$ correspondem, respectivamente, aos valores mínimo e máximo observados da variável x no conjunto de dados.

O *pipeline* computacional desenvolvido para preparar os dados realiza sumariamente as seguintes etapas:

1. Amostragem de n subconjuntos com 6 pares (M, R), extraídos aleatoriamente de cada curva, respeitando as regiões observacionais.

2. Aplicação do critério de ordenação dos pares (M, R) por minimização do comprimento de curva, com orientação fixada pelo menor valor de massa.

3. Consolidação do DF contendo 15 colunas, das quais 12 armazenam os pares amostrados de (M, R) e 3 correspondem aos parâmetros Γ, seguida da conversão dos dados para o tipo numérico `float32`.

4. Divisão dos dados consolidados em treinamento (60%), validação (20%) e teste (20%).

5. Inserção de ruído gaussiano no conjunto de treinamento, com variâncias $\sigma_M^2 = 0{,}01$ e $\sigma_R^2 = 0{,}03$, simulando incertezas observacionais em massa e raio.

6. Unificação do conjunto original de treinamento com sua versão ruidosa, seguida de embaralhamento global das amostras.

7. Normalização das variáveis de entrada (massa e raio) e dos parâmetros-alvo (Γ) via transformação linear para o intervalo $[0, 1]$.

[23] Os normalizadores ajustados são armazenados em arquivos externos, possibilitando a reprodutibilidade da normalização nas fases de treinamento e aplicação do modelo.

3 Redes Neurais

Cabe-nos, ao início deste capítulo, apresentar uma breve revisão histórica sobre o surgimento das Redes Neurais artificiais seguido dos principais aspectos das redes do tipo *Feedforward* (Rosenblatt, 1958; Rumelhart; Hinton; Williams, 1986; Bishop, 1995). Essa abordagem visa introduzir além dos conceitos básicos sobre a temática, elementos que darão suporte à compreensão do futuro paralelo a ser explorado.

Nesse contexto, a comparação que será realizada possui como foco evidenciar a natureza das particularidades do modelo de RN probabilístico que posteriormente será implementado. Por último, abordaremos as especificidades topológicas do modelo e seus respectivos resultados acerca do treinamento.

3.1 O Nascimento das Redes Neurais Artificiais

Entre as diversas vertentes do campo de AM, a Inteligência Artificial (IA) consolidou-se como uma subárea sendo um dos temas mais "quentes"da atualidade. Algo que há pouco pertencia somente aos cenários de ficção, hoje, com uma vasta gama de aplicações, trata-se de uma realidade cotidiana[1] do ser humano.

Figurando como um dos principais alicerces desse domínio, as IAs são compostas por RNs profundas inspiradas pela constituição biológica do cérebro. Embora o exponencial crescimento de poder computacional direcionado a essas redes tenha sido visto mais recentemente, os sucessivos avanços teóricos dessa temática remontam à década de 1940 (Mcculloch; Pitts, 1943).

Dentre os trabalhos pioneiros até a sua consolidação, um dos marcos iniciais foi a proposta de Warren McCulloch e Walter Pitts, que apresentaram o primeiro modelo artificial de um neurônio biológico. Eles demonstraram que, com um número suficiente dessas unidades conectadas operando de forma sincronizada, seria possível representar qualquer função computável. Todavia, como não foi desenvolvido um mecanismo para

[1] Mesmo que operem de forma invisível aos usuários, sistemas de recomendação, mecanismos de análise de crédito e ferramentas de reconhecimento de imagem constituem aplicações amplamente difundidas na atual sociedade.

ajustar os pesos das conexões, esse modelo não dispunha de um processo de treinamento e otimização.

Esse cenário recebeu uma importante contribuição do neuropsicólogo Donald Hebb, que, em seu livro *The Organization of Behavior*, propôs o conceito de aprendizado em RNs biológicas. Segundo sua teoria, as conexões sinápticas se fortalecem quando dois neurônios são ativados simultaneamente, caracterizando um mecanismo fundamental para o processo de aprendizado (Hebb, 1949).

Na década seguinte, influenciado pelo trabalho de Hebb, Frank Rosenblatt ofereceu outra importante contribuição às pesquisas da área ao propor uma RN orientada computacionalmente, empregando um método inovador de aprendizagem supervisionada. Representado na Fig.(3.1), denominado como *Perceptron*, o mesmo consistia em um modelo cognitivo de unidades sensoriais/entradas conectadas a uma única camada de neurônios acrescidos de sinapses ajustáveis.

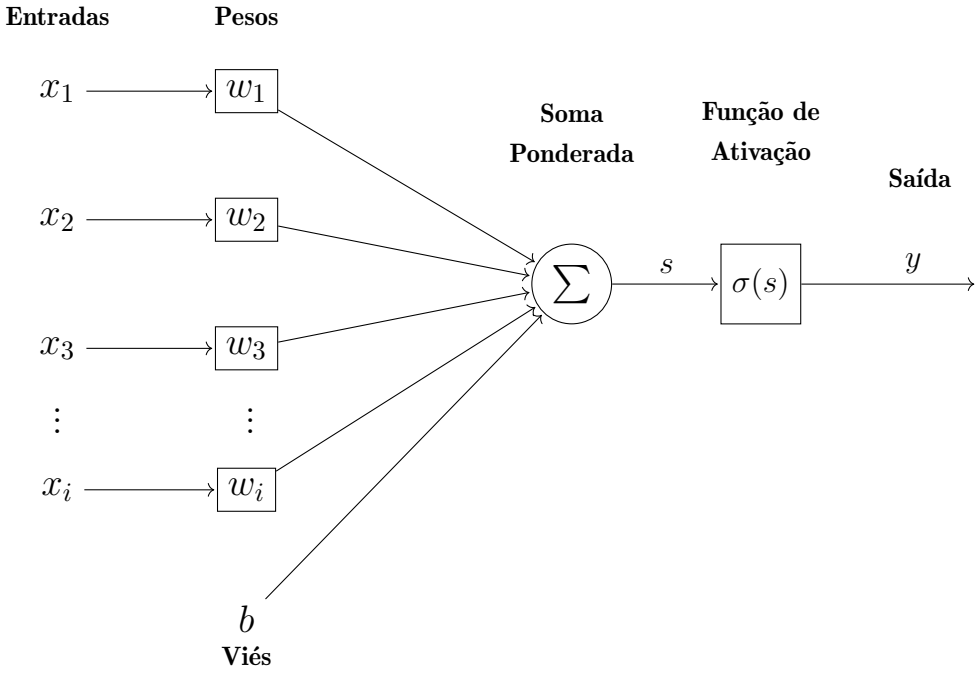


FIGURA 3.1 – Modelo Perceptron: múltiplas entradas são ponderadas por seus respectivos pesos e somadas ao viés, produzindo um valor intermediário. Este valor é então processado pela função de ativação, gerando a saída final do modelo.

Com essa nova estrutura, o modelo adquiriu a capacidade de identificar e aprender padrões a partir dos dados fornecidos[2], possibilitando, assim, a atribuição de classificações a novas amostras (Rosenblatt, 1958). Essa propriedade, posteriormente generalizada em

[2]Limitado a padrões linearmente separáveis.

modelos mais complexos, constitui a base da maioria das aplicações contemporâneas de RNs.

Na Fig.(3.1) ilustra-se a arquitetura básica de um *Perceptron*. O vetor de características $(x_1, x_2, \ldots, x_i)$ é ponderado por pesos $(w_1, w_2, \ldots, w_i)$. Cada w_j regula a influência da respectiva entrada na decisão, e o viés b desloca o limiar da ativação, ajustando a fronteira de decisão. As contribuições são agregadas em $s = \sum_{j=1}^{i} w_j x_j + b$, o qual é processado pela função de ativação $\sigma(s)$, que é capaz de introduzir não linearidade ao modelo. A saída $y = \sigma(s)$ representa a predição do perceptron.

Embora novas contribuições tenham sido feitas (von der Malsburg, 1973; Willshaw; von der Malsburg, 1976), no período que se seguiu, uma série de limitações de cunho teórico e tecnológico levaram as RNs artificiais a passarem por um momento de declínio.

Em particular, assinalando a abertura dessa temporada, Marvin L. Minsky e Seymour Papert chamaram a atenção para uma série de problemas em seu livro *Perceptrons*, demonstrando que tais modelos de camada única são incapazes de aprender funções não lineares limitando severamente seu poder de representação (Minsky; Papert, 1969).

Foi apenas na década de 1980 que as RNs voltaram a ganhar popularidade. Esse movimento teve entre suas obras precursoras o trabalho de John Hopfield, que mostrou como essas redes podem armazenar e recuperar padrões de forma associativa, impulsionando o interesse renovado no campo, especialmente nas redes recorrentes (Hopfield, 1982). Diversos outros avanços fundamentais surgiram nessa época, incluindo a popularização do algoritmo de *backpropagation*, redescoberto e difundido por David E. Rumelhart, Geoffrey E. Hinton e Ronald J. Williams (Goodfellow; Bengio; Courville, 2016).

Desse período fértil, importa salientar que, em reconhecimento à contribuição decisiva dessas ideias para o progresso das RNs e, por consequência, do AM, Hopfield e Hinton foram agraciados com o Nobel de Física de 2024.

Representado na Fig.(3.2), modelos como o Perceptron Multicamadas já empregavam regras de atualização de pesos, porém, o algoritmo de *backpropagation* introduziu um procedimento sistemático para retropropagar o erro por várias camadas.

Isso permitiu que redes com arquiteturas maiores aprendessem de forma mais eficiente, refletindo diretamente em modelos de desempenho superior (Rumelhart; Hinton; Williams, 1986). Nesse ambiente favorável, diante de sua eficácia em solucionar com êxito uma variedade de problemas, as RNs persistiram e ampliaram seu escopo de aplicação, sendo utilizadas em diversas tarefas como, por exemplo, reconhecimento de fala e imagens, recomendação de produtos e serviços, análise de risco, robótica avançada e condução autônoma.

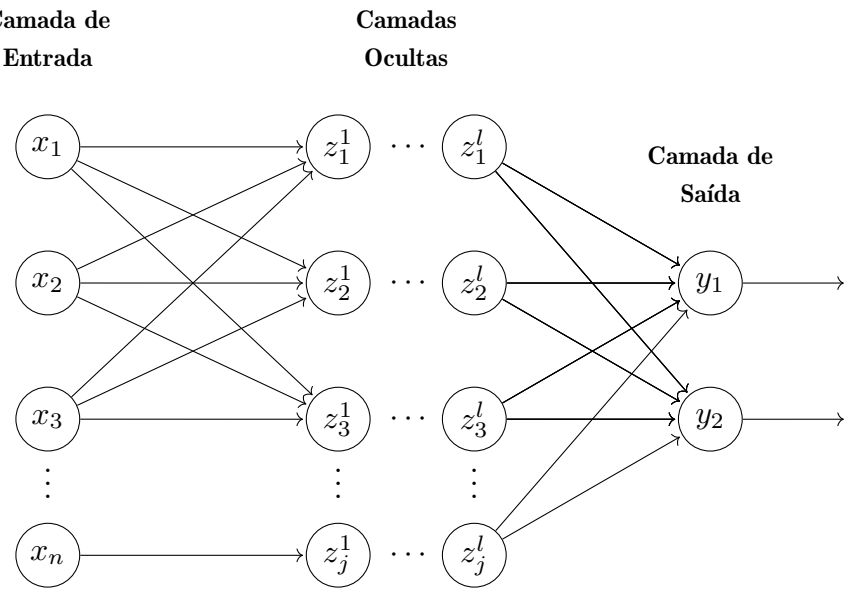


FIGURA 3.2 – Modelo Perceptron Multicamadas: composto por camadas de entrada, ocultas e de saída. As variáveis de entrada são processadas sucessivamente através de combinações ponderadas e funções de ativação, permitindo ao modelo capturar relações não lineares. A presença de múltiplas camadas ocultas amplia sua capacidade de representação, gerando as predições finais na camada de saída.

A Fig.(3.2) ilustra a estrutura de um Perceptron Multicamadas, constituído por camadas de entrada, ocultas e de saída. Na camada de entrada, cada componente x_j representa uma variável do vetor de entrada $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, a qual é propagada para a primeira camada oculta como $z_j^0 = x_j$. Nas camadas ocultas, cada neurônio i da camada l ($l \geq 1$) realiza uma combinação linear das saídas da camada anterior, computando a soma ponderada $s_i^l = \sum_j w_{ij}^l z_j^{l-1} + b_i^l$, seguida pela aplicação da função de ativação, produzindo $z_i^l = \sigma(s_i^l)$. Esse procedimento é repetido em todas as camadas ocultas subsequentes, de acordo com a regra geral $z_i^l = \sigma(\sum_j w_{ij}^l z_j^{l-1} + b_i^l)$. Por fim, na camada de saída, cada neurônio k computa sua saída a partir das ativações da última camada oculta (denotada por L), segundo $y_k = \sigma(\sum_j w_{kj}^{L+1} z_j^L + b_k^{L+1})$.

A popularidade crescente das RNs deve-se especialmente à sua excepcional aptidão em processar grandes volumes de dados e aprender representações altamente complexas. Um exemplo nítido desse potencial, o qual já integra a realidade de muitas pessoas, são as IAs generativas tais como o ChatGPT, Gemini, Llama e DeepSeek, também conhecidos como modelos de linguagem de grande escala.

Os famosos *Large Language Models* utilizam a arquitetura *transformer*, um avanço significativo no campo do *Deep Learning* (Aprendizagem Profunda), permitindo que o processamento e a geração de texto sejam realizados de forma extremamente contextualizada e coerente (Vaswani *et al.*, 2017).

3.1.1 Componentes e Topologias

Visto esse panorama promissor, antes de iniciarmos as discussões a respeito das bases conceituais e técnicas vinculadas à estruturação de nosso modelo bayesiano, é necessário, inicialmente, abordarmos algumas definições mais gerais deste contexto.

Contudo, apesar da ampla diversidade de possibilidades existentes no campo, notadamente, os elementos apresentados nesta seção têm como finalidade estabelecer um conhecimento básico e, dentre eles, os essenciais para desenvolvimento do modelo que será proposto mais a frente. Essa delimitação visa garantir clareza e consistência no desenvolvimento teórico que sustenta nossa abordagem. Para tal, elencamos:

Neurônios: Também denominados nós, são unidades fundamentais de processamento que recebem um ou mais *inputs* (dados de entrada) e produzem um *output* (dado de saída). Cada neurônio realiza uma soma ponderada das entradas recebidas, adiciona[3] um termo de viés e, em seguida, aplica uma função de ativação ao resultado para gerar sua saída.

Pesos: Valores numéricos associados a cada conexão de entrada, definindo a contribuição relativa de cada sinal de entrada na soma ponderada efetuada pelo neurônio. Esses parâmetros são ajustados durante o treinamento da rede pelo algoritmo de aprendizado.

Viés: Corresponde a um termo adicional associado a cada neurônio, cuja função é ajustar o limiar de ativação. Tecnicamente, o viés é somado ao valor da soma ponderada antes da aplicação da função de ativação, permitindo deslocar o limiar de ativação e, assim, ampliar a capacidade de ajuste do modelo aos dados.

Função de ativação: Recebendo como argumento a soma ponderada acrescida do viés, define a saída do neurônio. Pode ser linear ou não-linear, sendo estas últimas essenciais para permitir a modelagem de relações complexas. Exemplos comuns incluem a função sigmoid, ReLU (Unidade Linear Retificada) e *tanh* (tangente hiperbólica).

Função de perda: Quantifica o erro entre a saída prevista pela rede e o valor esperado. Ela sintetiza o desempenho do modelo durante o treinamento e orienta o ajuste dos parâmetros. A escolha da função de perda depende do tipo de problema e dos dados considerados. Entre as mais comuns estão o Erro Quadrático Médio, a Entropia Cruzada e o Erro Absoluto Médio.

Em termos gerais, a topologia de uma rede neural refere-se à sua configuração estrutural, ou seja, à maneira como os neurônios estão organizados em camadas e ao padrão de conexões que define o fluxo de informações ao longo da rede. Diferentes topologias determinam distintas formas de processamento e capacidades de representação. Entre as principais formas de arquiteturas, destacam-se:

[3] Quando previsto na arquitetura em questão.

Feedforward: configuram o modelo mais elementar e amplamente utilizado, no qual o fluxo de informação é estritamente unidirecional, propagando-se sequencialmente da camada de entrada até a camada de saída, sem a existência de ciclos ou realimentações. Essa organização favorece estabilidade computacional e é largamente aplicada em tarefas de classificação e regressão.

Recorrentes: caracterizam-se pela presença de conexões retroativas que permitem o retorno de informações a camadas anteriores, criando ciclos internos na rede. Essa arquitetura confere à rede a capacidade de modelar dependências temporais e sequências de dados, sendo particularmente adequada para processamento de séries temporais, linguagem natural e sinais dinâmicos.

Convolucionais: estruturadas por meio de conexões locais e compartilhamento de pesos, reduzem significativamente o número de parâmetros e exploram a estrutura espacial dos dados de entrada. São especialmente eficazes no processamento de imagens, reconhecimento de padrões visuais e, mais recentemente, têm sido adaptadas a outras modalidades de dados estruturados.

Dentre essas variantes, a feedforward será a que adotaremos na construção de nosso modelo. Nesse tipo de rede, as camadas são, em geral, densamente conectadas, no sentido de que cada neurônio de uma camada estabelece conexões com todos os outros da camada seguinte. Os neurônios são organizados em três tipos de camadas principais, classificadas de acordo com sua posição e função na arquitetura:

Camada de entrada: Responsável por receber os dados iniciais e repassá-los para o processamento subsequente na rede. É composta por neurônios que correspondem diretamente aos atributos do conjunto de dados, de modo que cada componente do vetor de entrada associa-se a um nó específico.

Camadas ocultas: Situadas entre a camada de entrada e a camada de saída, nelas os neurônios realizam transformações sobre os sinais recebidos, sejam oriundos da camada de entrada ou de outras camadas ocultas. Após o processamento, os resultados são transmitidos às camadas posteriores.

Camada de saída: Conforme a natureza da tarefa, essa camada pode conter um ou mais neurônios, com ou sem função de ativação, sendo responsável por fornecer o resultado final da rede.

3.1.2 Aprendizagem Supervisionada e Retropropagação

Visto a estrutura geral e os principais elementos que compõem uma RN feedforward, podemos agora discutir o método de aprendizagem e as operações básicas envolvidas no

processo de treinamento, as quais darão base para o desenvolvimento seguinte.

A aprendizagem supervisionada consiste em um paradigma de treinamento no qual um modelo é ajustado a partir de um conjunto de exemplos rotulados, isto é, de pares $(\mathbf{x}^{(i)}, y^{(i)})$, onde $\mathbf{x}^{(i)} \in \mathbb{R}^n$ representa o vetor de entrada e $y^{(i)}$ a respectiva saída ou rótulo desejado. O objetivo é construir uma função $f(\mathbf{x}; \theta)$, parametrizada por θ, capaz de mapear as entradas $\mathbf{x}$ para saídas preditas $\hat{y}$, minimizando o erro com relação aos valores reais y.

No contexto de RNs, cada neurônio executa duas operações fundamentais: uma soma ponderada dos sinais de entrada seguida da aplicação de uma função de ativação.

Seja $\mathbf{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ o vetor de entrada do neurônio, $\mathbf{w} = (w_1, w_2, \ldots, w_n)^{\mathrm{T}}$ o vetor de pesos associados e b o termo de viés. A soma ponderada é dada por:

$$s = \mathbf{w}^{\mathrm{T}}\mathbf{x} + b. \tag{3.1}$$

Em seguida, o resultado s é transformado pela função de ativação $\sigma(\cdot)$, resultando na saída do neurônio:

$$z = \sigma(s). \tag{3.2}$$

Durante o treinamento supervisionado, os parâmetros $\theta = (\mathbf{w}, b)$ são ajustados de modo a minimizar uma função de perda $C(y, \hat{y})$, que quantifica a discrepância entre a saída predita $\hat{y} = z$ e a saída real y. A atualização dos parâmetros é, em geral, realizada por algoritmos de otimização baseados em gradiente, como o método do gradiente descendente:

$$\theta \leftarrow \theta - \eta \nabla_\theta C(y, \hat{y}), \tag{3.3}$$

onde η é a taxa de aprendizado e $\nabla_\theta C$ denota o gradiente da função de perda em relação aos parâmetros.

Este processo iterativo de ajuste permite que a rede aprenda, a partir dos dados rotulados, o mapeamento mais adequado entre entradas e saídas, com o intuito de generalizar para novos dados não observados.

- **Backpropagation**

Embora não seja uma característica da arquitetura, devido a sua eficiência para calcular o gradiente da função de perda, o backpropagation tornou-se o método padrão para o treinamento de redes feedforward parametrizadas de forma diferenciável.

Conforme ele será parte[4] da estrutura de nosso modelo, nessa seção iremos explanar o

[4]Aplicado de forma implícita, através do método `tf.GradientTape.gradient()` do framework ten-

mesmo em mais detalhes para compreendermos sua integração no processo de atualização dos pesos na rede.

Baseado na descrição da Fig.(3.3), onde $l-1$ e l representam duas camadas consecutivas quaisquer em uma RN, destacamos a estrutura local sobre a qual o algoritmo de backpropagation atua: cada saída z_j^{l-1} é ponderada pelos pesos w_{ij}^l e somada ao viés b_i^l, resultando na soma ponderada s_i^l, cuja ativação gera z_i^l. Esse esquema aponta o fluxo forward básico, de uma conexão, sobre o qual o algoritmo de backpropagation opera para o ajuste dos parâmetros.

