



**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO
CEARÁ
IFCE *CAMPUS* MARACANAÚ
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

LEVI CORDEIRO CARVALHO

**EXPLICAÇÕES PARA REDES NEURAIS BASEADAS EM RACIOCÍNIO
ABDUTIVO**

**MARACANAÚ - CE
2021**

LEVI CORDEIRO CARVALHO

**EXPLICAÇÕES PARA REDES NEURAIS BASEADAS EM RACIOCÍNIO
ABDUTIVO**

Trabalho de Conclusão de Curso apresentado ao
Curso de Bacharelado em Ciência da Computa-
ção do Instituto Federal de Educação, Ciência
e Tecnologia do Ceará -IFCE - *Campus* Mara-
canaú como requisito parcial para obtenção do
Título de Bacharel em Ciência da Computação.
Orientador: Prof. Dr. Thiago Alves Rocha.

MARACANAÚ - CE

2021

NOME COMPLETO

TÍTULO DO TRABALHO

Esta Monografia foi julgada adequada para obtenção do título de Licenciado em matemática e aprovada em sua forma final pelo departamento de Matemática do Instituto Federal do Ceará-*Campus Cedro*.

Aprovado em: _____ / _____ / _____

BANCA EXAMINADORA

Prof. Me. Luiz Fernando Ramos Lemos (Orientador)
IFSULDEMINAS - *Campus Inconfidentes*.

Prof.(a). Ma. Mikaelle Barboza Cardoso
IFCE - *Campus Sobral*

Prof. Dr. João Nunes de Araújo Neto
IFCE - *Campus Cedro*

DEDICATÓRIA

À minha mãe, ...

“Astronarta libertado
Minha vida me urtrapassa
Em quarqué rota que eu faça.”

(Dois mil e um - Tom Zé)

AGRADECIMENTOS

Graças à vida, que me deu tanto...

Ao Instituto Federal de Educação, Ciência e Tecnologia do Ceará - *Campus* Cedro, todos os servidores, professores e alunos.

Não esqueça de agradecer às instituições que lhe forneceram algum tipo de financiamento ao longo da graduação!!!

RESUMO

Resumo em português

Palavras-chave: Matemática. Educação. Função Afim. Função Definida por Partes. PDI.

ABSTRACT

English abstract.

Keywords: Mathematics. Education. Affine Function. Piecewise Function. DIP.

LISTA DE ILUSTRAÇÕES

Quadro 1	— LIVROS ANALISADOS	15
----------	-------------------------------	----

LISTA DE FIGURAS

Figura 1	— Representação gráfica da função afim	16
----------	--	----

LISTA DE CÓDIGOS

Código 1	— Método da Bissecção	18
----------	---------------------------------	----

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Introdução do tema	12
1.2	Resumo dos artigos selecionados	13
2	MOTIVAÇÃO INICIAL E JUSTIFICATIVA FUNDAMENTADA DO ESTUDO	15
3	FUNDAMENTAÇÃO TEÓRICA	16
3.1	Teoria dos Registros das Representações Semióticas	16
3.2	Exemplo de seção.	17
3.2.1	<i>Exemplo de subseção</i>	<i>17</i>
3.2.2	<i>Imagem da função afim</i>	<i>17</i>
3.2.3	<i>Zero da função</i>	<i>17</i>
3.2.4	<i>Exemplo de subseção</i>	<i>17</i>
3.3	Outra seção	18
4	ABORDAGEM AO PROBLEMA	19
5	CONSIDERAÇÕES FINAIS	20
ANEXO A	Apêndice	21
ANEXO A.1	Texto auxiliar do trabalho	22
REFERÊNCIAS		24

1 INTRODUÇÃO

1.1 Introdução do tema

As redes neurais artificiais (do inglês, *Artificial Neural Network* - ANN) são utilizadas em diversas aplicações para resolução de problemas, como classificação de imagens, detecção de objetos e reconhecimento de fala. Para que esses algoritmos alcancem um resultado satisfatório, um dos principais requisitos é realizar o treinamento da rede com o uso de um conjunto de dados que representem bem os casos do mundo real, com o objetivo de garantir a generalização do aprendizado para dados não vistos.

Apesar do grande sucesso das redes neurais, elas podem ser classificadas como um algoritmo caixa preta, sendo, basicamente, o funcionamento computado para a aquisição da resposta da ANN, dado um conjunto de entradas, não humanamente interpretável. Essa falta de explicações para o seu comportamento pode prejudicar sistemas críticos que permitem apenas uma pequena margem para falhas, como aplicações médicas e financeiras. Além disso, com o avanço de leis de proteção de dados observável em diversas regiões do mundo, como na União Europeia (COMISSÃO EUROPEIA, 2018), está ficando cada vez mais necessário disponibilizar uma explicação do que está sendo realizado com os dados dos seus respectivos donos.

Os métodos heurísticos são as principais abordagens utilizadas com o intuito de disponibilizar explicações para as saídas das redes neurais atualmente, como o LIME (RIBEIRO; SINGH; GUESTRIN, 2016) e ANCHOR (RIBEIRO; SINGH; GUESTRIN, 2018). Porém, de acordo com o trabalho de Ignatiev, Narodytska e Marques-Silva (2019-b), como eles exploram o espaço da instância localmente, esses algoritmos não acarretam em respostas que possuam garantias formais, resultando em explicações que podem não ser verdadeiras para toda instância pertencente ao espaço de instâncias.

Fundamentado na importância de disponibilizar explicações decorrentes de algoritmos que carregam garantias formais, este trabalho possui o objetivo de apresentar uma abordagem baseada em lógica utilizando o raciocínio abduutivo para a aquisição de explicações para redes neurais artificiais.

1.2 Resumo dos artigos selecionados

O trabalho de Ignatiev, Narodytska e Marques-Silva (2019-a) apresenta um algoritmo *model-agnostic*, significando que funciona em qualquer modelo desde que ele possa ser representado por um sistema de raciocínio de restrição e que consultas de implicação possam ser decididas por um oráculo. Essa abordagem utiliza o raciocínio abdutivo para computar explicações mínimas para modelos de aprendizagem de máquina com garantias formais, fornecendo explicações de cardinalidade mínima ou subconjunto mínima. A abordagem basicamente codifica um modelo M de aprendizagem de máquina como um conjunto de restrições lineares F em alguma teoria a partir de um algoritmo 0-1 MILP (FISCHETTI; JO, 2018), então, dado uma instância C associado com a predição E tal que $C \wedge F = E$ (equivalente à $C = (F \rightarrow E)$), é computada a explicação mínima C' (dado C) que é implicante principal de $F \rightarrow E$.

O artigo de Fischetti e Jo (2018) propõe um algoritmo que realiza a codificação de uma rede neural profunda como um 0-1 MILP que usa restrições de indicação para evitar o uso da notação de *Big-M*, utilizando variáveis de ativação binárias para impor as implicações lógicas. Foi realizado a modelagem de aplicações que usavam redes com ReLUs e *max/average pooling*. Esse modelo codificado em restrições lineares não é suscetível à treinamento, pois ele se torna bilinear nessa configuração. Foi realizado experimentos do algoritmo em dois problemas de aprendizagem de máquina: visualização de características e aprendizagem de máquina adversário.

O trabalho de Tjeng, Xiao e Tedrake (2019) implementa um algoritmo MILP que codifica as partes lineares (camadas que usam transformações lineares ou funções que possuam partes lineares, como ReLU e *max pooling*) de uma rede neural como restrições lineares, minimizando o número de variáveis binárias presentes no problema MILP e melhorando o condicionamento numérico. Para a formulação das funções de ativação ReLU é considerado que há 3 fases: a unidade está estavelmente inativa, estavelmente ativa e instável. Essa abordagem foi realizada na aplicação de exemplos adversariais sendo de duas a três vezes de ordem de magnitude mais rápida que o estado da arte, permitindo um aumento significativo do tamanho das redes neurais codificadas.

O artigo de Ignatiev, Narodytska e Marques-Silva (2019-b) descreve os experimentos realizados que questionam as explicações de modelos de aprendizagem de máquina dadas por métodos heurísticos, que computam uma explicação explorando localmente o espaço da instância. Para realização desses experimentos é utilizado abordagens baseadas em lógica com o cálculo de implicantes principais por meio do raciocínio abdutivo, realizando os testes em *boosted trees*. Os algoritmos desenvolvidos possuem os objetivos de acessar a qualidade das explicações locais, reparar as explicações locais que são otimistas (quando existem instâncias no espaço de instâncias que a explicação

computada falha) e refinar as explicações locais que são pessimistas (quando alguma literal pertencente à explicação computada é irrelevante e pode ser descartada).

O trabalho de Katz et al. (2017) implementa o algoritmo Reluplex, um SMT *solver* para a teoria da aritmética linear real que tem o objetivo de verificar redes neurais profundas utilizando uma técnica baseada no Simplex, porém adaptada para lidar com a função de ativação não-convexa ReLU sem simplificações no seu funcionamento original, apenas permitindo que suas entradas e saídas sejam temporariamente inconsistentes e corrigidas conforme a execução do algoritmo. O Reluplex foi avaliado em 45 redes neurais profundas desenvolvidas como um protótipo inicial do sistema anti-colisão de última geração para aeronaves não tripuladas ACAS Xu.