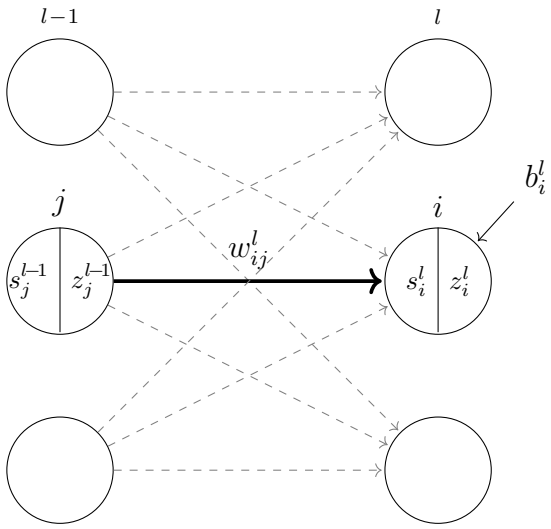


FIGURA 3.3 – Ilustração da conexão entre duas camadas consecutivas de uma RN, representadas pelos índices $l-1$ e l. Cada z_j^{l-1} da camada anterior é ponderada pelo respectivo w_{ij}^l e somada ao termo b_i^l, resultando na soma ponderada s_i^l, que posteriormente é processada pela função de ativação para produzir a saída z_i^l. Esta estrutura define o fluxo forward básico sobre o qual o algoritmo de backpropagation opera para ajuste dos parâmetros.

Inicialmente, por convenção, chamando w_{ij}^l o peso que liga a saída z_j^{l-1} ao neurônio i da camada l, a soma ponderada escreve-se

$$s_i^l = \sum_{j=1}^{n_{\text{I}}^l} w_{ij}^l \, z_j^{l-1} + b_i^l, \qquad i = 1, \ldots, n_{\text{C}}^l. \tag{3.4}$$

Aqui n_{I}^l é o número de entradas provenientes da camada $l-1$ e n_{C}^l é o número de neurônios da camada l. O termo s_i^l resulta do somatório das ativações z_j^{l-1} multiplicadas pelos respectivos pesos w_{ij}^l, ao qual se adiciona o viés b_i^l.

Observe que cada peso w_{ij}^l está associado à conexão que leva a ativação z_j^{l-1} da camada anterior ao neurônio i da camada l. Assim, a soma ponderada s_i^l resulta do produto entre

sorflow (TensorFlow Developers, 2024).

a linha i da matriz de pesos $\mathbf{W}^{\,l} \in \mathbb{R}^{n_{\mathsf{c}}^{l} \times n_{\mathsf{I}}^{l}}$ e o vetor-coluna das ativações $\mathbf{z}^{\,l-1} \in \mathbb{R}^{n_{\mathsf{I}}^{l}}$.

$$\mathbf{s^l} = \begin{bmatrix} s_1^l \\ \vdots \\ s_{n_{\mathsf{c}}^l}^l \end{bmatrix}, \qquad \mathbf{b^l} = \begin{bmatrix} b_1^l \\ \vdots \\ b_{n_{\mathsf{c}}^l}^l \end{bmatrix}, \qquad \mathbf{z^{l-1}} = \begin{bmatrix} z_1^{l-1} \\ \vdots \\ z_{n_{\mathsf{I}}^l}^{l-1} \end{bmatrix}, \qquad \mathbf{w^l} = \left[w_{ij}^l\right].$$

Dessa forma, a operação de soma ponderada para toda a camada é escrita como

$$\mathbf{s^l} = \mathbf{w^l}\,\mathbf{z^{l-1}} + \mathbf{b^l}. \tag{3.5}$$

Cada componente de $\mathbf{s^l}$ é, portanto, obtido pelo produto interno da linha correspondente de $\mathbf{w^l}$ com $\mathbf{z^{l-1}}$, seguido da adição do respectivo viés, onde n_{c}^l denota o número de neurônios da camada l.

Sendo y o vetor com a saída produzida pela rede e $\hat{y}$ o vetor com a saída esperada, definimos[5] como função de perda

$$C\left(y, \hat{y}\right) = \frac{1}{2MN} \sum_{k}^{N} \sum_{i}^{M} \left(y_{ik} - \hat{y}_{ik}\right)^2, \tag{3.6}$$

a soma dos quadrados dos erros, com M sendo o número de saídas da RN e N o número de amostras de treinamento. A fração que multiplica esse somatório é um fator de normalização.

Com o gradiente representando um vetor do espaço de parâmetros, dada a função de perda na Eq.(3.6), na prática precisamos apenas calcular as derivadas parciais que representam seus componentes, visto que a forma de atualização dos parâmetros da rede seguem a regra:

$$\bar{w}_{ij}^l \longleftarrow w_{ij}^l - \eta \frac{\partial C}{\partial w_{ij}^l}, \tag{3.7}$$

$$\bar{b}_{i}^l \longleftarrow b_{i}^l - \eta \frac{\partial C}{\partial b_{i}^l}, \tag{3.8}$$

onde $\partial C/\partial w_{ij}^l$ e $\partial C/\partial b_i^l$ são as derivadas da função de perda em relação aos pesos e bias e o parâmetro η representa a taxa de aprendizado que determina a velocidade com que a RN aprende. Um η muito alto pode fazer com que a rede passe por soluções ótimas, enquanto um η muito baixo pode fazer com que a rede demore muito para aprender ou fique presa em soluções sub-ótimas.

[5]Nesta seção adotamos uma taxa de aprendizagem efetiva $\eta = \eta/|B|$, já reescalonada pelo inverso do tamanho do *mini-batch*. Entenda por mini-batch o subconjunto de $|B|$ amostras extraído do conjunto de treinamento e processado em uma única iteração do algoritmo de descida do gradiente.

Os termos $\bar{w}_{ij}^l$ e $\bar{b}_i^l$ são composto por seus respectivos valores da iteração anterior, corrigidos de valor proporcional ao gradiente. O sinal negativo nessas equações indicam que estamos indo na direção contrária à do gradiente, conforme mencionado intenção de minimizar uma função de perda.

Como precisamos calcular as derivadas parciais da função de perda sobre todas as camada da rede, notando que y_{ik} é saída da rede z_i^L para o vetor de inputs da coluna k, ou seja x_k, temos que esse é o único termo dependente dos parâmetros w_{ij}^l e b_i^l, relevantes a derivação.

Dessa forma, sabendo que a derivada das somas é o mesmo que a soma das derivadas, fazendo uma manipulação algébrica sobre a Eq.(3.6), podemos obter que

$$\frac{\partial C}{\partial w_{ij}^l} = \frac{1}{MN} \sum_k^N \sum_i^M \frac{\partial}{\partial w_{ij}^l} \frac{1}{2} \left(y_{ik} - \hat{y}_{ik} \right)^2 , \tag{3.9}$$

$$\frac{\partial C}{\partial b_i^l} = \frac{1}{MN} \sum_k^N \sum_i^M \frac{\partial}{\partial b_i^l} \frac{1}{2} \left(y_{ik} - \hat{y}_{ik} \right)^2 . \tag{3.10}$$

Focando nossa atenção nos últimos termos dentro dos somatórios dessas equações, substituindo y_{ik} por $z_i^L(x_k)$ e aplicando a regra da cadeia,

$$\begin{aligned} \frac{\partial}{\partial w_{ij}^l} \frac{1}{2} \left(z_i^L(x_k) - \hat{y}_{ik} \right)^2 &= \left(z_i^L(x_k) - \hat{y}_{ik} \right) \left(\frac{\partial}{\partial w_{ij}^l} z_i^L(x_k) - \frac{\partial}{\partial w_{ij}^l} \hat{y}_{ik} \right) \\ &= \left(z_i^L - \hat{y}_i \right) \frac{\partial}{\partial w_{ij}^l} z_i^L . \end{aligned} \tag{3.11}$$

Eliminamos o termo $\hat{y}_{ik}$ da equação devido à sua independência em relação aos pesos e bias. Além disso, para simplificar a notação, suprimimos o sub-índice k que conecta a amostra específica e a dependência do vetor x_k.

Continuando o passo de derivação sobre o termo que contem z_i^L,

$$\frac{\partial}{\partial w_{ij}^l} z_i^L = \frac{\partial z_i^L}{\partial s_i^l} \frac{\partial s_i^L}{\partial w_{ij}^l}, \tag{3.12}$$

reescrevemos a Eq.(3.11),

$$\begin{aligned}\frac{\partial}{\partial w_{ij}^l}\frac{1}{2}\left(z_i^L(x_k)-\hat{y}_{ik}\right)^2 &= \overbrace{\left(z_i^L-\hat{y}_i\right)}^{\text{Derivada do erro}} \overbrace{\frac{\partial z_i^L}{\partial s_i^l}}^{\substack{\text{Derivada da função}\\ \text{de ativação}}} \frac{\partial s_i^L}{\partial w_{ij}^l} \\ &= \left(z_i^L-\hat{y}_i\right)\sigma'(s_i^l)\frac{\partial s_i^L}{\partial w_{ij}^l} \\ &= \delta_i^L\frac{\partial s_i^L}{\partial w_{ij}^l} \end{aligned} \tag{3.13}$$

onde o termo δ_i^L, composto pela multiplicação das derivadas do erro e da função de ativação, representa a contribuição do neurônio i da camada L da perda a qual estamos calculando.

Desenvolvido o passo anterior, podemos reescrever a derivada da função de perda com relação aos pesos como,

$$\frac{\partial C}{\partial w_{ij}^l} = \delta_i^l\frac{\partial s_i^l}{\partial w_{ij}^l}, \tag{3.14}$$

assim como, em processo análogo, aos bias

$$\frac{\partial C}{\partial b_i^l} = \delta_i^l\frac{\partial s_i^l}{\partial b_i^l}. \tag{3.15}$$

Obtidas as equações acima, podemos calcular as derivadas parciais da função de perda com relação a qualquer camada da RN apenas variando a definição de δ_i^l. Sendo que para a última camada temos a dependência do erro, enquanto as demais irão depender do δ da camada posterior.

Aprofundando em mais detalhes como calcular as Eqs.(3.14 - 3.15) para todas as camadas envolvidas no processo, inicialmente definiremos a quantidade δ_i^l como sendo a participação do neurônio i da camada l na derivada do erro.

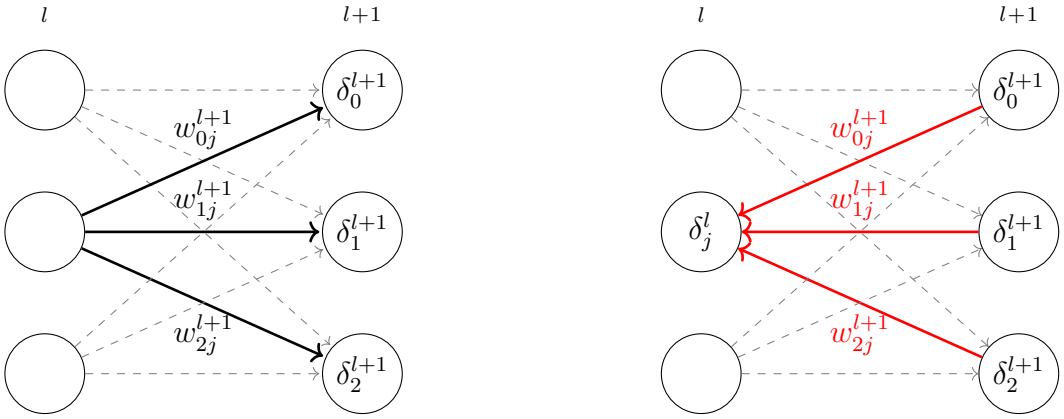
Conforme já vimos, partindo da última camada L,

$$\delta_i^L = \left(z_i^L-\hat{y}_i\right)\sigma'(s_i^L), \tag{3.16}$$

composto pela diferença entre as saídas da camada L e seus respectivos valores esperados multiplicada pela derivada da função de ativação para as entradas do neurônio.

No próximo passo, para reproduzirmos os δ com relação às camadas intermediárias vejamos os seguintes diagramas na Fig.(3.4), assumindo que os δ já foram calculados sob a camada $l+1$, onde os pesos w_{ij} conectam as entradas da camada $l+1$ com um neurônio

específico da camada l. Na forma matricial $\mathbf{w^{l+1}}$ refere-se aos pesos da coluna j da matriz de pesos da camada $l+1$.



(a) Conexões forward utilizadas na retropropagação

(b) Fluxo backward

FIGURA 3.4 – Cálculo dos termos de erro no backpropagation. Em (a), são mostradas as conexões forward da rede, representadas pelos pesos w_{ij}^{l+1} que conectam as camadas l e $l+1$. Os valores de erro δ_i^{l+1} já calculados na camada $l+1$ estão indicados em cada neurônio. Em (b), o fluxo backward (em vermelho) ilustra a propagação dos erros δ para a camada l, onde cada termo δ_i^{l+1} é ponderado pelos respectivos pesos w_{ij}^{l+1}, e o cálculo de δ_j^l é completado multiplicando essa soma ponderada pela derivada da função de ativação, conforme Eq.(3.18).

Destacados em vermelho na Fig.(3.4b) os fluxos de propagação backward , para calcularmos δ_j^l, temos que

$$\delta_0^{l+1} w_{0j}^{l+1} + \delta_1^{l+1} w_{1j}^{l+1}, \cdots, \delta_n^{l+1} w_{nj}^{l+1}, \tag{3.17}$$

onde os pesos w_{ij}^{l+1} servem como fator de ponderação regulando a proporção que cada δ_i^{l+1} tem sobre esse passo. Como os pesos envolvidos nesse processo são os mesmos do sentido forward de propagação da rede, podemos escrever que

$$\delta_{\mathbf{j}}^{\mathbf{l}} = \left(\mathbf{w_j^{l+1}}\right)^{\mathrm{T}} \delta^{\mathbf{l+1}}\, \sigma'\left(\mathbf{s_j^l}\right), \tag{3.18}$$

define a relação para o cálculo de $\delta_{\mathbf{j}}^{\mathbf{l}}$ em sua forma matricial. Vale-se destacar que o resultante dessa equação é um escalar, uma vez que sua composição é dada através da multiplicação dos respectivos termos na camada $l+1$ da versão transposta da coluna j da matriz de pesos pelo vetor coluna δ e o escalar consequente da derivação da função de ativação sob a soma ponderada.

Voltando aos termos com derivadas nas Eqs.(3.14 - 3.15), utilizando a relação dada

pela Eq.(3.4),

$$s_i^l = w_{i0}^l z_0^{l-1} + w_{i1}^l z_1^{l-1} + w_{i2}^l z_2^{l-1} + \cdots + w_{ij}^l z_j^{l-1} + b_i^l, \tag{3.19}$$

claramente obtemos para a derivada da soma ponderada com relação a um peso específico

$$\frac{\partial s_i^l}{\partial w_{ij}^l} = z_j^{l-1}, \tag{3.20}$$

resultado esse devido ao termo que descreve a saída do neurônio j da camada $l-1$, ser o único elemento do somatório que multiplica o peso em particular, ou seja todos os demais elementos não contribuem para esse cálculo. Seguindo a mesma lógica para a derivada com relação ao bias, temos que

$$\frac{\partial s_i^l}{\partial b_i^l} = 1. \tag{3.21}$$

Com esses resultados, podemos reescrever as derivadas da função de perda

$$\frac{\partial C}{\partial w_{ij}^l} = \delta_i^l \, z_j^{l-1}, \tag{3.22}$$

onde o produto após a igualdade é formado pelo δ da camada de destino e a ativação do neurônio de origem referente a sinapse da camada $l-1$. Já para a derivada com relação ao bias

$$\frac{\partial C}{\partial b_i^l} = \delta_i^l, \tag{3.23}$$

sendo simplesmente definido pelo δ do respectivo neurônio. As duas equações acima, são, portanto, as formas finais para as derivadas parciais da função de perda com relação aos pesos e bias, enquanto a Eq.(3.16) e Eq.(3.18) definem o cálculo dos δ da última e demais camadas respectivamente.

Se substituirmos o termo δ da Eq.(3.22) por seu equivalente dado na Eq.(3.16), obtemos que

$$\frac{\partial C}{\partial w_{ij}^L} = \left(z_i^L - \hat{y}_i\right) \sigma'(s_i^L) \, z_j^{L-1}. \tag{3.24}$$

Essa expressão geral admite uma importante simplificação em situações específicas, quando há compatibilidade entre a função de perda e a função de ativação da camada de saída, fenômeno conhecido na literatura como *Loss-Activation Compatibility*.

Por exemplo, supondo que a camada de saída utilize uma função de ativação sigmoid, dada por

$$\sigma(s) = \frac{1}{1 + e^{-s}}, \tag{3.25}$$

e a função de perda adotada seja a entropia cruzada binária (cross-entropy), definida como

$$C = -\left[\hat{y}_i \ln(z_i^L) + (1 - \hat{y}_i) \ln(1 - z_i^L)\right], \tag{3.26}$$

ocorre um cancelamento exato entre os termos provenientes da derivada da função de perda e da derivada da função de ativação.

Podemos verificar isso, inicialmente, expandindo o cálculo do termo δ_i^L, como

$$\delta_i^L = \frac{\partial C}{\partial s_i^L} = \frac{\partial C}{\partial z_i^L} \cdot \frac{\partial z_i^L}{\partial s_i^L}. \tag{3.27}$$

Para a derivada da função de perda C com relação à saída z_i^L

$$\frac{\partial C}{\partial z_i^L} \quad = \quad -\left(\frac{\hat{y}_i}{z_i^L} - \frac{1 - \hat{y}_i}{1 - z_i^L}\right). \tag{3.28}$$

Em seguida, para a função de ativação, sua derivada em relação à soma ponderada é

$$\frac{\partial z_i^L}{\partial s_i^L} = \sigma'(s_i^L) = z_i^L(1 - z_i^L). \tag{3.29}$$

Substituindo esses resultados na expressão da derivada total, simplificando-se os termos,

$$\delta_i^L = z_i^L - \hat{y}_i, \tag{3.30}$$

eliminando a necessidade de computar a derivada $\sigma'(s_i^L)$.

Portanto, a atualização dos pesos para a camada de saída passa a depender unicamente do erro residual e da ativação da camada anterior:

$$\frac{\partial C}{\partial w_{ij}^L} = \left(z_i^L - \hat{y}_i\right) z_j^{L-1}. \tag{3.31}$$

Esse comportamento analítico representa uma propriedade de grande importância prática, pois simplifica os cálculos de backpropagation e melhora a estabilidade numérica durante o treinamento.

3.2 Rede Neural Bayesiana

Com a evolução das pesquisas em AM na década de 1990, os métodos bayesianos ganharam projeção por sua capacidade em quantificar, de forma explícita, a incerteza associada às previsões dos modelos.

Esse interesse crescente resultou em trabalhos pioneiros como os de Radford Neal e David MacKay que demonstraram como a inferência bayesiana poderia ser conjugada à flexibilidade das RNs (Neal, 1996; MacKay, 1992), inaugurando o conceito de Redes Neurais Bayesianas (RNBs). O livro *Bayesian Learning for Neural Networks*, de Neal, firmou as bases teóricas e computacionais dessa integração, estabelecendo uma ponte consistente entre estatística probabilística e modelos de alto poder de representação.

Desde então, dois avanços foram decisivos para tornar as RNBs viáveis em cenários de larga escala: (i) técnicas de inferência aproximada, como a variacional e a Monte Carlo via dinâmica Hamiltoniana, que reduzem o custo de amostragem posterior; e (ii) *frameworks* modernos, notadamente o *TensorFlow* aliado ao *TensorFlow Probability*, que oferecem primitivas diferenciáveis para distribuições, *kernels* de MCMC (Markov Chain Monte Carlo) e camadas bayesianas nativas (Abadi *et al.*, 2016; Dillon *et al.*, 2017). Esses desenvolvimentos viabilizam a especificação de arquiteturas complexas com poucas linhas de código, ao mesmo tempo em que preservam a rastreabilidade matemática dos componentes probabilísticos.

Também não podemos deixar de mencionar que a adoção da abordagem bayesiana confere uma série de vantagens estruturais, as quais iremos destacar a seguir, que vão além do mero ajuste de hiperparâmetros:

Incorporação de conhecimento: Os *priors* podem codificar resultados empíricos anteriores ou restrições físicas, permitindo inferência cumulativa em que novos dados refinam, em vez de contradizer, o estado de conhecimento vigente.

Saídas probabilísticas: A solução é dada por distribuições *a posteriori*, fornecendo densidades, intervalos de credibilidade e probabilidades diretas de eventos, úteis para decisões baseadas em risco.

Robustez em amostras pequenas: Ao reduzir a variância das estimativas, a informação do *prior* estabiliza o aprendizado quando o número de exemplos é limitado, evitando que o modelo se ajuste a flutuações aleatórias.

Quando esses princípios são incorporados a uma RN, surgem ganhos adicionais. Além da regularização implícita fornecida pelos *priors*, que mitigam o sobre-ajuste em conjuntos de dados ruidosos ou de baixa dimensão, a própria estrutura hierárquica da rede permite codificar diretamente conhecimento disciplinar, como simetrias ou restrições físicas. O

resultado dessa combinação é especialmente valioso em contextos nos quais avaliar precisamente "quão certo" (nível de confiança atribuído a cada predição) é tão fundamental quanto o valor previsto. Em se tratando de cenários desafiadores, como no problema aqui estudado, não basta apenas predizer valores consistentes com as leis físicas; torna-se igualmente essencial estabelecer intervalos de credibilidade rigorosos e dispor de métricas objetivas para avaliar hipóteses concorrentes. Tais características são cruciais diante da complexidade envolvida na análise astrofísica, em que as observações são escassas e heterogêneas.

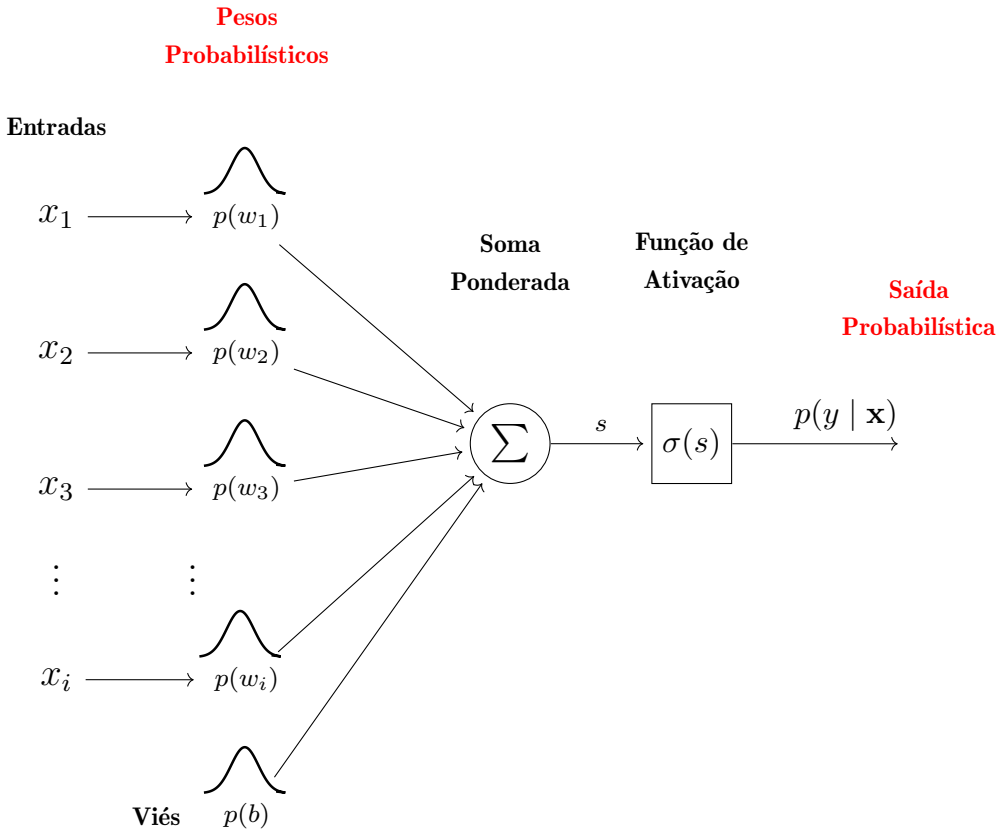


FIGURA 3.5 – Modelo de Neurônio Bayesiano: as entradas são combinadas com pesos tratados como parâmetros incertos, para os quais se especificam distribuições *a priori*. A soma ponderada resultante é processada por uma função de ativação, gerando uma saída probabilística $p(y \mid \mathbf{x})$.

Na Fig.(3.5), caracterizando a essência das diferenças entre RNs determinísticas e probabilísticas, podemos observar a representação idealizada de um neurônio utilizado por essas redes. Note que, neste esquema, não temos mais um conjunto de pesos exatos w_i que definem uma conexão como visto na Fig.(3.1). São atribuídas distribuições a priori $p(w_j)$ e $p(b)$, expressando o conhecimento prévio ou suposições sobre seus valores antes da observação dos dados. As entradas $(x_1, x_2, \ldots, x_i, x_b)$ são ponderadas por esses pesos, gerando a combinação linear $s = \sum_{j=1}^{i} w_j x_j + w_b x_b$, cuja incerteza decorre da própria modelagem dos parâmetros. O termo s é então transformado por uma função de ativação

$\sigma(s)$, resultando em uma saída probabilística $p(y \mid \mathbf{x})$, que incorpora tanto a predição quanto a incerteza associada ao modelo diante dos dados observados.

Apresentado a ideia conceitual por de trás desse tipo de modelo probabilístico, podemos sumarizar os principais elementos que caracterizam uma RN como bayesiana:

Pesos probabilísticos: Cada parâmetro é tratado como variável aleatória.

Distribuição *a priori* explícita: Define-se $p(\boldsymbol{\theta})$ antes de observar os dados.

Inferência da *posteriori*: Estima-se $p(\boldsymbol{\theta} \mid \mathcal{D})$ ou uma aproximação variacional.

Incerteza: O modelo produz médias e intervalos de credibilidade sobre os resultados.

Regularização estatística: Treinamento combina perda e influência da *prior*.

No decorrer do atual capítulo, todos esses elementos que foram listados serão abordados em maiores detalhes.

3.2.1 Camada Variacional

Conforme, em breve, passaremos a discutir os aspectos técnicos vinculados à estruturação de nosso modelo, a seguir, apresentamos a fundamentação teórica que integra a Estatística Bayesiana às Redes Neurais, estabelecendo o elo conceitual que sustenta a inferência variacional adotada neste trabalho.

Sejam, respectivamente, os vetores que reúnem todos os pesos e viéses:

$$\mathbf{w} := (w_1, \ldots, w_K)^\top, \tag{3.32}$$

$$\mathbf{b} := (b_1, \ldots, b_L)^\top. \tag{3.33}$$

Podemos empilhá-los, definindo um único vetor de parâmetros:

$$\boldsymbol{\theta} := (\mathbf{w}, \mathbf{b})^\top \in \mathbb{R}^d. \tag{3.34}$$

Dando o primeiro passo em direção a uma RN formada a partir de conceitos probabilísticos, assumimos que esses parâmetros seguem uma distribuição normal multivariada:

$$\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{3.35}$$

onde $\boldsymbol{\mu} \in \mathbb{R}^d$ é o vetor de médias; $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$ é a matriz de covariância[6], simétrica e definida positiva, por construção.

[6]Seus elementos diagonais contêm as variâncias de cada componente de $\boldsymbol{\theta}$ e os termos fora da diagonal representam covariâncias, indicando correlações lineares entre os componentes (Murphy, 2012).

Como padrão, para qualquer vetor aleatório $\mathbf{x} \in \mathbb{R}^d$, a representação funcional desse tipo de distribuição possui a forma,

$$\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^d \, |\boldsymbol{\Sigma}|}} \exp\Big[-\tfrac{1}{2} \, (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\Big], \tag{3.36}$$

onde $|\boldsymbol{\Sigma}|$ denota o determinante de $\boldsymbol{\Sigma}$.

Com base na Eq.(3.36), especificamos a distribuição *a priori* (ou simplesmente *prior*):

$$p(\boldsymbol{\theta}) = \mathcal{N}\big(\boldsymbol{\theta} \mid \boldsymbol{\mu}_p, \; \boldsymbol{\Sigma}_p\big), \tag{3.37}$$

que incorpora o conhecimento preliminar sobre os parâmetros antes de observarmos dados.

Após observar o conjunto $\mathcal{D}$, interessamo-nos pela distribuição *a posteriori*

$$p(\boldsymbol{\theta} \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid \boldsymbol{\theta}) \, p(\boldsymbol{\theta})}{p(\mathcal{D})}. \tag{3.38}$$

Diante da intratabilidade da distribuição *a posteriori* exata, optamos por uma abordagem de inferência variacional (Jordan *et al.*, 1999). Esse método consiste em definir uma família paramétrica de distribuições aproximadoras $q(\boldsymbol{\theta})$, e buscar, dentro dessa família, aquela cujos parâmetros minimizem a divergência (em geral, a divergência de Kullback-Leibler) em relação à posteriori verdadeira. Desta forma, transformamos o problema de inferência em um problema de otimização.

Para nosso modelo, especificamos a mesma família variacional de nossa *prior*, uma densidade normal multivariada:

$$q(\boldsymbol{\theta}) = \mathcal{N}\big(\boldsymbol{\theta} \mid \boldsymbol{\mu}_q, \; \boldsymbol{\Sigma}_q\big), \tag{3.39}$$

na qual os parâmetros $\boldsymbol{\mu}_q$ e $\boldsymbol{\Sigma}_q$ são aprendidos diretamente durante o treinamento da rede, de modo a ajustar a aproximação $q(\boldsymbol{\theta})$ à verdadeira posteriori $p(\boldsymbol{\theta} \mid \mathcal{D})$.

Assim, enquanto $(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p)$ representam o conjunto de hiperparâmetros do *prior*, transmitindo o conhecimento prévio disponível, os parâmetros $(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q)$ configuram a aproximação variacional, ambos respeitando a mesma estrutura funcional da densidade normal multivariada estabelecida na Eq.(3.36) (Murphy, 2012).

- **Decomposição de Cholesky**

Ao adotarmos essa representação para os parâmetros $\boldsymbol{\theta}$, estabelecendo que cada distribuição, *prior* ou *a posteriori* variacional, é especificada por um vetor de médias $\boldsymbol{\mu}$ e uma matriz de covariâncias $\boldsymbol{\Sigma}$, a representação de $\boldsymbol{\Sigma}$ através da decomposição de Cholesky torna os cálculos em alta dimensão computacionalmente viáveis e estáveis. A seguir,

abordaremos essa técnica, destacando suas vantagens e aplicação em nossa estrutura.

Consistindo, basicamente, de um processo de fatoração, a decomposição de Cholesky expressa a matriz de covariância como o produto de uma matriz A por sua própria transposta (Kingma; Welling, 2013), isto é

$$\boldsymbol{\Sigma} = AA^{\mathrm{T}}, \tag{3.40}$$

em que a matriz A é triangular inferior e única, desde que $\boldsymbol{\Sigma}$ seja definida positiva. A utilização dessa decomposição garante automaticamente a positividade em $\boldsymbol{\Sigma}$, condição essencial para a validade dos cálculos probabilísticos na RN (Blundell *et al.*, 2015).

Outro ganho no uso dessa técnica está em evitar procedimentos numéricos custosos, tais como a extração explícita da raiz quadrada ou a inversão direta de matrizes, simplificando a implementação e proporcionando maior eficiência computacional.

Definindo uma dos principais elementos que caracterizam uma RN como probabilística, ao longo do treinamento da rede, diferentes conjuntos de parâmetros $\boldsymbol{\theta}$ são amostrados de forma:

$$\boldsymbol{\theta} = \boldsymbol{\mu} + A\,\mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d), \tag{3.41}$$

onde $\mathbf{z}$ é ruído[7] padrão independente. Esse mecanismo, chamado de reparametrização, permite as estimativas do modelo e consequentemente os cálculos de erros e gradientes necessários no treinamento junto ao algoritmo backpropagation. Temos como resultado da implementação desse método um processo mais simples e estável para gerar amostras sobre as distribuições. Aspecto fundamental para capturar adequadamente a incerteza associada aos parâmetros $\boldsymbol{\theta}$ (pesos e viés) do modelo.

Em conformidade com a notação que estabelecemos, as dimensões dos parâmetros vetoriais discutidos[8] e os componentes estruturais da RN são relacionados segundo as igualdades:

$$K = n_{\mathsf{c}}^{l} \times n_{\mathsf{I}}^{l}, \tag{3.42}$$

$$L = n_{\mathsf{c}}^{l}, \tag{3.43}$$

$$d^{l} = K + L = n_{\mathsf{c}}^{l}\left(n_{\mathsf{I}}^{l} + 1\right), \tag{3.44}$$

com o sobre-índice representando a l-ésima camada da rede, estas expressões decorrem diretamente da forma como definimos inicialmente o vetor de parâmetros $\boldsymbol{\theta}$, responsável

[7] Trata-se de um vetor, definido através da chamada de um gerador de número aleatório sobre uma distribuição normal univariada ($\mu = 0$, $\sigma^2 = 1$), possibilitando o processo de amostragem.

[8] Tomado os parâmetros da rede como uma forma de distribuição, é sistematicamente apresentado uma série de definições dentre as Eqs.(3.32 - 3.41), as quais estabelecem as dimensões vetoriais mencionadas.

pela ligação entre a parametrização probabilística e a estrutura da RN.

Total de Parâmetros: Dado a forma das distribuição escolhidas para nossa modelo, cada amostra de $\boldsymbol{\theta}^l$, gerada durante o processo de inferência, possui como quantidade de parâmetros:

$$n_\theta = \underbrace{d_l}_{\text{parâmetros livres de } \boldsymbol{\mu}^l} + \underbrace{\frac{d_l(d_l+1)}{2}}_{\text{parâmetros livres de } A^l} . \tag{3.45}$$

Aqui, o segundo termo da soma reflete o número de parâmetros associados à matriz A^l, a qual, por ser definida como triangular inferior, possui apenas metade de suas entradas efetivamente livres[9]. Durante o treinamento, a rede aprende esses n_θ parâmetros com o objetivo de aproximar[10] a verdadeira distribuição *a posteriori*, permitindo a modelagem explícita da incerteza em suas predições associadas aos pesos.

- **Divergência de Kullback-Leibler**

A divergência de Kullback-Leibler ($\mathrm{D_{KL}}$) quantifica, de forma integral, o desvio entre as distribuições $p(\boldsymbol{\theta})$ e $q(\boldsymbol{\theta})$. Em RNBs, essa divergência atua como regularizador estatístico, constrangendo a distribuição variacional a permanecer aderente ao conhecimento codificado na *prior*, salvo quando as evidências observadas indicam com clareza a necessidade de afastamento (Kullback; Leibler, 1951; Cover; Thomas, 2006; Murphy, 2012).

Para uma arquitetura com M camadas variacionais independentes, a contribuição total da divergência é

$$\mathrm{D_{KL}} \;=\; \sum_{j=1}^{M} \mathrm{KL}\big[q(\boldsymbol{\theta}^j) \,\|\, p(\boldsymbol{\theta}^j)\big] \;=\; \sum_{j=1}^{M} \mathrm{E}_{q(\boldsymbol{\theta}^j)}\left[\log \frac{q(\boldsymbol{\theta}^j)}{p(\boldsymbol{\theta}^j)}\right], \tag{3.46}$$

onde $\boldsymbol{\theta}^j$ representa o vetor de parâmetros associados à j-ésima camada variacional (CV).

A Eq.(3.46) compõe a função definida como o Limite Inferior da Evidência, também conhecido como ELBO (*Evidence Lower Bound*), dada pela seguinte expressão (Blei; Kucukelbir; McAuliffe, 2017):

$$\mathrm{ELBO} \;=\; \sum_{i=1}^{N} \mathrm{E}_{q(\boldsymbol{\theta})}\left[\log p(\mathbf{y}_i \mid \mathbf{x}_i, \boldsymbol{\theta})\right] \;-\; \mathrm{D_{KL}}, \tag{3.47}$$

em que o primeiro termo é o somatório da esperança de N amostras, sob q, do logaritmo da verossimilhança e o segundo termo penaliza a divergência em relação *a prior*. Assim,

[9] Refere-se aos parâmetros que o modelo, de fato, tem disponível para serem treinados.
[10] Via inferência variacional, ajustando os valores de $(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q)$.

maximizar a ELBO concilia fidelidade aos dados com aderência à estrutura probabilística imposta pela distribuição *a priori*.

Conforme definimos o mesmo tipo de distribuição para representar nossa *prior* e *a posteriori* variacional, a divergência admite expressão analítica:

$$\mathrm{D_{KL}} \;=\; \frac{1}{2}\left[\log\frac{|\boldsymbol{\Sigma}_p|}{|\boldsymbol{\Sigma}_q|} \;-\; d \;+\; \mathrm{tr}\big(\boldsymbol{\Sigma}_p^{-1}\boldsymbol{\Sigma}_q\big) \;+\; (\boldsymbol{\mu}_p-\boldsymbol{\mu}_q)^{\mathrm{T}}\boldsymbol{\Sigma}_p^{-1}(\boldsymbol{\mu}_p-\boldsymbol{\mu}_q)\right], \tag{3.48}$$

onde d denota a dimensionalidade de $\boldsymbol{\theta}$. Esta forma fechada é diferenciável, compatibiliza-se naturalmente com o processo de reparametrização (Eq.3.41) e, portanto, integra-se sem custo adicional ao processo de retropropagação.

Em resumo, a $\mathrm{D_{KL}}$ é a ponte que vincula a inferência bayesiana com o aprendizado supervisionado em RNs, promovendo um equilíbrio entre a adequação aos dados observados e a preservação da estrutura probabilística imposta pelo *prior*. Seu papel como regularizador adaptativo é crucial para controlar a complexidade do modelo e evitar sobreajuste em cenários com incerteza significativa.

Como elemento final dentre as características que destacamos inicialmente, seja $\boldsymbol{\phi}$ o vetor de parâmetros variacionais que especifica a distribuição $q_{\boldsymbol{\phi}}(\boldsymbol{\theta})$. Considerando o conjunto de dados de treino $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N}$, define-se a função-perda

$$\mathcal{J}(\boldsymbol{\phi}) \;=\; -\mathrm{ELBO}(\boldsymbol{\phi}) \;=\; \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}_{q_{\boldsymbol{\phi}}(\boldsymbol{\theta})}\Big[-\log p(\mathbf{y}_i \mid \mathbf{x}_i, \boldsymbol{\theta})\Big] \;+\; \frac{1}{N}\,\mathrm{D_{KL}}(\boldsymbol{\phi}), \tag{3.49}$$

onde o primeiro termo da Eq.(3.49) corresponde à média, sobre o conjunto completo, da perda de log-verossimilhança calculada sob a aproximação variacional, enquanto o segundo regulariza a divergência entre *a posteriori* aproximada e o *prior*.

Ao minimizar $\mathcal{J}(\boldsymbol{\phi})$ maximiza-se, de modo equivalente, a ELBO, equilibrando fidelidade aos dados e coerência com o conhecimento prévio. Detalhes operacionais relativos a amostragem, reparametrização e atualização estocástica de gradientes serão apresentados no capítulo dedicada ao procedimento de treinamento.

3.3 Modelo

Utilizando a Eq.(3.44) que fornece o número de dimensões d^l, sabendo que a uma CV de nosso modelo será composta pelo número de neurônios $n_{\mathsf{c}}^l = 2$, que recebem o número

de entradas $n_{\text{I}}^l = 12$ cada, logo

$$d^l = 2\left(12 + 1\right) = 26. \tag{3.50}$$

Com o valor de d^l, através da Eq.(3.45), obtemos como total

$$n_\theta^l = 26 + \frac{26 \cdot 27}{2} = 26 + 351 = 377, \tag{3.51}$$

parâmetros treináveis, em uma CV.

Embora representadas pelo mesmo tipo de distribuição, cada uma das Eqs.(3.37, 3.39) possui um propósito específico em relação aos cálculos realizados ao longo do processo treinamento. Como sabemos, idealmente, a distribuição *prior* deve conter informações que reflitam algum conhecimento prévio sobre os dados, enquanto *a posteriori* deverá, nesse caso, modelar.

Para tanto, segmentamos a estrutura preditiva de nosso modelo em 3 partes, replicando o fluxo de entrada de dados para cada CV. Feita essa subdivisão, definimos a relação do conhecimento prévio sobre as distribuições, atribuindo aos parâmetros $\boldsymbol{\mu}_p$ e diag($\boldsymbol{A}_p$) de cada CV (Hoffman *et al.*, 2013; Blei; Kucukelbir; McAuliffe, 2017), os respectivos valores estatísticos de média e desvio padrão referentes aos dados-alvo utilizados no processo de treinamento.

Essa divisão, além de permitir o ajuste pontual de cada *prior*, tem como objetivo manter um processo de aprendizagem que reflita a especificidade da saída e ao mesmo tempo a correlação entre elas. De forma mais clara, nossa proposta é que, em paralelo, cada um dos três ramos definidos na RN possa ter um treinamento com maior foco sobre a previsão do respectivo/específico parâmetro $(\Gamma_1, \Gamma_2, \Gamma_3)$ da EdE em questão.

Com isso, totalizando 1164 parâmetros treináveis para o modelo, cada ramo é composto por uma CV e um nó de saída sigmoidal; as amostras dessas saídas são posteriormente reunidas em um vetor final único.

Em RNs, uma função de ativação $\sigma : \mathbb{R} \to \mathbb{R}$ é aplicada elemento-a-elemento aos potenciais $\mathbf{z} = \mathbf{W}\mathbf{x} + \mathbf{b}$ produzidos por uma camada linear. Sua presença é essencial, pois introduz a não-linearidade necessária para que a rede não se reduza a uma simples transformação linear, possibilitando assim a modelagem de relações complexas entre entradas e saídas. Além disso, a função de ativação permite impor restrições sobre a faixa de valores da saída, garantindo que esta pertença a domínios específicos, como no caso de probabilidades em $[0, 1]$ ou desvios padrão positivos.