Este trabalho é fundamentado no artigo de Ignatiev, Narodytska e Marques-Silva (2019-a), que utiliza um 0-1 MILP (FISCHETTI; JO, 2018) para a codificação da rede neural como um conjunto de restrições lineares em uma teoria, então são computadas explicações para os seus resultados com o uso do raciocínio abdutivo. A necessidade de tais explicações computadas carregando uma garantia formal é evidenciada em Ignatiev, Narodytska e Marques-Silva (2019-b). Um dos objetivos iniciais é comparar os possíveis resultados desse algoritmo utilizando diferentes MILPs e estratégias de codificação (verificação) das redes neurais, como os implementados em Tjeng, Xiao e Tedrake (2019) e Katz et al. (2017).

2 MOTIVAÇÃO INICIAL E JUSTIFICATIVA FUNDAMENTADA DO ESTUDO

Apresente o que motivou o estudo e a justificativa da relevância e necessidade do estudo. O Quadro 1 é uma exemplo de quadro.

Quadro 1 – LIVROS ANALISADOS

Referência para citar no texto	Título do livro	Autor/Autores
Livro 1	Matemática Completa	Bonjorno, Giovanni Jr e Paulo Câmara
Livro 2	Matemática: Contexto e Aplicações	Luiz Roberto Dante
Livro 3	Matemática	Emanuel Paiva
Livro 4	Matemática: Ciência e Aplicações	Gelson Iezzi, Osvaldo Dulce, David Degenszajn, Roberto Périco e Nilse de Almeida
Livro 5	Matemática para compreender o mundo	Kátia Stocco Smole e Maria Ignez Diniz
Livro 6	Fundamentos de Matemática Elementar	Gelson Iezzi e Carlos Marukami

Fonte: Elaborado pela autor(a).

Um exemplo de lista de itens.

Livro 1 - Neste livro, ... Ao final das seções percebemos razoável variação de exercícios resolvidos e propostos.

Livro 2 - No segundo livro, ...

Livro 3 - Nesse exemplar ...

Livro 4 - ??), ...

Livro 5 - As autoras abordam

Livro 6 - Neste livro, ...

3 FUNDAMENTAÇÃO TEÓRICA

Apresente um resumo das teorias utilizadas de forma a facilitar o acesso ao leitor do trabalho aos pré-requisitos para o entendimento do trabalho.

3.1 Teoria dos Registros das Representações Semióticas

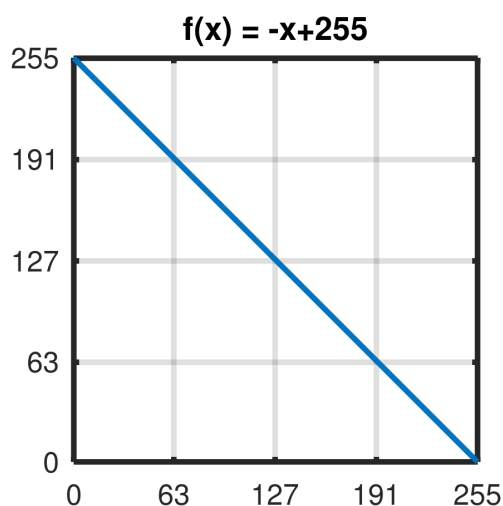
Ensinar é uma tarefa ... O psicólogo e filósofo Raymond Duval, desenvolveu a Teoria dos Registros de Representação Semiótica - (TRRS)...

A *conversão* para ??) é (exemplo de citação):

[...] a conversão de uma representação se refere às operações em que o registro inicial é transformado em outro registro; por essa razão, é considerada como uma “transformação externa”. Por exemplo, ao utilizarmos a linguagem algébrica para representar a frase “o dobro de um número resulta em oito”, estamos realizando uma conversão do registro dado na língua natural para o registro dado na linguagem algébrica (??, p. 6).

Exemplo de figura.

Figura 1 – Representação gráfica da função afim



Fonte: Elaborada pelo autor

Assim, a Figura 1..

3.2 Exemplo de seção.

??, p. 81), define Função como:

Definição 3.1 *Dado dois conjuntos A e B , não vazios, uma relação f de A em B recebe o nome de aplicação de A em B ou função definida em A com imagens em B se, e somente se, para todo $x \in A$ existe um só $y \in B$ tal que $(x, y) \in f$.*

$$f \text{ aplicado de } A \text{ em } B \iff (\forall x \in A, \exists |y \in B| (x, y) \in f) \quad (3.1)$$

3.2.1 Exemplo de subseção

??, p. 100), define função afim como:

Definição 3.2 *Uma aplicação de \mathbb{R} em \mathbb{R} com $a \neq 0$ e cada $x \in \mathbb{R}$ associa o elemento $(a \cdot x + b) \in \mathbb{R}$.*

$$f(x) = a \cdot x + b \quad \text{com} \quad (a \neq 0) \quad (3.2)$$

3.2.2 Imagem da função afim

??, p. 105) diz que:

reta permitindo a análise de que todos os valores de y estão relacionados com x .