Diversas escolhas são possíveis (ReLU, *tanh*, *softplus* etc.), mas neste trabalho optou-se pelo uso de duas funções com papéis distintos. A primeira delas é a **ativação exponencial**,

empregada na camada `DenseVariational`, parametrizada por médias μ e escalas s dos pesos. Para assegurar a condição $s > 0$, define-se

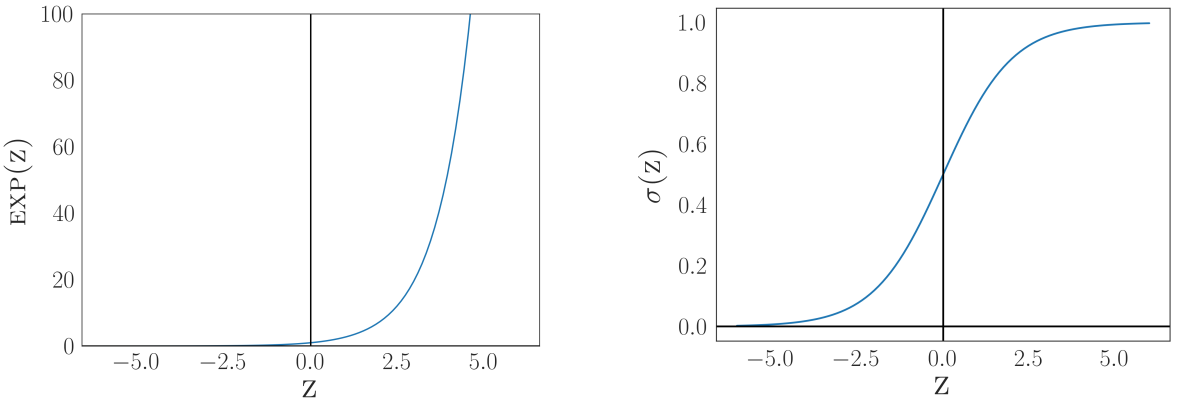
$$\sigma_{\exp}(z) = \exp(z), \qquad z \in \mathbb{R}. \tag{3.52}$$

Essa escolha faz com que a faixa da função seja $(0, \infty)$, assegurando positividade sobre os parâmetros amostrados. Sua derivada coincide com a própria função, $\sigma'_{\exp}(z) = \sigma_{\exp}(z)$, sendo sempre estritamente positiva e evitando zonas mortas no espaço de parâmetros. Do ponto de vista bayesiano, o mapeamento de $\log s$ para s permite que a otimização ocorra em todo $\mathbb{R}$, respeitando ao final a restrição estatística de $s > 0$. A segunda função adotada é a **sigmoid**, utilizada para a previsão dos parâmetros Γ_j em cada ramo j, definida por

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \qquad z \in \mathbb{R}, \tag{3.53}$$

o que produz valores $\hat{y}^j \in (0, 1)$. Sua faixa, portanto, é adequada para grandezas normalizadas ou probabilísticas. A derivada da sigmoid é dada por $\sigma'(z) = \sigma(z)\big(1 - \sigma(z)\big)$, atingindo valor máximo em $z = 0$ e decrescendo nas extremidades, o que pode ocasionar saturação, mas sem relevância quando z permanece em regime moderado. No contexto do presente trabalho, os parâmetros Γ_j foram previamente reescalonados para o intervalo unitário antes do treinamento, e a sigmoid retorna diretamente valores nessa mesma escala, eliminando a necessidade de transformações adicionais na saída.

Assim, a ativação exponencial assegura a positividade do desvio padrão variacional, enquanto a sigmoid garante que as predições finais permaneçam normalizadas em $[0, 1]$, compatíveis com o pré-processamento realizado. Ambas cumprem, portanto, os papéis fundamentais de introduzir não-linearidade e impor restrições adequadas de faixa exatamente nos pontos da arquitetura onde tais propriedades são exigidas.



(a) Ativação Exponencial

(b) Ativação Sigmoid

FIGURA 3.6 – (a) Função ativação exponencial $\exp(z)$, utilizada nas camadas variacionais do modelo. (b) Função ativação sigmoid $\sigma(z) = (1 + e^{-z})^{-1}$, utilizada nas camadas densas de saída do modelo para mapear as predições ao intervalo $(0, 1)$.

TABELA 3.1 – Descrição das camadas da estrutura do modelo

Ramo	Camada	Ativação	Neurônios	Finalidade	$\boldsymbol{\theta}$ (treináveis)
–	0	–	12	entrada de atributos (M, R)	–
–	1	BatchNorm	–	normalização em lote	24
1	2	exponencial	2	extrair features para Γ_1	377
	3	sigmoid	1	estimar Γ_1	3
2	4	exponencial	2	extrair features para Γ_2	377
	5	sigmoid	1	estimar Γ_2	3
3	6	exponencial	2	extrair features para Γ_3	377
	7	sigmoid	1	estimar Γ_3	3
–	8	–	–	agrupar previsões	–

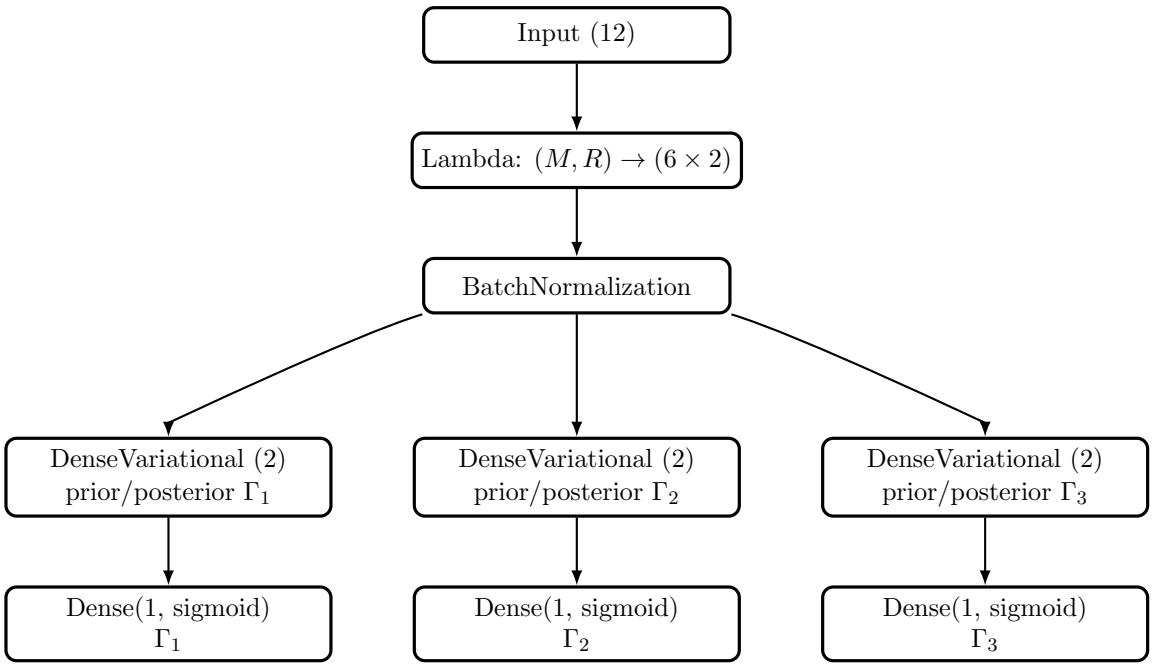


FIGURA 3.7 – Arquitetura da rede neural bayesiana adotada. A entrada de dimensão 12 é reorganizada em uma estrutura (6×2) por uma camada Lambda, seguida de normalização em lote para estabilizar as variáveis. Em seguida, a rede se divide em três ramos paralelos, cada um com uma camada DenseVariational de 2 unidades, parametrizada por distribuições a priori e a posteriori. Esses ramos aprendem representações distintas e produzem, via camada densa com ativação sigmoid, as previsões dos parâmetros (Γ_1, Γ_2, Γ_3). O treinamento combina perda Huber em cada saída com regularização pelo termo KL, equilibrando ajuste e complexidade do modelo.

3.4 Treinamento

Na seção anterior contornamos alguns aspectos relativos ao treinamento desse tipo de modelo, onde, inclusive, um deles foi indicado como um dos elementos que o caracterizam como bayesiano. Na Eq.(3.49) apresentamos a forma clássica da função que dever ser minimizada durante o treinamento do modelo. Baseada na Evidência Inferior Variacional, essa equação equilibra dois componentes essenciais: a verossimilhança dos dados sob os parâmetros amostrados e a $\mathrm{D_{KL}}$ que estabelece regularidade entre as distribuições $q(\boldsymbol{\theta})$ e $p(\boldsymbol{\theta})$.

Contudo, tal formato não é uma imposição rígida do arcabouço bayesiano. O elemento que define o caráter probabilístico do modelo não é a forma exata da verossimilhança, mas sim a presença da inferência sobre distribuições de pesos, a regularização via $\mathrm{D_{KL}}$, e a coerência estatística entre a função de perda adotada e uma distribuição plausível dos erros dos dados.

Neste trabalho, propomos uma generalização dessa abordagem utilizando a *Huber Loss*, que chamaremos de Perda Huber (PH), como substituto do logaritmo negativo da verossimilhança (Huber, 1964). Essa função de perda é particularmente útil em contextos onde os resíduos entre as predições do modelo e os valores reais apresentam comportamento errático, com a presença de outliers ou dispersões não modeladas adequadamente por uma distribuição normal.

A PH pode ser interpretada como uma aproximação contínua por partes do logaritmo da densidade de uma distribuição mista entre a normal e a de Laplace, apresentando um comportamento quadrático em torno da origem (para erros pequenos) e linear nas regiões periféricas (para erros grandes). Em termos probabilísticos, isso equivale a assumir que os resíduos $r = y - \hat{y}$ são centrados em zero, mas com caudas mais pesadas do que as da distribuição normal. Assim, a adoção da PH como função de verossimilhança implícita corresponde à modelagem de erros com maior robustez a valores extremos.

É importante destacar que essa escolha reflete uma nova hipótese sobre o comportamento dos dados, uma hipótese que, em diversas aplicações práticas, é mais realista do que a suposição de ruído gaussiano com variância constante.

Além disso, na literatura pode-se encontrar esse mesmo conceito aplicado, como por exemplo a substituição da verossimilhança modelada por uma distribuição normal sendo substituída diretamente pela função do Erro Quadrático Médio (EQM). Na Fig.(3.8) temos uma representação da relação existente entre o logaritmo negativo da verossimilhança de uma normal e a função de perda do tipo EQM, assim como a relação análoga entre a verossimilhança de uma distribuição de tipo Laplace e a função de PH.

A seguir, será sistematizado de forma objetiva o processo de treinamento do modelo

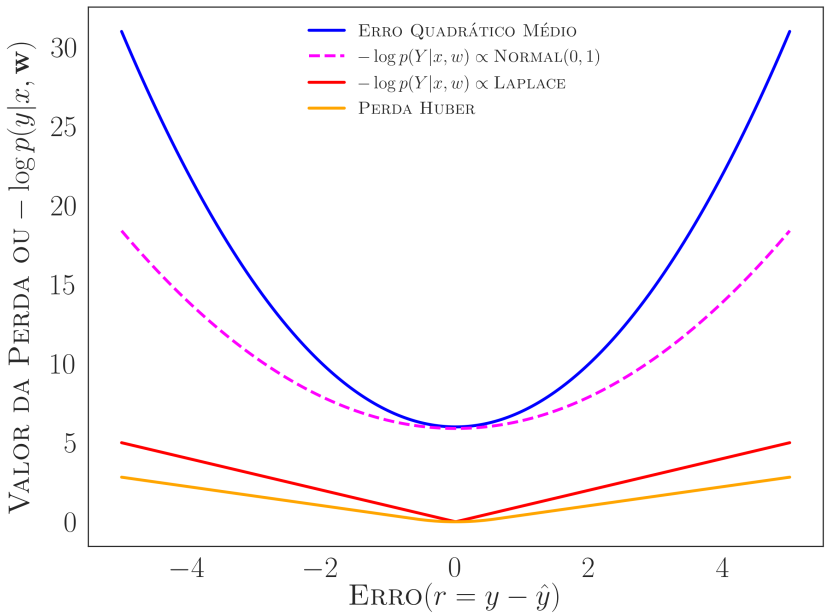


FIGURA 3.8 – Comparação entre as funções de perda determinísticas (EQM e Huber) e os logaritmos negativos das respectivas distribuições de verossimilhança (Normal e Laplace), ambos como função do erro residual $r = y - \hat{y}$. As curvas ilustram como essas perdas podem ser interpretadas como aproximações das respectivas log-verossimilhanças, justificando seu uso em substituição a modelagens probabilísticas explícitas em alguns cenários de aprendizado de máquina.

proposto, incluindo o fluxo de propagação das entradas e a composição da função de perda total utilizada para a otimização dos parâmetros. Os resultados obtidos a partir desse processo serão analisados na próxima seção.

Seja $\mathcal{D}_{\text{train}} = \left\{(\mathbf{x}_i, \mathbf{y}_i)\right\}_{i=1}^{N}$, $N \in \mathbb{N}_{>0}$, o conjunto de dados de treinamento, onde cada vetor-alvo $\mathbf{y}_i = \left(y_i^1, y_i^2, y_i^3\right)$ contém os parâmetros (Γ) que definem uma EdE.

Considerando a estrutura definida na Tabela 3.1, a RNB possui, então, três ramos independentes, indexados por $j \in \{1, 2, 3\}$, cada qual responsável pela predição de um Γ_j.

Fixado um tamanho para cada lote[11] de dados M, o número de mini-lotes por época de treinamento do modelo é $B = (N/M)$. Na época t $(t = 1, \ldots, T)$ o b-ésimo mini-lote é

$$\mathcal{B}_{(t,b)} = \left\{(\mathbf{x}_{t,b,m}, \mathbf{y}_{t,b,m})\right\}_{m=1}^{n_{t,b}}, \qquad b = 1, \ldots, B, \tag{3.54}$$

onde $n_{t,b} = |\mathcal{B}_{(t,b)}| = \min\left\{M,\ N - (b-1)M\right\}$ denota a cardinalidade do lote, isto é, o número de pares $(\mathbf{x}, \mathbf{y})$ que o compõem. Cada época executa B iterações de atualização sobre os parâmetros da rede.

[11] O lote (ou mini-lote) em aprendizagem de máquina é uma técnica que divide o conjunto de dados de treinamento em pequenos subconjuntos, usados para atualizar os pesos do modelo durante o treinamento.

Perda total por lote: Para $\mathcal{B}_{(t,b)}$ calcula-se a perda total

$$\mathcal{L}_{(t,b)} = \sum_{j=1}^{3} \alpha_{(t)}^{j}\, \bar{\mathcal{H}}_{(t,b)}^{j} \;+\; \frac{1}{N} \sum_{j=1}^{3} \mathrm{D}_{\mathrm{KL}}^{j}, \tag{3.55}$$

onde $\alpha_{(t)}^{j} \geq 1$ é o peso dinâmico do ramo j no início da época t e $\bar{\mathcal{H}}_{(t,b)}^{j}$ é a função de PH média do ramo j no lote (t, b).

Perda de Huber: Para cada par $(\mathbf{x}_{t,b,m}, \mathbf{y}_{t,b,m}) \in \mathcal{B}_{(t,b)}$, denote por $\hat{y}_{t,b,m}^{j}$ a predição do ramo j. O erro residual é $u_{t,b,m}^{j} = y_{t,b,m}^{j} - \hat{y}_{t,b,m}^{j}$. Definindo um limiar $\delta > 1$, a perda de Huber é

$$\mathcal{H}(u) = \begin{cases} \frac{1}{2}\, u^2, & |u| \leq \delta; \quad \text{Erro Quadrático Médio (EQM),} \\ \delta\big(|u| - \frac{1}{2}\delta\big), & |u| > \delta; \quad \text{Erro Absoluto Médio (EAM).} \end{cases} \tag{3.56}$$

A média por ramo no mini-lote resulta em

$$\bar{\mathcal{H}}_{(t,b)}^{j} = \frac{1}{n_{t,b}} \sum_{m=1}^{n_{t,b}} \mathcal{H}_{(u_{t,b,m})}^{j}. \tag{3.57}$$

Otimizador: Os parâmetros variacionais $\boldsymbol{\phi} = \{\phi_j\}_{j=1}^{3}$ que parametrizam $q_{\boldsymbol{\phi}}(\boldsymbol{\theta}^j)$ são ajustados com AdamW $\big(\beta_1 = 0.9,\ \beta_2 = 0.999,\ \varepsilon = 10^{-8}\big)$ e taxa inicial η. Para cada iteração n, que em nosso caso é equivalente ao número de mini-lotes $\mathcal{B}_{(t,b)}$, as atualizações escalarizadas seguem

$$g^{(n)} \;=\; \partial_{\boldsymbol{\phi}} \mathcal{L}_{(t,b)}, \tag{3.58}$$

$$m^{(n)} \;=\; \beta_1 m^{(n-1)} + (1-\beta_1)\, g^{(n)}, \qquad \hat{m}^{(n)} = \frac{m^{(n)}}{1-\beta_1^{n}}, \tag{3.59}$$

$$v^{(n)} \;=\; \beta_2 v^{(n-1)} + (1-\beta_2)\big(g^{(n)}\big)^2, \qquad \hat{v}^{(n)} = \frac{v^{(n)}}{1-\beta_2^{n}}, \tag{3.60}$$

$$\phi^{(n+1)} \;=\; \phi^{(n)} - \eta\, \frac{\hat{m}^{(n)}}{\sqrt{\hat{v}^{(n)}} + \varepsilon} \;-\; \eta\, \lambda_{\mathrm{WD}}\, \phi^{(n)}, \tag{3.61}$$

onde λ_{WD} é o coeficiente de *weight-decay* desacoplado do gradiente (característico do AdamW). A normalização adaptativa $\sqrt{\hat{v}^{(n)}}$ estabiliza o treinamento em alta dimensão, fundamental quando se amostram centenas de pesos por passo.

Ajuste dinâmico: Para cada época t define-se a PH média por ramo

$$\bar{\mathcal{H}}_{(t)}^{j} \;=\; \frac{1}{B} \sum_{b=1}^{B} \bar{\mathcal{H}}_{(t,b)}^{j}, \qquad j \in \{1, 2, 3\},$$

onde $\bar{\mathcal{H}}^{j}_{(t,b)}$ é a média por lote já introduzida na Eq. (3.57). Calcula-se então a referência média entre os ramos

$$\bar{\mathcal{H}}_{(t)} \;=\; \frac{1}{3}\sum_{j=1}^{3} \bar{\mathcal{H}}^{j}_{(t)}.$$

Com fator de adaptação $\lambda = 10^{-4}$, os pesos dinâmicos para a época seguinte são atualizados por

$$\alpha^{j}_{(t+1)} = \begin{cases} \alpha^{j}_{(t)}\big(1+\lambda\big), & \bar{\mathcal{H}}^{j}_{(t)} > \bar{\mathcal{H}}_{(t)}, \\ \max\big\{1,\, \alpha^{j}_{(t)}\big(1-\lambda/2\big)\big\}, & \text{caso contrário.} \end{cases}$$

Dessa forma, ramos cujo erro em PH excede a média recebem α^{j} ligeiramente maior, enquanto ramos mais precisos sofrem uma redução controlada jamais abaixo de 1), atualizando as contribuições na Eq. (3.55).

Esse procedimento completo, da formação dos mini-lotes à evolução dinâmica dos pesos de perda, tem como objetivo que o modelo aprenda simultaneamente $\Gamma_1, \Gamma_2, \Gamma_3$ de forma equilibrada respeitando as incertezas inerentes às camadas variacionais e a robustez conferida pela PH. Durante o processamento de cada mini-lote $\mathcal{B}_{(t,b)}$ o vetor de entrada, composto por 6 pares (M, R) que caracterizam as seis ENs, é primeiro reorganizado pela operação de *interleaving*[12], resultando em $\mathbf{x} \in \mathbb{R}^{12}$ compartilhado pelos três ramos da rede; em seguida, $\mathbf{x}$ é padronizado pela camada `BatchNormalization`[13].

Cada ramo j amostra 377 pesos variacionais a partir de $q(\boldsymbol{\theta}^{j})$ numa CV `DenseVariational` com $n_c^{l} = 2$ neurônios. A saída dessa camada alimenta um nó denso[14] com ativação `sigmoid`, produzindo a previsão $\hat{y}^{j}$ para o correspondente parâmetro Γ_j.

Para cada linha $(\mathbf{x}_{t,b,m}, \mathbf{y}_{t,b,m})$ calcula-se o residual $u^{j}_{t,b,m}$ e a PH elementar $\mathcal{H}(u^{j}_{t,b,m})$ (Eq. 3.56); a média no lote fornece $\bar{\mathcal{H}}^{j}_{(t,b)}$ (Eq. 3.57). Somando-se a $\mathrm{D_{KL}}$ devolvida pela CV, obtém-se a $\mathcal{L}_{(t,b)}$ via Eq.(3.55), em que o termo da divergência é dividido por N para manter a regularização compatível com o tamanho do conjunto de treinamento. Essa perda única é então passada ao otimizador AdamW (Loshchilov; Hutter, 2019), que atualiza os parâmetros variacionais $\boldsymbol{\phi}$ conforme Eq.(3.61).

Ao término dos B mini-lotes da época t, computa-se a PH média de cada ramo $\bar{\mathcal{H}}^{j}_{(t)} = B^{-1}\sum_b \bar{\mathcal{H}}^{j}_{(t,b)}$ e a média global $\bar{\mathcal{H}}_{(t)} = \frac{1}{3}\sum_j \bar{\mathcal{H}}^{j}_{(t)}$. Com o incremento $\lambda = 10^{-4}$, os pesos dinâmicos são atualizados por $\alpha^{j}_{(t+1)}$, o qual foi descrito anteriormente.

[12] Operação que rearranja os elementos do vetor de entrada (originalmente com 6 valores de M e em sequencia os de R) em uma sequência alternada $(M_0, R_0, M_1, R_1, \ldots, M_5, R_5)$, formando os pares M-R que representam diretamente as ENs do conjunto.

[13] A camada padroniza o vetor de atributos de cada mini-lote (subtrai a média e divide pelo desvio padrão) e aplica um reescalonamento e viés aprendíveis, reduzindo diferenças de escala entre características e acelerando o treinamento.

[14] Uma camada densa, também chamada de nó denso, aplica uma combinação linear dos dados de entrada, seguida por uma função de ativação, conectando todos os neurônios da entrada à saída.

Podemos sumarizar todo o processo de treinamento, para cada $\mathcal{B}_{(t,b)}$ (Eq. 3.54), nos seguintes passos:

1. **Distribuição de *Inputs*:** O vetor de entrada contém 6 pares (M, R) que, após a operação de *interleaving*, geram um tensor $\mathbf{x} \in \mathbb{R}^{12}$ comum aos três ramos da rede. Esse tensor é normalizado pela camada `BatchNormalization`.

2. **Amostragem - CV:** Cada ramo j executa uma CV com $n_{\mathtt{c}}^{l} = 2$ neurônios (Tabela 3.1). São amostrados 377 parâmetros variacionais segundo $q(\boldsymbol{\theta}^{j})$.

3. **Predição - Γ_j:** O resultado produzido pela CV alimenta um nó denso com função de ativação `sigmoid`, produzindo a saída $\hat{y}^{j}$ para o respectivo Γ_j da EdE.

4. **Cálculo de perdas:** Para cada amostra do lote calcula-se o residual $u^{j}_{t,b,m}$ e a PH é agregada segundo a Eq.(3.57); a $\mathrm{D}^{j}_{\mathrm{KL}}$ de cada ramo é automaticamente devolvida pela CV e somada conforme Eq.(3.55) dividida por N para preservar a escala da regularização.

5. **Otimização:** A perda total $\mathcal{L}_{(t,b)}$ alimenta o otimizador AdamW, que executa o passo de atualização dos parâmetros variacionais $\boldsymbol{\phi}^{(n)} \rightarrow \boldsymbol{\phi}^{(n+1)}$ via Eq. (3.61).

Após os B mini-lotes da época t:

6. **Ajuste dinâmico:** Calcula-se a PH média de cada ramo $\bar{\mathcal{H}}^{j}_{(t)}$ e a média global $\bar{\mathcal{H}}_{(t)}$ Com $\lambda = 10^{-4}$, o parâmetro $\alpha^{j}_{(t+1)}$ ajusta os pesos comparando os valores de PH médio de cada ramo ao global no final de cada época.

Essas etapas podem ser visualizadas através da Fig.(3.9), onde é exibido um fluxograma que direciona o treinamento da rede.

3.4.1 Análise dos Resultados do Treinamento

Uma vez detalhada a arquitetura proposta e o fluxo de treinamento, o processo de aprendizado foi conduzido ao longo de 100 épocas, utilizando mini-lotes de tamanho $5{,}12 \times 10^{4}$. Os resultados, resumidos nas Figs.(3.10, 3.12) e nas Tabelas 3.2, 3.3, permitem destacar três aspectos principais que caracterizam o desempenho do modelo:

a) **Convergência e estabilidade:** A curva da Perda Huber total (Fig. 3.10) apresenta decaimento exponencial nas primeiras 15 épocas, estabilizando-se em torno de $3{,}8 \times 10^{-2}$. A proximidade entre as curvas de treinamento e validação demonstra boa

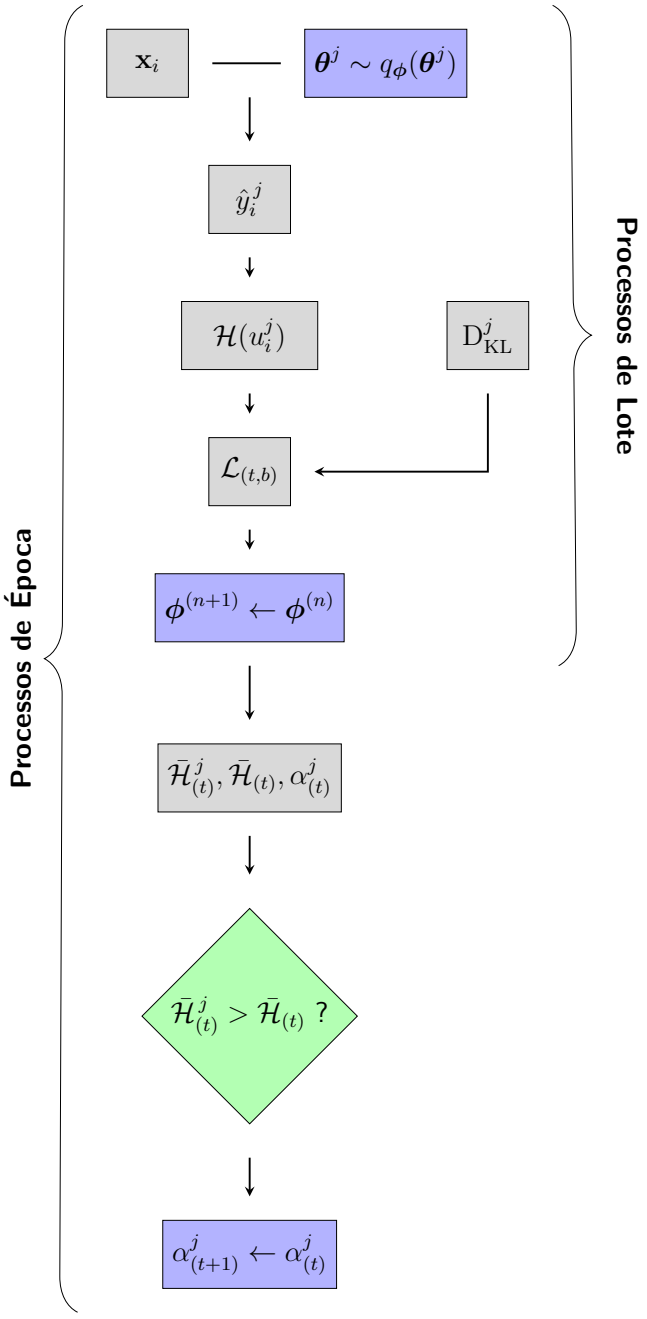


FIGURA 3.9 – Fluxograma de Treinamento: Descreve o fluxo de processos que ocorre por lote de dados e em uma época de treinamento.

capacidade de generalização, sem indícios significativos de sobreajuste para o horizonte analisado. Além disso, a baixa variabilidade nas últimas iterações reforça a estabilidade do processo de otimização.

b) **Desempenho por ramo:** As perdas individuais de cada saída, mostradas na Fig.(3.11), evidenciam uma hierarquia de dificuldades: a primeira saída, associada a Γ_1, converge para valores próximos de 1×10^{-3}; a segunda (Γ_2) estabiliza-se em torno de 1×10^{-2}; e a terceira (Γ_3) mantém-se na ordem de 3×10^{-2}. Ou seja, há diferenças de uma a duas ordens de grandeza entre os ramos, refletidas nas métricas do conjunto

de teste (Tabela 3.3), em que a REQM cresce de $3{,}38 \times 10^{-2}$ para $2{,}51 \times 10^{-1}$ entre a primeira e a terceira saída. Essa disparidade sugere que a predição de Γ_3, relacionada às regiões de maior densidade da EdE, envolve maior complexidade intrínseca e maior sensibilidade a variações residuais, enquanto Γ_1, ligado a densidades mais baixas, é aprendido de forma muito mais estável.

c) **Impacto do ajuste dinâmico:** O fator de ponderação α (Fig.(3.12)) apresenta variações sutis, controladas pelos coeficientes $\lambda_+ = 1{,}0 \times 10^{-2}$ e $\lambda_- = 5{,}0 \times 10^{-3}$. Observa-se aumento gradual de α_3, associado ao ramo de maior erro, atingindo $\approx 1{,}009$ na última época; α_1 sofre leve redução, enquanto α_2 permanece praticamente estável. Em termos de aprendizado, cada termo de perda $\bar{\mathcal{H}}^j$ é ponderado por α^j na Eq. (3.55); portanto, mesmo sendo otimizada uma única função de perda global, o gradiente parcial de cada ramo é efetivamente amplificado ou atenuado conforme o desempenho relativo. O crescimento de α_3 intensifica os ajustes dessa saída, enquanto a redução em α_1 previne sobrecorreção em um ramo já estável. O mecanismo, assim, redistribui o foco do modelo de forma equilibrada, sem necessidade de otimizadores separados e sem comprometer a estabilidade global da atualização.

A Tabela 3.2 resume a PH final do ciclo de treinamento, confirmando a hierarquia de dificuldade já observada entre as saídas. Para avaliar a relevância absoluta desses erros, definimos

$$\Omega_j \;=\; \max(y_j) \;-\; \min(y_j), \qquad p_j \;=\; \frac{\mathrm{EAM}_j}{\Omega_j} \;\times 100\,\%, \tag{3.62}$$

onde Ω_j representa a amplitude do alvo Γ_j no conjunto de teste e p_j corresponde ao Erro Absoluto Médio (EAM) expresso como fração percentual dessa amplitude.

Aplicando-se as amplitudes obtidas (2,83, 3,50, 3,00) e os valores de EAM da Tabela 3.3, obtêm-se $p_1 \approx 0{,}88\,\%$, $p_2 \approx 2{,}37\,\%$ e $p_3 \approx 6{,}83\,\%$. Dessa forma, mesmo no cenário mais desafiador (Γ_3), o erro médio permanece substancialmente inferior a $10\,\%$ da variação observada nos dados, enquanto para Γ_1 esse valor é inferior a $1\,\%$.

Esses resultados evidenciam que a estratégia adotada, combinação de PH com regularização bayesiana via $\mathrm{D_{KL}}$, garante aprendizado consistente mesmo em presença de resíduos potencialmente heterocedásticos[15]. A baixa taxa de erro relativo em Γ_1 confirma a capacidade do modelo de capturar padrões estáveis em regiões de menor complexidade física, enquanto o desempenho mais modesto em Γ_3 indica maior sensibilidade à variabilidade intrínseca dos dados de alta densidade.

[15]Termo utilizado para indicar que resíduos/erros não têm variância constante. Em nosso contexto, faz referência a discrepância dos valores dos erros conectados a diferentes regimes de densidade.

O leve desbalanceamento entre os ramos, refletido nos diferentes valores de p_j, encontra respaldo no ajuste dinâmico dos fatores α, que redistribuem o foco da otimização em direção às saídas de maior erro. Ainda assim, a análise sugere que ajustes no limiar δ da PH ou modificações específicas na arquitetura voltada à predição de Γ_3 poderiam reduzir essa disparidade.

Em síntese, a abordagem proposta conduz a uma convergência rápida, com generalização adequada e erros relativos compatíveis com a incerteza das observações astrofísicas atuais. Nas próximas seções, investigaremos de forma mais detalhada as distribuições dos pesos variacionais e os intervalos de credibilidade associados às estimativas dos parâmetros Γ_j, aprofundando a análise da incerteza preditiva do modelo.

Cada camada variacional é inicializada com um *prior* empírico[16]

$$p(\boldsymbol{\theta}^j) \;=\; \mathcal{N}\big(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\big), \tag{3.63}$$

onde o vetor de médias $\boldsymbol{\mu}_j = (\bar{\Gamma}_j, \ldots)$ é ancorado na média dos alvos do ramo j e a matriz de covariância $\boldsymbol{\Sigma}_j$ é parametrizada via decomposição de Cholesky, $\boldsymbol{\Sigma}_j = A_j A_j^{\mathrm{T}}$.

Aqui, A_j é uma matriz triangular inferior, onde seus elementos diagonais codificam as variâncias associadas a cada saída. Já os elementos[17] abaixo da diagonal representam correlações potenciais entre os parâmetros, permitindo que a rede as aprenda durante o treinamento. Essa escolha garante que $\boldsymbol{\Sigma}_j$ seja sempre simétrica definida positiva e preserva a expressividade multivariada da distribuição.

Para estabilizar a fase inicial e evitar correlações artificiais, adotamos como ponto de partida uma forma diagonal e isotrópica[18], isto é,

$$A_j = \mathrm{diag}(\sigma_{\Gamma_j}, \ldots, \sigma_{\Gamma_j}), \tag{3.64}$$

onde σ_{Γ_j} reflete a dispersão dos alvos correspondentes. Essa escala comum concentra a maior parte da massa de probabilidade em torno de valores plausíveis, compatíveis com a variabilidade real dos dados. Contudo, a presença de elementos fora da diagonal em A_j possibilita que, ao longo do aprendizado, o modelo flexibilize essa estrutura inicial e incorpore correlações relevantes entre os parâmetros da CV.

Na prática, a posterior variacional implementada em `VariableLayer` inicia próxima de $\mathcal{N}(\mathbf{0}, \mathbf{I})$, produzindo um pequeno desalinhamento inicial em relação ao prior empí-

[16] Valores estatísticos derivados diretamente dos dados de treinamento do modelo, são aplicados na construção das distribuições.

[17] Tais elementos são inicializados aleatoriamente.

[18] O uso do termo *isentrópico* visa apenas indicar que a pré-definição recaiu sobre a diagonal de A_j. Isso não implica que A_j seja estritamente diagonal, o que corresponderia a uma distribuição independente e uniforme. No nosso caso, a estrutura geral permanece multivariada, permitindo a captura de correlações entre parâmetros.

rico. Esse desajuste é corrigido pelos dados durante o treinamento, enquanto o termo de regularização $\mathrm{D}_{\mathrm{KL}}^{j}/N$[19] mantém o custo associado em escala reduzida. Observa-se na Fig.(3.13) que a divergência normalizada permanece no intervalo $7{,}0 \times 10^{-6}$ a $1{,}8 \times 10^{-5}$ ao longo das 100 épocas, indicando que o prior estabiliza a otimização sem sobrepor-se à verossimilhança induzida pela PH.

Do ponto de vista dinâmico, a evolução de D_{KL} ao longo das épocas revela uma tendência ascendente acompanhada por flutuações locais. Esse comportamento indica que a regularização não impõe rigidez excessiva, tendo em vista ela permite oscilações de curto prazo, mas conduz a um crescimento consistente da divergência. Observa-se ainda que, após aproximadamente a época 60, a taxa de aumento torna-se mais acentuada, sugerindo intensificação da adaptação da rede na fase final do treinamento. Já as distribuições *a posteriori* dos pesos, vistas na Fig.(3.14), evidenciam realocação diferenciada de capacidade entre os ramos. Note que a distribuição referente a Γ_1 permanece mais concentrado, com variância reduzida sem ramificações pronunciadas. A de Γ_2 apresenta indícios de bimodalidade, sugerindo flexibilidade intermediária. Em termos probabilísticos, significa que essa camada é mais expressiva, permitindo ajustes diferenciados que podem refletir diferentes caminhos para acomodar os dados. Por último, relacionada a Γ_3, podemos ver maior dispersão e assimetria, compatível com a complexidade de altas densidades. Observe que ao contrário da primeira camada, não há colapso único e, diferentemente da segunda, os modos não são tão bem definidos.

Em síntese, o *prior* empírico multivariado, aliado à decomposição de Cholesky, atua como ancoragem estatística inicial e garante coerência probabilística, enquanto o baixo peso relativo da D_{KL} permite afastamentos parcimoniosos quando respaldados pela redução de erro. O resultado é uma adaptação controlada, em que a flexibilidade para aprender correlações emerge progressivamente, reforçando a leitura de que a combinação entre PH e regularização bayesiana promove consistência sem sacrificar expressividade.

A Tabela 3.3 reporta REQM e EAM por saída no conjunto de dados de teste, consolidando a hierarquia de dificuldades já identificada pelas curvas de perda. Observa-se que a terceira saída apresenta erros superiores às demais. Em termos de escala, $\mathrm{REQM}_{\Gamma_3}/\mathrm{REQM}_{\Gamma_1} \approx 7{,}30$ e $\mathrm{EAM}_{\Gamma_3}/\mathrm{EAM}_{\Gamma_1} \approx 7{,}95$, caracterizam uma diferença próxima de uma ordem de grandeza entre os extremos.

Além disso, observa-se que a razão EAM/REQM cresce de forma monotônica entre os ramos ($\approx 0{,}75$, $0{,}77$, $0{,}82$). Esse aumento indica que, em Γ_3, os resíduos apresentam caudas relativamente mais pesadas, característica típica de cenários de maior variabilidade. Tal comportamento é precisamente o que motiva a adoção da PH, uma vez que essa função de perda reduz a influência de outliers e lida de maneira mais adequada com heteroce-

[19] Com N denotando o número total de amostras do conjunto de treinamento.

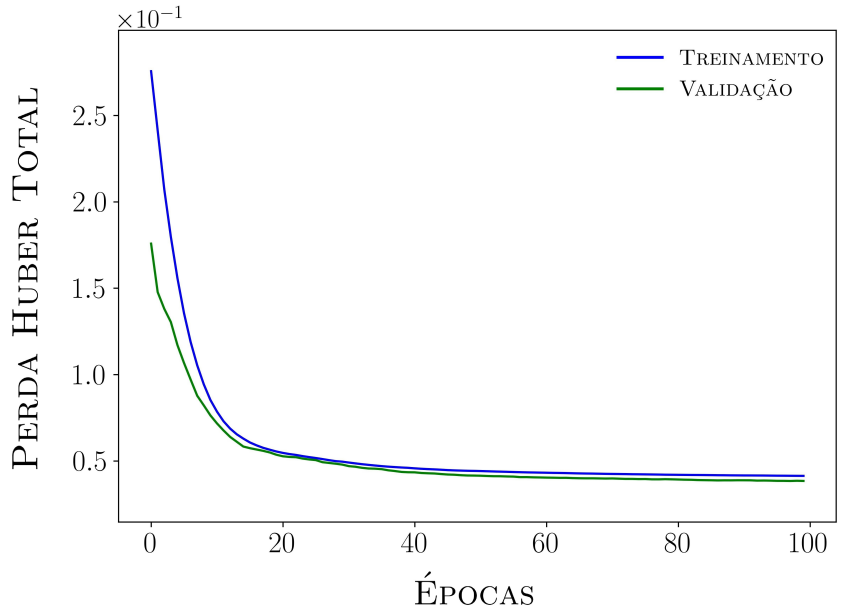


FIGURA 3.10 – Evolução da PH total ao longo das 100 épocas. Observa-se decaimento exponencial nas primeiras quinze épocas e convergência suave em torno de $3{,}8 \times 10^{-2}$. As curvas de treinamento (azul) e validação (verde) permanecem próximas, sugerindo boa generalização e ausência de sobreajuste significativo.

dasticidade, resultando em ajustes mais consistentes ao longo dos diferentes regimes de densidade.

Ao expressar o EAM como fração da amplitude do alvo Ω_j (Sec. 3.2), obtêm-se $p_1 \approx 0{,}88\,\%$, $p_2 \approx 2{,}37\,\%$ e $p_3 \approx 6{,}83\,\%$. Ou seja, mesmo no caso mais desafiador, o erro médio permanece bem abaixo de 10% da variabilidade observada, enquanto em Γ_1 fica abaixo de 1%. Em termos físicos, isso é consistente com o fato de que Γ_3 governa a região de alta densidade da EdE, mais pobremente informada por observáveis e mais sensível a pequenas variações estruturais, ao passo que Γ_1 (baixa densidade) é mais bem condicionado pelos dados disponíveis.