3.2.3 Zero da função

Vejamos que $f(x)$ é crescente pois na medida que os valores em x vão aumentando, as suas respectivas imagens também crescem.

3.2.4 Exemplo de subseção

??, p. 118), resume em:

A função afim:

$$f(x) = a \cdot x + b \text{ anula-se para } x = -\frac{b}{a}. \quad (3.3)$$

Para $x > -\frac{b}{a}$, temos:

$$\begin{cases} \text{se } a > 0 \text{ então } f(x) = a \cdot x + b > 0 \\ \text{se } a < 0 \text{ então } f(x) = a \cdot x + b < 0 \end{cases} \quad (3.4)$$

Isto é, $x > -\frac{b}{a}$ a função $f(x) = a \cdot x + b$ tem sinal de a .

Para $x < -\frac{b}{a}$, temos:

$$\begin{cases} \text{se } a > 0 \text{ então } f(x) = a \cdot x + b < 0 \\ \text{se } a < 0 \text{ então } f(x) = a \cdot x + b > 0 \end{cases} \quad (3.5)$$

Isto é, para $x < -\frac{b}{a}$ a função $f(x) = a \cdot x + b$ tem o sinal de ‘-a’ (sinal contrário ao de a).

Exemplo de função definida por partes:

Exemplo 3.1 Seja $f : \mathbb{R} \rightarrow \mathbb{R}$ definida por:

$$f(x) = \begin{cases} x & \text{se } 0 \leq x \leq 128 \\ 128 & \text{se } 128 \leq x \leq 256 \\ x - 128 & \text{se } c.c \end{cases} \quad (3.6)$$

3.3 Outra seção

Exemplo de código.

Código 1 — Método da Bissecção

```
function xm=mb(f,xp,xn) % metodo da bissecao para zero de funcoes
xm=(xp+xn)/2;
y=f(xm);
while(abs(y)>0.01) % enquanto |y|>ep -> laço de repeticao
    if(y>0) % se y maior que 0
        xp=xm;
    else % senao
        xn=xm;
    end
    xm=(xp+xn)/2;
    y=f(xm);
end
end
```

Fonte: Elaborada pela autor(a) (GNU Octave)

4 ABORDAGEM AO PROBLEMA

Apresente a sua proposta de abordagem ao problema ou discussão da questão do trabalho monográfico.

5 CONSIDERAÇÕES FINAIS

Apresente suas considerações finais.

ANEXO A Apêndice

ANEXO A.1 Texto auxiliar do trabalho

O apêndice deve ser autoral, textos externos devem ser colocados como anexo.

REFERÊNCIAS

COMISSÃO EUROPEIA. *Reforma de 2018 das regras de proteção de dados da UE*. 2018. Disponível em: <https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes_en.pdf>.

FISCHETTI, M.; JO, J. Deep neural networks and mixed integer linear optimization. *Constraints*, v. 23, p. 296–309, Jul. 2018. Disponível em: <<https://link.springer.com/article/10.1007/s10601-018-9285-6>>.

IGNATIEV, A.; NARODYTSKA, N.; MARQUES-SILVA, J. Abduction-based explanations for machine learning models. *Proceedings of the AAAI Conference on Artificial Intelligence*, v. 33, n. 01, p. 1511–1519, Jul. 2019-a. Disponível em: <<https://ojs.aaai.org/index.php/AAAI/article/view/3964>>.

IGNATIEV, A.; NARODYTSKA, N.; MARQUES-SILVA, J. On validating, repairing and refining heuristic ML explanations. *CoRR*, abs/1907.02509, 2019–b. Disponível em: <<http://arxiv.org/abs/1907.02509>>.

KATZ, G. et al. Reluplex: An efficient SMT solver for verifying deep neural networks. In: *Computer Aided Verification*. Springer International Publishing, 2017. p. 97–117. Disponível em: <https://doi.org/10.1007/978-3-319-63387-9_5>.

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. "why should i trust you?": Explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery, 2016. (KDD '16), p. 1135–1144. ISBN 9781450342322. Disponível em: <<https://doi.org/10.1145/2939672.2939778>>.

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. Anchors: High-precision model-agnostic explanations. In: MCILRAITH, S. A.; WEINBERGER, K. Q. (Ed.). *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*. AAAI Press, 2018. p. 1527–1535. Disponível em: <<https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16982>>.

TJENG, V.; XIAO, K. Y.; TEDRAKE, R. Evaluating robustness of neural networks with mixed integer programming. In: *International Conference on Learning Representations*.

[s.n.], 2019. Disponível em: <<https://openreview.net/forum?id=HyGIdiRqtm>>.