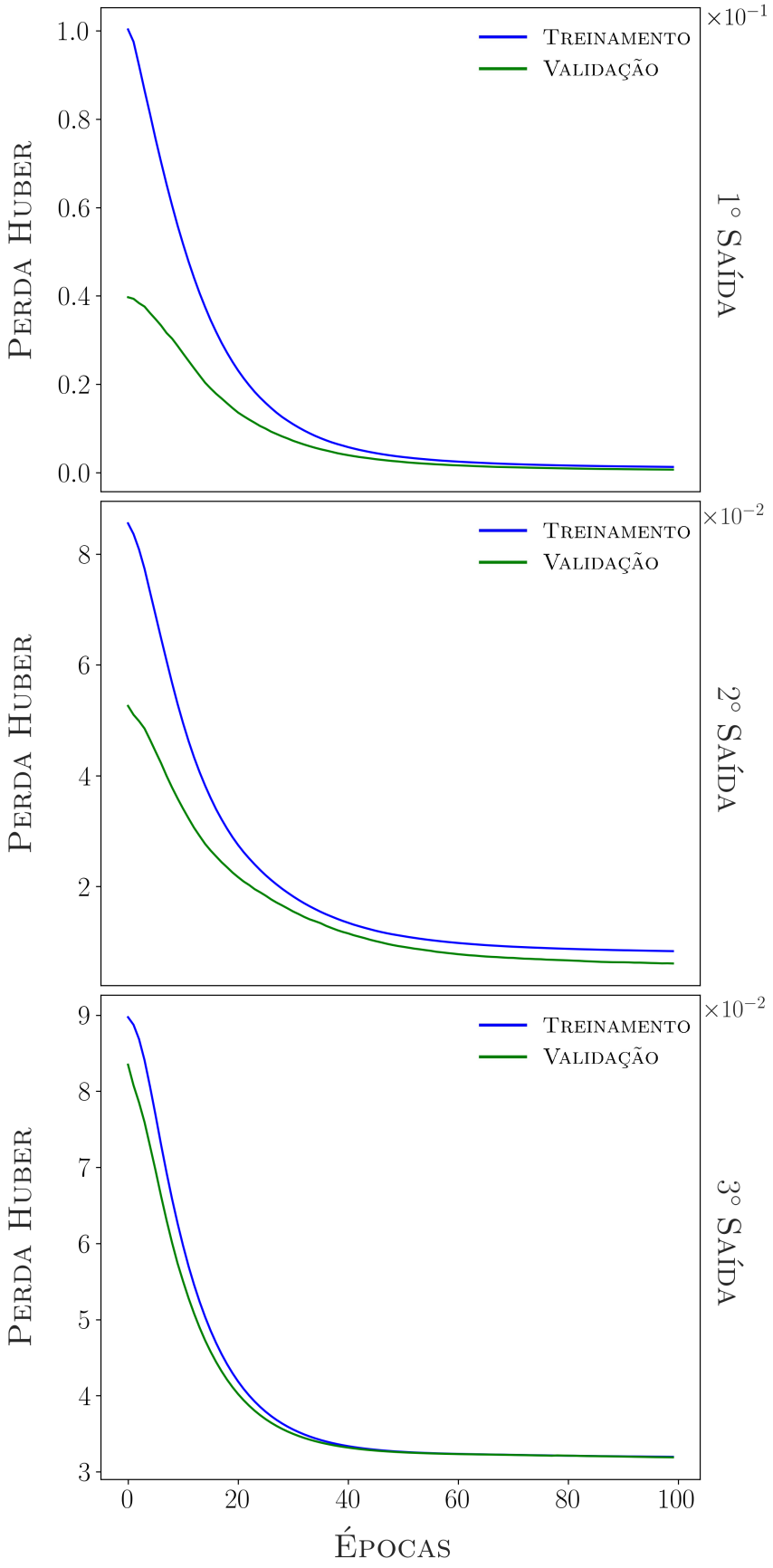


FIGURA 3.11 – PH média de cada saída ao longo das 100 épocas. O painel triplo exibe, respectivamente, o erro calculado referente a previsão dos parâmetros Γ_1, Γ_2 e Γ_3. As curvas de treinamento (azul) e validação (verde) convergem rapidamente, porém para patamares distintos com o seguintes valores aproximados: 10^{-3} (1º ramo), 10^{-2} (2º) e 3×10^{-2} (3º). As diferenças finais refletem a complexidade relativa de cada parâmetro.

Saída	Perda Final
Geral	0,038 034
1	0,000 591
2	0,005 895
3	0,031 529

TABELA 3.2 – PH na última iteração do treinamento. A linha Geral é a soma das perdas das três saídas. O erro residual da ordem de 10^{-4} a 10^{-3} no 1º/ 2º ramos e aproximadamente 10^{-2} no 3º ramo.

Saída	REQM	EAM
1	0,034 39	0,025 81
2	0,108 59	0,083 73
3	0,251 13	0,205 27

TABELA 3.3 – Métricas do modelo sobre o conjunto de dados de teste para cada saída. A escala relativa entre ramos é marcada: $\mathrm{REQM}_{\Gamma_3}/\mathrm{REQM}_{\Gamma_1} \sim 7{,}30$ e $\mathrm{EAM}_{\Gamma_3}/\mathrm{EAM}_{\Gamma_1} \sim 7{,}95$. As razões EAM/REQM ($\sim$ 0,75, 0,77, 0,82) indicam caudas residuais mais pesadas à medida que avançamos para os resultados atrelados ao parâmetro Γ_3 que pertence a região de maior densidade no modelo.

A Fig.(3.12) mostra o logaritmo decimal de α_j ao longo das épocas. As variações são deliberadamente modestas. Por exemplo, α_3 alcança $\sim 1{,}009$, o que corresponde a uma modulação multiplicativa de $\sim 0{,}9\%$ (com $\log_{10}(1{,}009) \sim 3{,}9 \times 10^{-3}$). Esse regime conservador evita instabilidades e, ainda assim, desloca o foco da otimização com maior eficiência em direção ao ramo que possui maior contribuição no erro. Em conjunto com a análise das perdas, o padrão observado sugere que valores ligeiramente maiores de λ poderiam acelerar o reequilíbrio entre ramos, ao custo potencial de maior oscilação. A escolha corrente privilegiou estabilidade global.

A trajetória de $\mathrm{D}_{\mathrm{KL},j}/N$ na Fig.(3.13) confirma que o desvio da posterior variacional em relação ao prior empírico permanece em nível residual (intervalo $\sim 7\times10^{-6}$ a $\sim 1{,}8\times10^{-5}$, um aumento de $\sim 2{,}6\times$ ao longo do treinamento), compatível com a leitura anterior de que o prior está bem calibrado. Esse comportamento indica que a regularização não domina

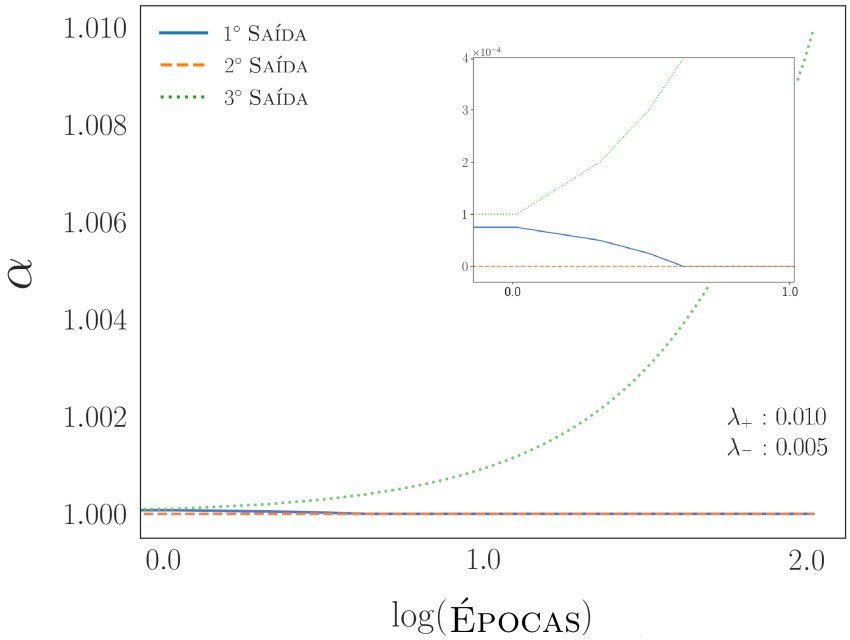


FIGURA 3.12 – Evolução do $\log_{10}$ do fator de modulação α_j por ramo ($j \in \{1, 2, 3\}$) ao longo de 100 épocas. Observa-se aumento progressivo de α_3, discreta redução de α_1 e estabilidade de α_2. A modulação é propositalmente pequena (variações $\lesssim 1\%$), garantindo redistribuição de gradientes sem comprometer a estabilidade do otimizador.

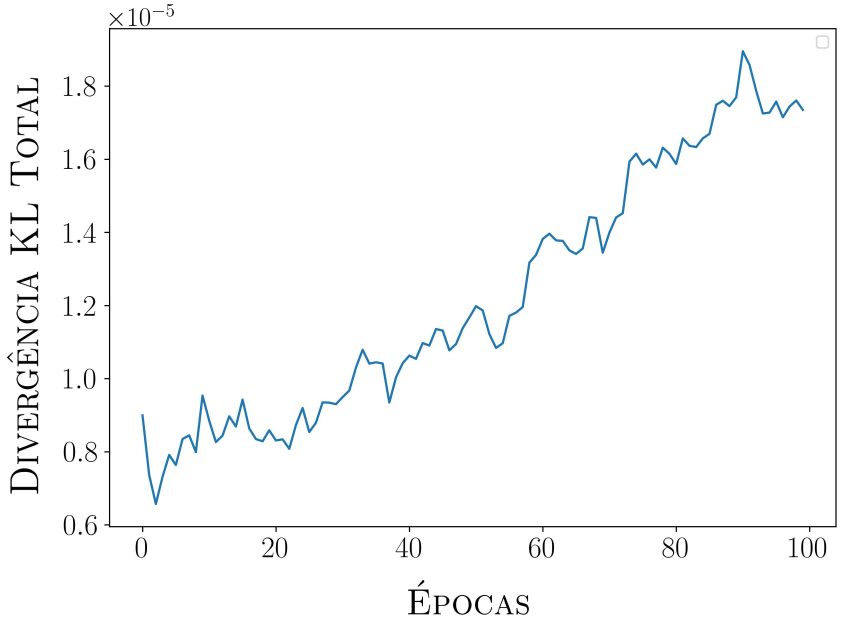


FIGURA 3.13 – Trajetória de $D_{KL,j}/N$ (posterior variacional vs. prior empírico) resultante da soma refente as três CVs ao longo de 100 épocas. Os valores iniciam em torno de $7{,}0 \times 10^{-6}$ e crescem gradualmente até aproximadamente $1{,}8 \times 10^{-5}$. Esse aumento lento e controlado indica que o custo de complexidade se mantém baixo, enquanto a rede realiza ajustes consistentes ao prior, sem sinais de instabilidade ou divergência abrupta.

a função objetivo nem inibe o ajuste induzido pela Perda Huber.

Os histogramas das distribuições a posteriori dos pesos na Fig.(3.14) sintetizam como cada CV utiliza o orçamento de complexidade. A CV de Γ_1 permanece concentrada em torno de zero, com variância decrescente, o que caracteriza uma tarefa mais estável. Já a CV de Γ_2 apresenta um leve deslocamento e sinais de bimodalidade, sugerindo maior flexibilidade. Por fim, a CV de Γ_3 evidencia assimetria e alargamento, indicando uma tarefa mais custosa e heterogênea. Esse gradiente de complexidade se mostra coerente com as métricas da Tabela 3.3 e com a dinâmica dos α_j exibida na Fig.(3.12). Do ponto de vista prático, duas extensões podem ser consideradas para o ramo de Γ_3: ampliar moderadamente a largura da CV, aumentando o número de unidades ou permitindo covariâncias mais expressivas na posterior variacional; e, de forma complementar, adotar priors com caudas mais pesadas caso a análise de resíduos confirme a persistência de assimetrias, sempre ponderando os impactos sobre a estabilidade.

3.5 Aplicações e Cenários de Teste

Nesta seção avaliamos a aplicabilidade do modelo treinado em dois cenários distintos, enfatizando que se trata de um preditor não determinístico. Diferentemente de abordagens puramente pontuais, exploramos explicitamente a incerteza das previsões induzida pela distribuição posterior (variacional) dos pesos. Consequentemente, além das estimativas centrais, reportamos intervalos de confiança (IC). Tal caracterização é essencial para julgar a robustez e a capacidade de generalização do modelo em usos astrofísicos, nos quais a escassez e a heterogeneidade observacionais são a regra.

Para cada vetor de entrada $\mathbf{x} \in \mathbb{R}^{12}$, realizamos $n \in \mathbb{N}_{>0}$ amostragens da rede ao longo da posterior variacional dos parâmetros. Em cada amostra i, extraímos $\theta^{(i)} \sim q(\theta)$ e avaliamos

$$y_i \;=\; f_{\theta^{(i)}}(\mathbf{x}), \qquad i = 1, \ldots, n, \tag{3.65}$$

onde $f_\theta(\cdot)$ denota a rede parametrizada. No caso vetorial, isto é, para $\Gamma_{i=1,2,3}$, cada amostra é registrada diretamente como um vetor $\mathbf{y}_i \in \mathbb{R}^3$. Essa representação conjunta preserva a estrutura estatística entre os parâmetros e, além disso, todos são necessários para compor a EdE.

A partir do conjunto de amostras $\{y_i\}_{i=1}^n$, calculamos a média preditiva,

$$\bar{y} \;=\; \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{3.66}$$

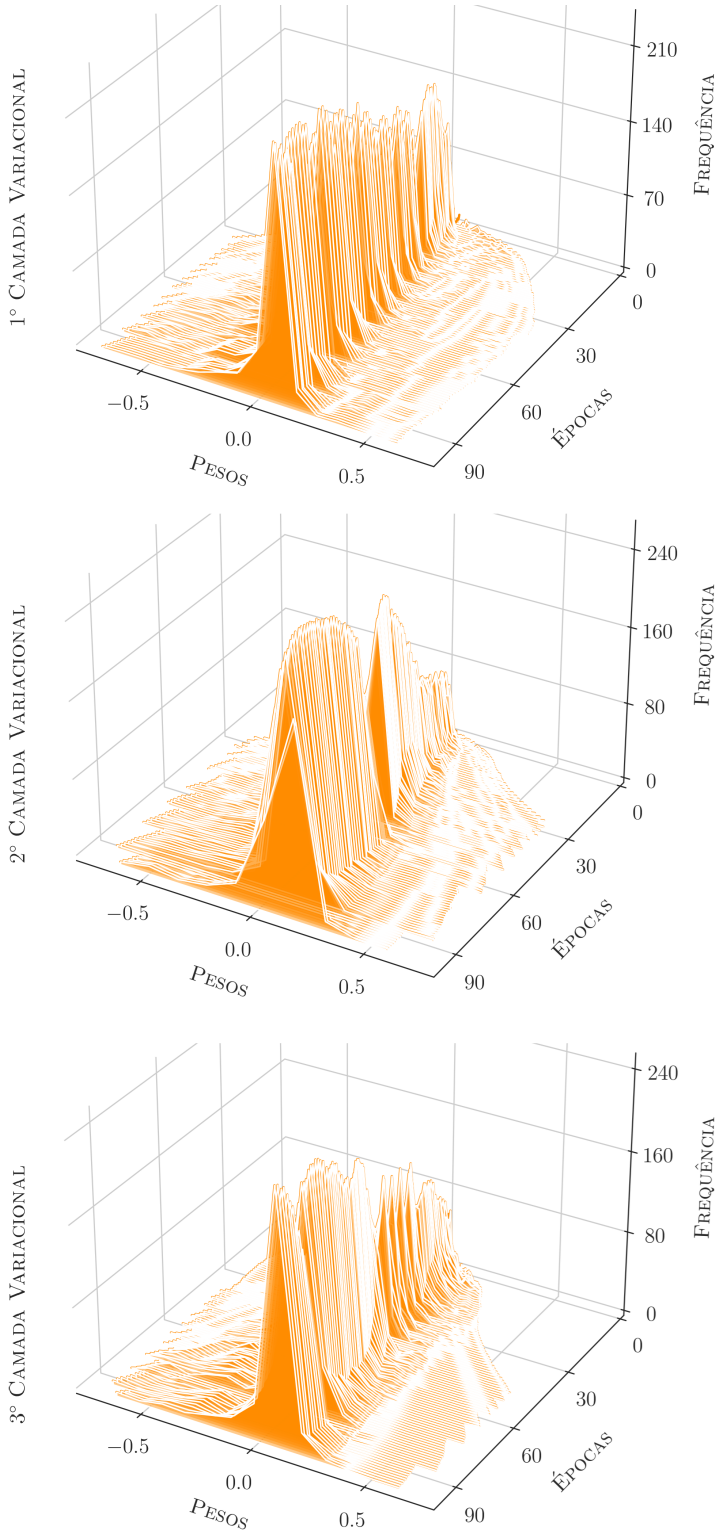


FIGURA 3.14 – Distribuições a posteriori dos pesos por CV ao longo do treinamento. 1ª CV: pico centrado em zero e variância decrescente; 2ª CV: deslocamento moderado e leve bimodalidade; 3ª CV: distribuição mais larga e assimétrica, com cauda direita pronunciada. Os padrões refletem a realocação de capacidade preditiva de acordo com a dificuldade de cada ramo e o papel moderador de D_{KL}.

bem como o desvio-padrão epistêmico,

$$\sigma_{\text{epi}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}. \tag{3.67}$$

No caso vetorial, as quantidades generalizam-se componente a componente, de modo a obter dispersões epistêmicas marginais para cada parâmetro Γ_i.

A incerteza epistêmica preditiva pode ser reportada por meio dos quantis empíricos da distribuição amostral:

$$\text{IC}^{\text{pred}}_{1-\alpha} \;=\; \big[\, Q_{\alpha/2}(\{y_i\}),\; Q_{1-\alpha/2}(\{y_i\}) \,\big], \tag{3.68}$$

o que reflete diretamente a variabilidade entre amostras da posterior. Quando a distribuição é aproximadamente simétrica, esse intervalo pode ser bem aproximado por $\bar{y} \pm z_{1-\alpha/2}\,\sigma_{\text{epi}}$.

As Eqs. (3.67) e (3.68) capturam a incerteza epistêmica, relacionada ao desconhecimento sobre os parâmetros θ. Não consideramos incerteza aleatória de observação, de modo que a largura dos intervalos de confiança aqui apresentados depende exclusivamente da dispersão entre os pesos amostrados.

Com esses elementos, torna-se possível avaliar a qualidade pontual das predições $\bar{y}^{(j)}$, reportar intervalos de confiança consistentes com a posterior variacional, analisar a estabilidade do modelo comparando larguras de intervalos entre diferentes cenários e saídas, e, de forma opcional, verificar a calibração por cobertura empírica. Neste último caso, compara-se a fração de pontos contidos nos intervalos de confiança com o nível nominal $1-\alpha$, podendo-se ainda utilizar métricas como PICP (taxa de cobertura) e MIW (largura média do intervalo) para caracterizar quantitativamente a qualidade dos intervalos preditivos.

3.5.1 Cenário de Validação

Como primeira verificação, avaliamos a capacidade da Rede Neural Bayesiana de recuperar, em um ambiente controlado, uma equação de estado de referência EdE_τ não utilizada no treinamento. O propósito deste cenário não é testar generalização sobre várias EdEs, mas sim aferir a precisão das predições médias em relação a uma curva conhecida e examinar a consistência das incertezas epistêmicas inferidas a partir da distribuição posterior variacional.

A EdE_τ foi resolvida via TOV e, a partir de sua curva massa-raio M-R_τ, selecionamos 6 pontos (M, R) de referência. Esses pares foram utilizados como entrada para a rede, que

então gerou 10^3 amostras preditivas por reamostragem dos pesos variacionais. As médias resultantes dessas amostras definem a equação de estado predita EdE_μ e a curva massa-raio correspondente M-R_μ. A Fig.(3.15) mostra que a curva média M-R_μ acompanha de forma consistente a referência M-R_τ, enquanto a Fig.(3.16) evidencia que a EdE_μ reproduz a EdE_τ com elevada fidelidade. Em ambos os casos, as regiões de incerteza são estreitas, indicando boa calibração epistêmica e confirmando que a dispersão observada provém majoritariamente da variabilidade do modelo.

Para quantificação detalhada, avaliamos pressões em três densidades características ρ_i $(i = 1, 2, 3)$, considerando a variável adimensional

$$\ell_i \;\equiv\; \log_{10}\Big(\frac{p(\rho_i)}{1\,\mathrm{MeV\,fm^{-3}}}\Big). \tag{3.69}$$

Os valores de referência foram $\ell_\tau = [\,1{,}248,\ 2{,}106,\ 3{,}243\,]$. As médias preditivas obtidas foram $\ell_\mu = [\,1{,}237,\ 2{,}108,\ 3{,}235\,]$, com dispersões $\sigma_\mu = [\,0{,}010,\ 0{,}042,\ 0{,}081\,]$. Convertendo novamente para o domínio linear, os erros relativos foram $\delta = [\,2{,}43\%,\ 0{,}36\%,\ 1{,}84\%\,]$, com maior discrepância em baixas densidades, como esperado.

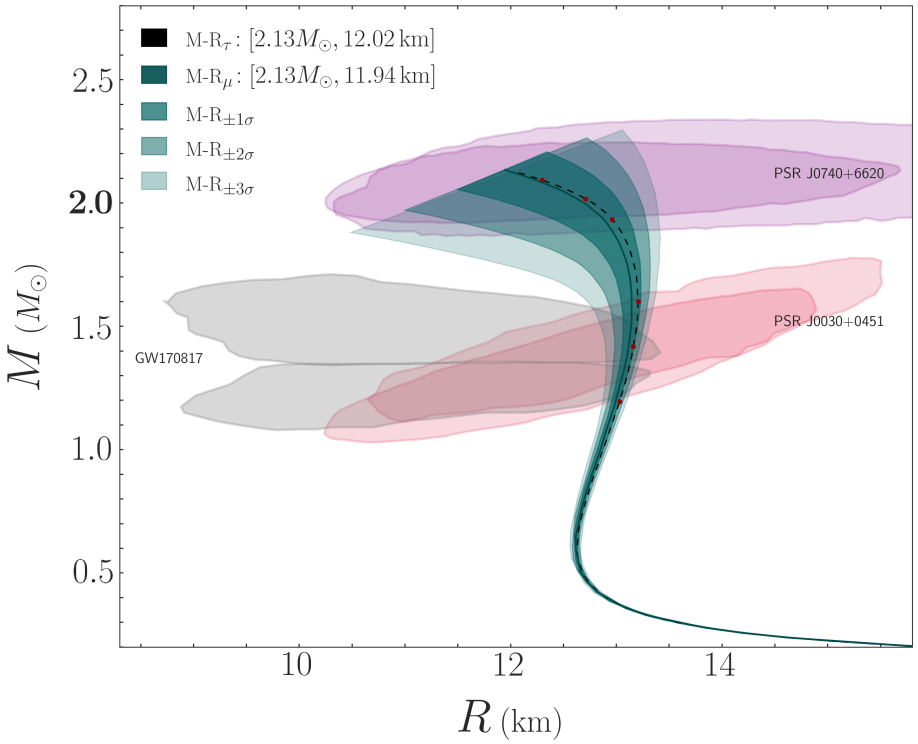


FIGURA 3.15 – Diagrama massa-raio no cenário de validação. Pontos vermelhos: pares (M, R) de referência obtidos da curva M-R$_\tau$ (solução TOV da equação de estado de teste). Linha sólida: curva predita M-R$_\mu$ a partir da EdE média EdE$_\mu$. Regiões em ciano: incertezas epistêmicas correspondentes às amostras da rede.

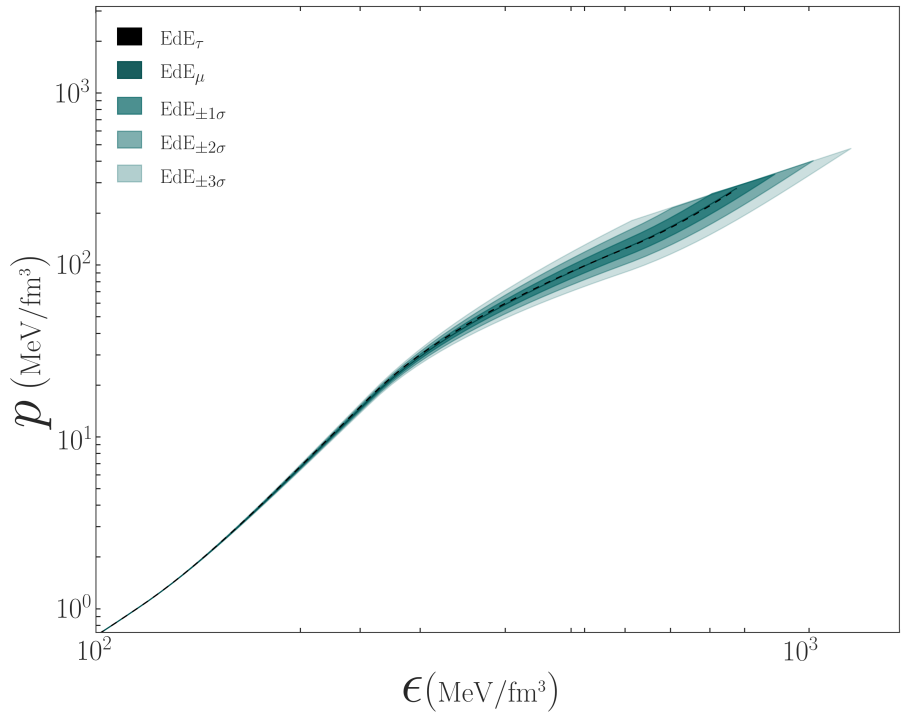


FIGURA 3.16 – Equação de estado média EdE$_\mu$ obtida das 10^3 amostras preditivas da rede. Em ciano, as regiões de incerteza epistêmica associadas à variabilidade entre amostras. Em preto, a equação de estado de referência EdE$_\tau$. Nota-se sobreposição quase completa entre as curvas, com faixas estreitas mesmo em altas densidades de energia.

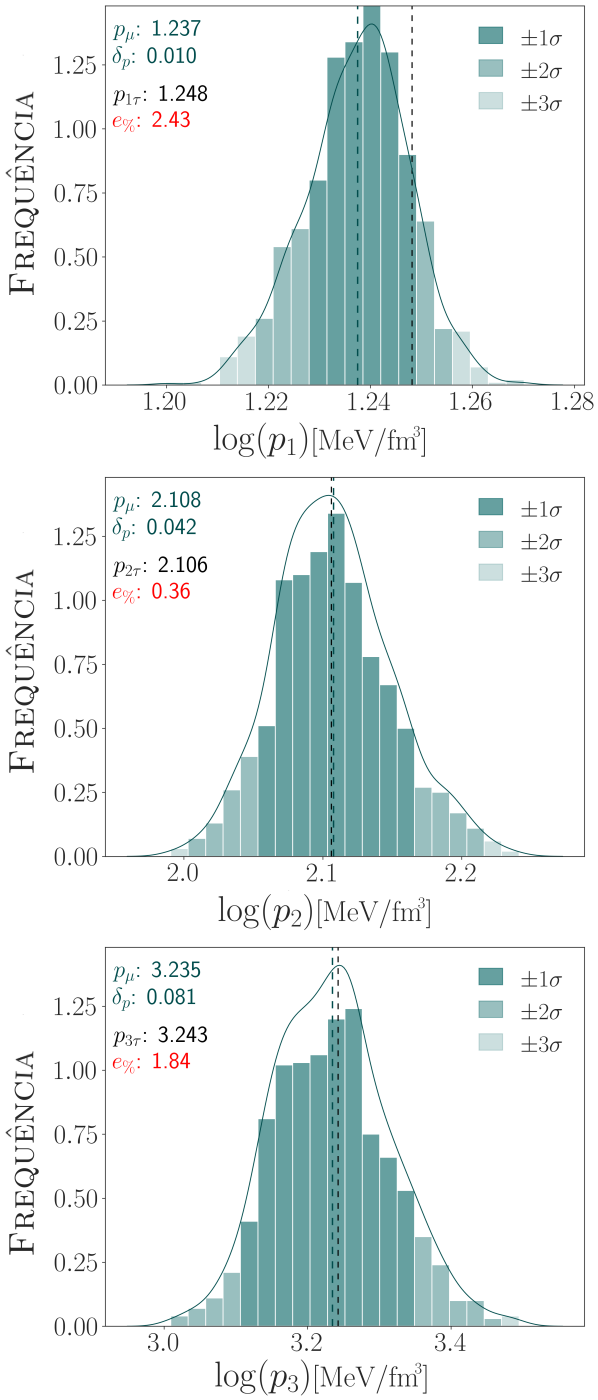


FIGURA 3.17 – Distribuições das 10^3 amostras preditivas para os três pontos característicos da EdE no cenário de validação. Linhas tracejadas: valores de referência ℓ_τ. Linhas sólidas: médias preditivas ℓ_μ. Faixas sombreadas: dispersões σ_μ. Em vermelho, erros relativos no domínio linear (2,43%, 0,36%, 1,84%). É importante notar que o valor final

Em síntese, o teste de validação demonstra que a rede recupera a EdE de referência com alta precisão e incertezas epistêmicas bem calibradas. A sobreposição entre curvas médias e alvos, associada à estreiteza das distribuições preditivas, confirma a robustez do modelo neste cenário. Ressaltamos, contudo, que os resultados se baseiam em seis

pontos extraídos de uma única EdE e não incluem ruído observacional, sendo portanto condicionados à classe de modelos PPG e ao mapeamento

$$(M, R)^{\times 6} \longmapsto (\Gamma_{i=1,2,3}) \longmapsto \text{EdE} \longmapsto \text{TOV}.$$

3.5.2 Cenário Realista

Avançando para um teste ancorado em dados observacionais, avaliamos a RNB sob um cenário realista construído a partir das medidas do NICER para os pulsares J0740+6620 e J0030+0451 (Tab. 3.4). O objetivo é verificar a consistência da predição média em torno de pares observados (M, R) e a calibração epistêmica das incertezas quando a informação física é mais escassa e potencialmente correlacionada.

Para cada pulsar, geramos outros dois pares de (M, R) adicionando ruído gaussiano da mesma ordem dos utilizados no treinamento, totalizando 6 pares que alimentam a RNB,

$$(M', R') = (M, R) + (\Delta M, \Delta R), \qquad (\Delta M, \Delta R) \sim \mathcal{N}\big(\mathbf{0},\, \mathrm{diag}(\sigma_M^2, \sigma_R^2)\big),$$

totalizando 6 pares que alimentam a RNB.

PSR	**Massa ($M_\odot$)**	**Raio (km)**
J0740+6620	2,072	12,39
J0030+0451	1,34	12,71

TABELA 3.4 – Pares M-R do NICER utilizados na montagem do cenário realista.

Amostramos 10^3 predições do pósterior variacional dos pesos. A equação de estado média EdE_μ e a respectiva solução TOV M-R_μ estão dispostas nas Figs.(3.18, 3.19). Com esse número de amostras, a variância da média é inferior a 0,1% da variância epistêmica, de modo que a largura dos intervalos exibidos é dominada pela incerteza do modelo, como no cenário de validação.

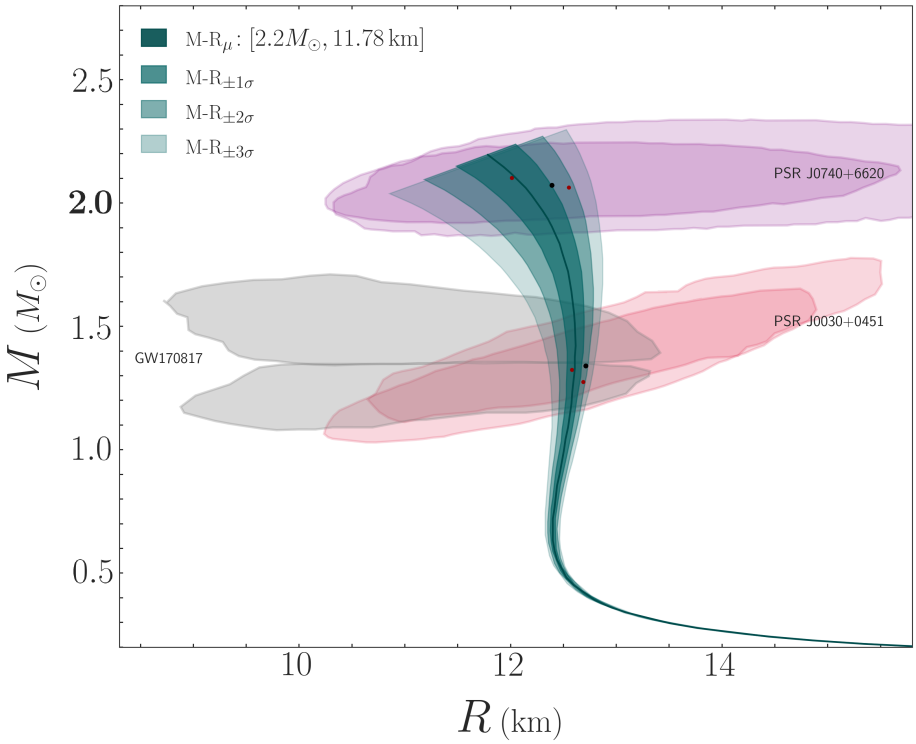


FIGURA 3.18 – Diagrama massa-raio para o cenário realista. Pontos vermelhos: pares (M, R) gerados a partir de ruído gaussiano sobre os dados do NICER (pontos pretos). Curva sólida: solução TOV obtida da EdE média EdE$_\mu$. Faixas ciano: intervalos epistêmicos induzidos pela amostragem dos pesos.

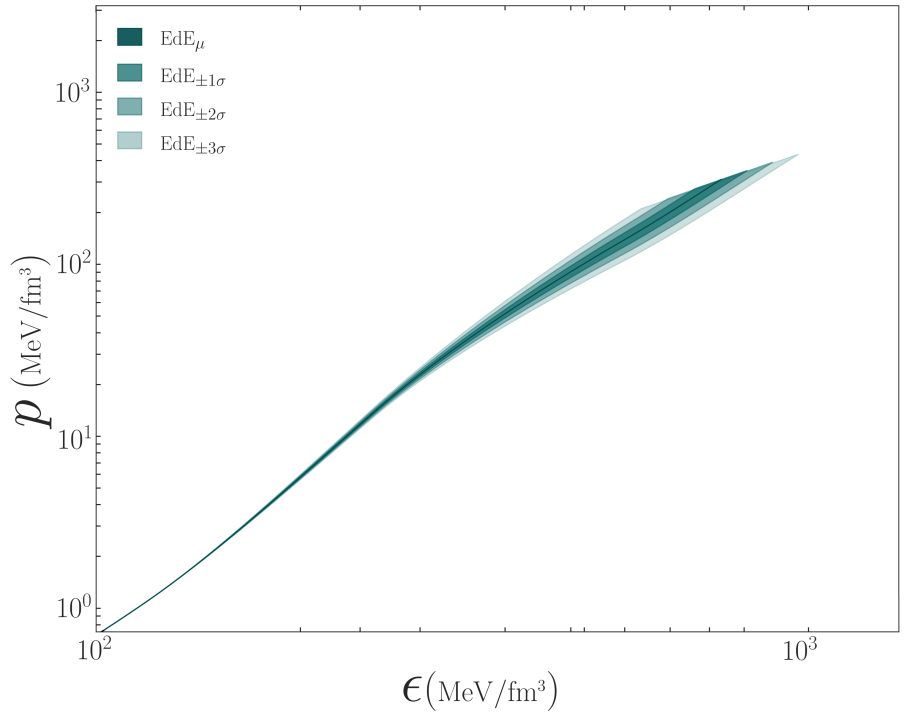


FIGURA 3.19 – Equação de estado média EdE$_\mu$ resultante das 10^3 amostras preditivas. Faixas ciano: incertezas epistêmicas correspondentes.

A Fig.(3.18) mostra que a curva média M-R$_\mu$ interpola suavemente os pares observados (pretos) e os sintéticos (vermelhos), mantendo largura radial típica moderada: $\lesssim 1{,}5\,\mathrm{km}$ para $M \gtrsim 1{,}8\,M_\odot$. A faixa 2σ cobre os seis pares utilizados e intersecta as regiões credíveis do NICER, com J0740+6620 situado entre 1σ e 2σ e J0030+0451 tipicamente dentro de 2σ. Incluímos também a região de credibilidade de 90% de GW170817 (cinza), a qual é cruzada pela faixa 2σ de M-R$_\mu$ em $M \sim 1{,}2$-$1{,}6\,M_\odot$ e $R \sim 11$-$13\,\mathrm{km}$, servindo como checagem independente.

Na Fig.(3.19), a EdE$_\mu$ apresenta bandas estreitas no intervalo de ϵ que domina os raios entre $\sim 1{,}3$ e $\sim 2{,}0\,M_\odot$. Para leitura consistente, os eixos são interpretados como $\log_{10}[\epsilon/(\mathrm{MeV\,fm^{-3}})]$ e $\log_{10}[p/(\mathrm{MeV\,fm^{-3}})]$, com ticks indicando o intervalo $\sim 10^2$–10^3 $\mathrm{MeV\,fm^{-3}}$

Os histogramas da Fig.(3.20) são aproximadamente gaussianos e simétricos, com médias e respectivo desvio-padrão para sobre os valore previstos para pressão:

$$\ell_\mu = \big[\,1{,}130,\ 2{,}145,\ 3{,}335\,\big] \quad \text{e} \quad \sigma_\mu = \big[\,0{,}009,\ 0{,}036,\ 0{,}059\,\big]. \tag{3.70}$$

Dois pontos merecem esclarecimento a fim de evitar interpretações equivocadas sobre os resultados numéricos e gráficos obtidos nos cenários analisados com a RNB.

Primeiro, é essencial ter em mente que o modelo não prevê diretamente pressões p_μ ou seus correspondentes ℓ_μ. O objetivo da rede é fornecer os parâmetros Γ_i associados às DIs da EdE. A partir desses parâmetros é que se calculam os valores de pressão apresentados, como os que aparecem nos histogramas.

Em segundo lugar, deve-se notar que certos valores de pressão, como p_3 exibido na Fig.(3.20), não aparecem nas curvas da EdE mostradas na Fig.(3.19). Esse efeito decorre do critério de **causalidade**, aplicado no processo de construção das equações de estado (visto através dos gráficos) e de integração das TOV, onde os valores são **truncados** quando o limite causal é atingido.

Comparado ao cenário de validação, nota-se leve redistribuição: ℓ_1 tende a valores menores, enquanto ℓ_2 e ℓ_3 crescem modestamente. Esse padrão é plausível, pois raios em $M \sim 1{,}3$-$1{,}6\,M_\odot$ respondem a pressões em $\rho \sim (1\text{-}2)\,n_0$, ao passo que a sustentação de $M_{\max} \gtrsim 2{,}0\,M_\odot$ exige maior rigidez em densidades mais altas.

Em síntese, o cenário realista indica que a RNB acomoda os alvos do NICER com incertezas epistêmicas moderadas, preserva a coerência física da EdE$_\mu$ e mantém compatibilidade com GW170817. A interseção sistemática de M-R$_\mu$ com as regiões credíveis do NICER, a largura radial $\lesssim 1{,}5\,\mathrm{km}$ em $M \gtrsim 1{,}8\,M_\odot$ e a estrutura quase-gaussiana das distribuições ℓ_i confirmam boa calibração epistêmica neste regime observacional.

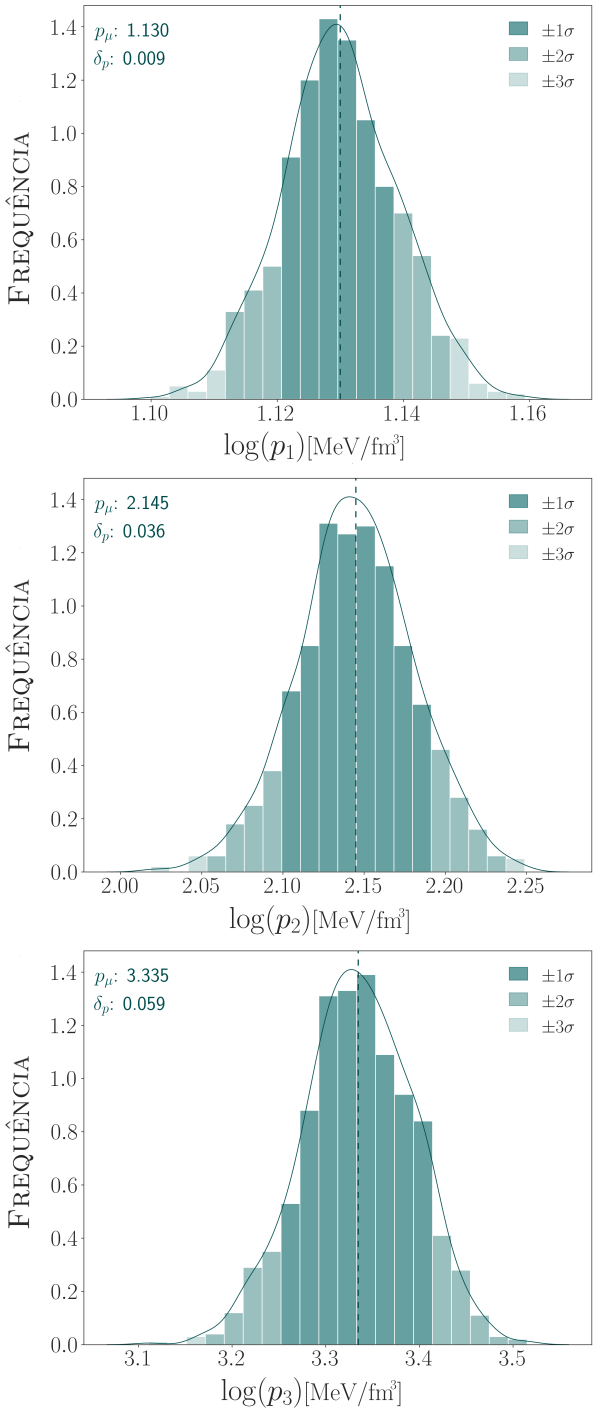


FIGURA 3.20 – Distribuições das 10^3 predições de $\log p$ em três densidades características ρ_i. Linhas sólidas: médias preditivas ℓ_μ; faixas sombreadas: dispersões σ_μ. A variância epistêmica típica $< 3\%$ corrobora as bandas estreitas vistas em M-R.

4 Conclusões

A principal contribuição deste trabalho foi demonstrar que a recuperação dos parâmetros de equações de estado para estrelas de nêutrons pode ser conduzida de forma confiável quando o problema é tratado como um sistema integrado, no qual a geração de dados, a preparação estatística e a modelagem bayesiana atuam de maneira coerente. Para viabilizar o treinamento da rede, tornou-se necessário construir um conjunto de pares (M, R) e respectivos parâmetros da equação de estado. Esse conjunto foi obtido a partir do formalismo politrópico por partes generalizado, garantindo continuidade de $p(\varepsilon)$, $dp/d\varepsilon$, diferenciabilidade, estabilidade na integração das equações de TOV. Essa base assegurou que o treinamento da rede fosse alimentado por exemplos fisicamente consistentes e estatisticamente bem estruturados.

A preparação do banco de dados preservou a geometria do problema M-R e distribuições estatísticas básicas. A amostragem repetida de pares (M, R), a ordenação canônica para evitar ambiguidades de permutação, a divisão estratificada e a injeção seletiva de ruído no treino aumentaram a robustez sem induzir viés. O resultado foi um *dataframe* com quase cinco milhões de entradas, equilibrado e reprodutível.

Sobre esse alicerce assentou-se uma Rede Neural Bayesiana enxuta, com apenas 377 parâmetros variacionais por ramo e um total de ~ 1164 parâmetros treináveis. As escolhas arquiteturais - ativações consistentes com a positividade de pressões, priors informativos ancorados em estatísticas físicas, perda de Huber para lidar com caudas pesadas e regularização KL controlada - permitiram alcançar estabilidade rápida no treinamento e prevenção de sobreajuste.

Os resultados obtidos organizam-se em três níveis:

- **Cenário controlado**: o erro absoluto médio, como fração da amplitude do alvo, ficou abaixo de 1 % em Γ_1, em torno de $\sim 2{,}4$ % em Γ_2 e $\sim 6{,}8$ % em Γ_3. O erro amostral da média foi residual, confirmando que a dispersão é dominada pela incerteza do modelo.
- **Cenário de validação**: com seis pares (M, R) de uma EdE de referência, observamos sobreposição quase completa entre EdE_μ e EdE_τ, bem como entre $\mathrm{M\text{-}R}_\mu$ e $\mathrm{M\text{-}R}_\tau$. Os erros relativos nos três pontos de pressão foram (2,43, 0,36, 1,84) %, e a largura típica

em R foi de ~ 1 km, atestando alta precisão e consistência epistêmica.

- **Cenário realista**: ancorado nos pulsares J0740+6620 e J0030+0451 (NICER), a curva média M-R_μ interpolou suavemente os pares observados e sintéticos, manteve largura radial $\lesssim 1,5$ km para $M \gtrsim 1,8\,M_\odot$ e intersectou tanto as regiões credíveis do NICER quanto o contorno de 90% de GW170817. No espaço pressão-densidade, as médias de $\log p$ foram [1,130, 2,145, 3,335] com dispersões menores que 3% da amplitude, revelando coerência física e compatibilidade multimensageira.

Em conjunto, esses três testes evidenciam que a RNB não apenas recupera equações de estado de referência em ambientes controlados, mas também acomoda medidas reais com incertezas moderadas e bem calibradas.

Algumas limitações foram identificadas. A saída independente por parâmetro pode subestimar correlações entre $(\Gamma_1, \Gamma_2, \Gamma_3)$; o limiar da perda de Huber interage com a normalização e merece ajuste fino; e as incertezas reportadas são estritamente epistêmicas, pois ruído observacional não foi explicitamente modelado. Esses pontos abrem caminhos claros para avanços: cabeças conjuntas com covariância plena, variacionais mais expressivas no regime denso para modular a pressão da KL e inclusão de heterocedasticidade para compor incertezas epistêmicas e aleatórias.

4.1 Considerações finais

O que se consolida aqui é um método: integrar física, dados e inferência em um quadro coerente gera ganhos que nenhuma etapa isolada alcança. Um formalismo físico simples e eficaz para geração de dados que respeita restrições, um *dataprep* que evita vazamentos e um preditor bayesiano parcimonioso e explícito na quantificação de incerteza foram suficientes para reconstruir a relação pressão-densidade, aderir às medidas do NICER, manter consistência com GW170817 e sustentar erros muito abaixo de 10% mesmo em condições desafiadoras.

O *framework* permanece modular e extensível. Pode incorporar ruído aleatório na saída para intervalos completos, explorar perdas alternativas, enriquecer a posterior em altas densidades, testar sensibilidades a ρ_0 e ao modelo de crosta, além de utilizar ruído correlacionado para avaliação formal de cobertura. Nada disso altera o essencial: o modelo mantém coerência e nível de precisão com o esperado, e também quantifica o quanto sabe - e onde precisa aprender. É justamente nas altas densidades, onde a incerteza epistêmica se alarga, que arquiteturas conjuntas, variacionais mais ricas e composição epistêmico+aleatório prometem os maiores dividendos científicos. Com esses incrementos, o caminho para análises integradas NICER-LIGO-Virgo e para observatórios de raios X de próxima geração fica não só aberto, como também tecnicamente bem fundamentado.

Referências

ABADI, M. *et al.* **Tensorflow**: large-scale machine learning on heterogeneous distributed systems. [S.l.]: USENIX Association, 2016. p. 265–283. ISBN 9781931971331. 71

ABBOTT, B. *et al.* GW170817: observation of gravitational waves from a binary neutron star inspiral. **Phys. Rev. Lett.**, **American Physical Society**, v. 119, n. 16, p. 161101, Oct 2017. 9, 51, 53

ALFORD, M. G.; SCHMITT, A.; RAJAGOPAL, K.; SCHÄFER, T. Quark matter in compact stars. **Reviews of Modern Physics**, **American Physical Society**, v. 80, n. 4, p. 1455–1515, 2008. 19

ANTONIADIS, J. *et al.* A massive pulsar in a compact relativistic binary. **Science**, v. 340, p. 6131, 2013. 20

BAADE, W.; ZWICKY, F. Cosmic rays from super-novae. **Proceedings of the National Academy of Sciences**, v. 20, n. 5, p. 259–263, 1934. 22

BAADE, W.; ZWICKY, F. On super-novae. **Proceedings of the National Academy of Science**, v. 20, n. 5, p. 254–259, may 1934. 22

BAYM, G.; FURUSAWA, S.; HATSUDA, T.; KOJO, T.; TOGASHI, H. New neutron star equation of state with quark-hadron crossover. **The Astrophysical Journal**, v. 885, n. 1, p. 42, 2019. 44

BISHOP, C. **Neural networks for pattern recognition**. USA: Oxford University Press, Inc., 1995. ISBN 0198538642. 57

BLEI, D.; KUCUKELBIR, A.; MCAULIFFE, J. Variational inference: a review for statisticians. **Journal of the American Statistical Association**, v. 112, n. 518, p. 859–877, 2017. 76, 78

BLUNDELL, C. *et al.* Weight uncertainty in neural networks. *In*: PROCEEDINGS OF THE 32nd INTERNATIONAL CONFERENCE ON MACHINE LEARNING. 32., 2015. **Proceedings** [...]. [S.l.]: PMLR, 2015. v. 37, p. 1613–1622. 75

CALLEN, H. B. **Thermodynamics and an Introduction to thermostatistics**. 2. ed. New York: John Wiley & Sons, 1985. ISBN 978-0471862567. 36

CHANDRASEKHAR, S. The maximum mass of ideal white dwarfs. **apj**, v. 74, p. 81, Jul 1931. 24

CLAYTON, D. **Principles of stellar evolution and nucleosynthesis**. [S.l.]: University of Chicago Press, 1983. ISBN 0226109526. 24

CLAYTON, D. **Principles of stellar evolution and nucleosynthesis**. [S.l.]: University of Chicago Press, 1983. ISBN 0226109526. 24

COONEY, A.; DEDEO, S.; PSALTIS, D. Neutron stars in $f(r)$ gravity with perturbative constraints. **Phys. Rev. D**, **American Physical Society**, v. 82, p. 064033, Sep 2010. 20

COVER, T. M.; THOMAS, J. A. **Elements of information theory**. London: Wiley-Interscience, 2006. ISBN 978-0471241959. 76

DEMOREST, P. *et al.* A two-solar-mass neutron star measured using shapiro delay. **Nature**, v. 467, p. 1081–1083, 2010. 20

DILLON, J. *et al.* **Tensorflow distributions**. [S.l.]: Wiley, 11 2017. 71

DOUCHIN, F.; HAENSEL, P. A unified equation of state of dense matter and neutron star structure. **Astron. Astrophys.**, v. 380, p. 151, 2001. 44

EINSTEIN, A. Die feldgleichungen der gravitation. **Sitzungsberichte der Preussischen Akademie der Wissenschaften**, Berlin, p. 844–847, Nov 1915. 26

FILHO, K. de S. O.; SARAIVA, M. de F. O. **Astronomia e astrofísica**. São Paulo: Livraria da Física, 2014. 23

FUJIMOTO, Y.; FUKUSHIMA, K.; MURASE, K. Extensive studies of the neutron star equation of state from the deep inference with the observational data augmentation. **Journal of High Energy Physics**, v. 03, p. 1–46, 2021. 20

GEORGE, D.; HUERTA, E. A. Deep neural networks to enable real-time multimessenger astrophysics. **Phys. Rev. D, American Physical Society**, v. 97, p. 044039, Feb 2018. 20

GOLD, T. Rotating neutron stars as the origin of the pulsating radio sources. **Nature**, v. 218, p. 731–732, 1968. 23

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learnind**. London: MIT Press, 2016. ISBN 978-0262035613. 59

GREIF, S. K.; RAAIJMAKERS, G.; HEBELER, K.; SCHWENK, A.; WATTS, A. L. Equation of state sensitivities when inferring neutron star and dense matter properties. **Monthly Notices of the Royal Astronomical Society**, v. 485, n. 4, p. 5363–5376, 03 2019. ISSN 0035-8711. 50

HEBB, D. **The organization of behavior**: a neuropsychological theory. New York: Wiley, 1949. (A Wiley book in clinical psychology). ISBN 9780471367277. 58

HEBELER, K. *et al.* Equation of state and neutron star properties constrained by nuclear physics and observation. **The Astrophysical Journal**, v. 773, n. 1, p. 11-14, aug 2013. 50

HEWISH, A.; BELL, S.; PILKINGTON, J.; SCOTT, P.; COLLINS, R. Observation of a rapidly pulsating radio source. **Nature**, v. 217, n. 5130, p. 709–713, 1968. 23

HOFFMAN, M. *et al.* Stochastic variational inference. **Journal of Machine Learning Research**, v. 14, n. 1, p. 1303–1347, 2013. 78

HOPFIELD, J. Neural networks and physical systems with emergent collective computational abilities. **Proceedings of the National Academy of Sciences of the United States of America**, v. 79, n. 8, p. 2554–2558, Apr 1982. 59

HUBER, P. Robust estimation of a location parameter. **The Annals of Mathematical Statistics**, v. 35, n. 1, p. 73–101, 1964. 81

JORDAN, M. I.; GHAHRAMANI, Z.; JAAKKOLA, T. S.; SAUL, L. K. An introduction to variational methods for graphical models. In: JORDAN, M. I. (Ed.). **Learning in Graphical Models**. Dordrecht: Kluwer Academic Publishers, 1999. p. 105–161. 74

KINGMA, D. P.; WELLING, M. Auto-encoding variational bayes. *In*: PROCEEDINGS OF THE 2nd INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS (ICLR), 2014. **Proceedings** [...]. [S.l], 2014. 75

KULLBACK, S.; LEIBLER, R. On information and sufficiency. **The Annals of Mathematical Statistics**, v. 22, n. 1, p. 79–86, 1951. 76

LANDAU, L.; LIFSHITZ, E. **Fluid mechanics**. 2. ed. New York: Pergamon Press, 1987. v. 6. (Course of Theoretical Physics, v. 6). ISBN 0080339328. 37

LATTIMER, J. M. Constraints on neutron star radii. **Annual Review of Nuclear and Particle Science**, v. 62, p. 485–515, 2012. 19

LATTIMER, J. M.; PRAKASH, M. Neutron star structure and the equation of state. **The Astrophysical Journal**, v. 550, n. 1, p. 426, 2001. 19

LOSHCHILOV, I.; HUTTER, F. Decoupled weight decay regularization. *In*: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS (ICLR), 2019. New Orleans. **Proceedings** [...]. New Orleans: ICLR, 6-9 May, 2019. 84

MACKAY, D. A practical bayesian framework for backpropagation networks. **Neural Computation**, v. 4, n. 3, p. 448–472, 1992. 71

MALSBURG, C. von der. Self-organization of orientation sensitive cells in striate cortex. **Biological Cybernetics**, v. 14, p. 85–100, 01 1973. 59

MAYER, J.; GOEPPERT, M. **Statistical mechanics**. New York: John Wiley & Sons, 1940. 37

MCCULLOCH, W.; PITTS, W. A logical calculus of ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, v. 5, p. 127–147, 1943. 57

MILLER, M. C. *et al.* PSR J0030+0451 mass and radius from nicer data and implications for the properties of neutron star matter. **The Astrophysical Journal Letters, The American Astronomical Society**, v. 887, n. 1, p. L24, Dec 2019. x, 20, 51, 53

MILLER, M. C. *et al.* The radius of PSR J0740+6620 from nicer and xmm-newton data. **The Astrophysical Journal Letters, The American Astronomical Society**, v. 918, n. 2, p. L28, Sep 2021. x, 20, 51, 53

MINSKY, M.; PAPERT, S. **Perceptrons**: an introduction to computational geometry. Cambridge, MA, USA: MIT Press, 1969. 59

MORAES, P. H. R. S. *et al.* **Compact astrophysical objects in $f(r,t)$ gravity**. [S.l]. 2018. 20

MURPHY, K. **The machine learning**: a probabilistic perspective. London: MIT Press, 2012. ISBN 978-0262018029. 73, 74, 76

NEAL, R. **Bayesian learning for neural networks**. New York: Springer, 1996. v. 118. (Lecture Notes in Statistics, v. 118). 71

O'BOYLE, M.; MARKAKIS, C.; STERGIOULAS, N.; READ, J. Parametrized equation of state for neutron star matter with continuous sound speed. **Physical Review D**, **American Physical Society**, v. 102, p. 083027, Oct 2020. 45, 46

O'CONNOR, E.; OTT, C. Black hole formation in failing core-collapse supernovae. **The Astrophysical Journal, The American Astronomical Society**, v. 730, n. 2, p. 70, 2011. 24

OERTEL, M.; HEMPEL, M.; KLÄHN, T.; TYPEL, S. Equations of state for supernovae and compact stars. **Reviews of Modern Physics**, v. 89, p. 015007, 2017. 19

OPPENHEIMER, J. R.; VOLKOFF, G. M. On massive neutron cores. **Phys. Rev.**, **American Physical Society**, v. 55, p. 374–381, Feb 1939. 23, 27

ÖZEL, F.; FREIRE, P. Masses, radii, and the equation of state of neutron stars. **Annual Review of Astronomy and Astrophysics**, v. 54, p. 401–440, 2016. 50

ÖZEL, F.; PSALTIS, D. Reconstructing the neutron-star equation of state from astrophysical measurements. **Physical Review D**, v. 80, p. 103003, 2009. 46, 54

PACINI, F. Energy emission from a neutron star. **Nature**, v. 216, n. 5115, p. 567–568, nov 1967. 23

RAAIJMAKERS, G. *et al.* Constraints on the dense matter equation of state and neutron star properties from nicer's mass-radius estimate of PSR J0740+6620 and multimessenger observations. **The Astrophysical Journal Letters**, v. 918, n. 2, p. L29, 2021. 20

RAITHEL, C.; ÖZEL, F.; PSALTIS, D. From neutron star observables to the equation of state. i. an optimal parametrization. **The Astrophysical Journal, The American Astronomical Society**, v. 831, Oct 2016. 46

RAITHEL, C.; ÖZEL, F.; PSALTIS, D. From neutron star observables to the equation of state. ii. bayesian inference of equation of state pressures. **The Astrophysical Journal, The American Astronomical Society**, v. 844, n. 2, p. 156, aug 2017. 46

RAITHEL, C. A.; MOST, E. R. Nonparametric inference of the dense matter equation of state with reliable uncertainty estimates. **Monthly Notices of the Royal Astronomical Society**, v. 517, p. 1234–1245, 2023. 50

READ, J. S. *et al.* Constraints on a phenomenologically parametrized neutron-star equation of state. **Phys. Rev. D, American Physical Society**, v. 79, Jun 2009. 46, 50

REZZOLLA, L.; ZANOTTI, O. **Relativistic Hydrodynamics**. Oxford: Oxford University Press, 2013. ISBN 9780198528906. 37, 38

RILEY, T. *et al.* A nicer view of PSR J0030+0451: Millisecond pulsar parameter estimation. **The Astrophysical Journal Letters**, v. 887, n. 1, p. L21, 2019. x, 20, 51, 53

RILEY, T. *et al.* A nicer view of the massive pulsar PSR J0740+6620 informed by radio timing and xmm-newton spectroscopy. **The Astrophysical Journal Letters**, v. 918, n. 2, p. L27, 2021. x, 20, 51, 53

RODRIGUES, H.; DUARTE, S.; OLIVEIRA, J. de. Massive compact stars as quark stars. **The Astrophysical Journal**, v. 730, n. 1, p. 31, 2011. 19

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological review**, v. 65, n. 6, p. 386–408, 1958. 57, 58

RUMELHART, D.; HINTON, G.; WILLIAMS, R. Learning representations by back-propagating errors. **Nature**, v. 323, n. 6088, p. 533–536, 1986. 57, 59

SCHWARZSCHILD, K. On the gravitational field of a mass point according to einstein's theory. **Abh. Konigl. Preuss. Akad. Wissenschaften Jahre 1906,92, Berlin,1907**, v. 1916, p. 189–196, jan 1916. 25

SHAPIRO, S. L.; TEUKOLSKY, S. A. **Black holes, white dwarfs, and neutron stars**: the physics of compact objects. New York: Wiley, 1983. ISBN 9780471873167. ix, 25, 32, 33

SOARES, B.; LENZI, C.; LOURENÃ§O, O.; DUTRA, M. Bayesian analysis of a relativistic hadronic model constrained by recent astrophysical observations. **Monthly Notices of the Royal Astronomical Society**, v. 525, p. 4347–4357, 2023. 19

TENSORFLOW DEVELOPERS. **TensorFlow Core API**: tf.GradientTape. 2024. Accessed: 2025-06-16. Disponível em: https://www.tensorflow.org/api docs/python/tf/GradientTape. 64

TOLMAN, R. C. Static solutions of einstein's field equations for spheres of fluid. **Phys. Rev., American Physical Society**, v. 55, p. 364–373, Feb 1939. 27

TOOPER, R. General relativistic polytropic fluid spheres. **The Astrophysical Journal**, v. 140, p. 434, aug 1964. 35

TRAVERSI, S.; CHAR, P. Structure of quark star: A comparative analysis of bayesian inference and neural network based modeling. **The Astrophysical Journal, The American Astronomical Society,** v. 905, n. 1, p. 9, dec 2020. 20

VASWANI, A. *et al.* Attention is all you need. *In*: PROCEEDINGS OF THE 31st INTERNATIONAL CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS. 31., 2017. **Proceedings** [...]. [S.l.]: Curran Associates, 2017. p. 6000–6010. 60

WILLSHAW, D.; MALSBURG, C. von der. How patterned neural connections can be set up by self-organization. **Proceedings of the Royal Society of London Series B**, v. 194, p. 431–445, 1976. 59

FOLHA DE REGISTRO DO DOCUMENTO

1. CLASSIFICAÇÃO/TIPO	2. DATA	3. DOCUMENTO N°	4. N° DE PÁGINAS
TD	22 de agosto de 2025	DCTA/ITA/TD-038/2025	110

5. TÍTULO E SUBTÍTULO:

Redes neurais bayesianas na inferência de equações de estado politrópicas para estrelas de nêutrons

6. AUTOR(ES):

Bruno da Silva Gonçalves

7. INSTITUIÇÃO(ÕES)/ÓRGÃO(S) INTERNO(S)/DIVISÃO(ÕES):

Instituto Tecnológico de Aeronáutica – ITA

8. PALAVRAS-CHAVE SUGERIDAS PELO AUTOR:

Equações de Estado; Estrelas de Nêutrons; Redes Neurais

9. PALAVRAS-CHAVE RESULTANTES DE INDEXAÇÃO:

Equações de estado; Estrelas de nêutrons; Redes neurais; Gravitação; Redes Bayesianas, Física nuclear; Física.

10. APRESENTAÇÃO: (**X**) **Nacional** ( ) **Internacional**

ITA, São José dos Campos. Curso de Doutorado. Programa de Pós-Graduação em Física. Área de Física Nuclear. Orientador: Prof. Dr. César Henrique Lenzi; co-orientadores: Profª. Drª. Mariana Dutra da Rosa Lourenço, Prof. Dr. Sérgio José Barbosa Duarte; Defesa em 23/07/2025. Publicada em 2025.

11. RESUMO:

A análise das equações de estado que descrevem a matéria no interior das estrelas de nêutrons contribui para a compreensão de dois pilares fundamentais da física, a matéria nuclear e a gravitação. Observações recentes, como o evento GW170817, marco inicial da era multimensageira com contrapartida gravitacional, e as medições de massa e raio realizadas pelo NICER para os pulsares PSR J0030+0451 e PSR J0740+6620, geralmente trazem evidências que desafiam explicações imediatas, desencadeando uma série de estudos dedicados à conexão entre teoria e observação. Somados ao avanço computacional, esses resultados motivaram a adoção de métodos de aprendizado de máquina aplicados à física da matéria densa. Neste contexto, o presente estudo aplicou redes neurais bayesianas à inferência de equações de estado politrópicas por partes. Para viabilizar o treinamento, foi gerado um banco de dados de quase cinco milhões de pares massa - raio, obtidos a partir do formalismo politrópico generalizado, que serviu como base algorítmica para a geração em larga escala dos dados de interesse. Sobre essa base, projetou-se uma arquitetura enxuta, com ~ 1164 parâmetros treináveis, *priors* físicos, perda de Huber e regularização KL, capaz de realizar inferência totalmente probabilística. Os testes mostraram alta precisão em cenários controlados (erros absolutos $< 10\%$), sobreposição quase completa com equações de estado de referência em cenários de validação, e consistência com observações reais como as dos estudos citados anteriormente. Em conjunto, os resultados consolidam um método modular e extensível, capaz de quantificar incertezas epistêmicas e acomodar restrições, fortalecendo a aplicação de redes neurais bayesianas na exploração da física e matéria das estrelas de nêutrons.

12. GRAU DE SIGILO:

(**X**) **OSTENSIVO** ( ) **RESERVADO** ( ) **SECRETO**