



Universidade Federal
do Rio de Janeiro
Escola Politécnica

TÍTULO

João Victor Almeida Davim

Projeto de Graduação apresentado ao curso de Engenharia de Computação e Informação da Escola Politécnica da Universidade Federal do Rio de Janeiro como parte dos requisitos necessários para a obtenção do grau de Engenheiro de Computação e Informação.

Orientadores: Priscila Machado Vieira Lima
Leopoldo André Dutra
Lusquino Filho

Rio de Janeiro
Agosto de 2021

TÍTULO

João Victor Almeida Davim

PROJETO SUBMETIDO AO CORPO DOCENTE DO CURSO DE ENGENHARIA DE COMPUTAÇÃO E INFORMAÇÃO DA ESCOLA POLITÉCNICA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE ENGENHEIRO DE COMPUTAÇÃO E INFORMAÇÃO.

Examinadores:

Prof. Nome do Primeiro Examinador Sobrenome, D.Sc.

Prof. Nome do Segundo Examinador Sobrenome, Ph.D.

Prof. Nome do Terceiro Examinador Sobrenome, D.Sc.

Prof. Nome do Quarto Examinador Sobrenome, Ph.D.

Prof. Nome do Quinto Examinador Sobrenome, Ph.D.

RIO DE JANEIRO, RJ – BRASIL

AGOSTO DE 2021

Almeida Davim, João Victor

Título/João Victor Almeida Davim. – Rio de Janeiro:
UFRJ/POLI – COPPE, 2021.

XI, 17 p.: il.; 29, 7cm.

Orientadores: Priscila Machado Vieira Lima

Leopoldo André Dutra Lusquino Filho

Projeto (graduação) – UFRJ/ Escola Politécnica/ Curso
de Engenharia de Computação e Informação, 2021.

Referências Bibliográficas: p. 15 – 16.

1. Primeira palavra-chave. 2. Segunda palavra-chave.
3. Terceira palavra-chave. I. Machado Vieira Lima,
Priscila *et al.* II. Universidade Federal do Rio de Janeiro,
Escola Politécnica/ Curso de Engenharia de Computação
e Informação. III. Título.

*A alguém cujo valor é digno
desta dedicatória.*

Agradecimentos

Gostaria de agradecer a todos.

Resumo do Projeto de Graduação apresentado à Escola Politécnica/DEL/UFRJ como parte dos requisitos necessários para a obtenção do grau de Engenheiro de Computação e Informação.

TÍTULO

João Victor Almeida Davim

Agosto/2021

Orientadores: Priscila Machado Vieira Lima

Leopoldo André Dutra Lusquino Filho

Curso: Engenharia de Computação e Informação

Apresenta-se, nesta tese, ...

Palavras-Chave: Primeira palavra-chave, Segunda palavra-chave, Terceira palavra-chave.

Abstract of the Undergraduate Project presented to Poli/DEL/UFRJ as a partial fulfillment of the requirements for the degree of Computer and Information Engineer.

TITLE

João Victor Almeida Davim

August/2021

Advisors: Priscila Machado Vieira Lima

Leopoldo André Dutra Lusquino Filho

Course: Computer and Information Engineering

In this work, we present ...

Keywords: .

Sumário

Lista de Figuras	x
Lista de Tabelas	xi
1 Introdução	1
2 Previsão de Séries Temporais	2
2.1 Definição do problema	2
2.2 Modelos	2
2.2.1 ARMA	2
2.2.2 ARIMA	3
2.3 Tecnologias	4
2.3.1 Biblioteca statsmodels	4
2.3.2 Biblioteca fbprophet	4
3 Redes Neurais sem Peso	5
3.1 WiSARD	6
3.2 Regression WiSARD	8
3.3 Representação da entrada	10
3.4 Biblioteca wisardpkg	11
4 Autoregressão com WiSARD	12
4.1 Janelas deslizantes	12
4.2 Média móvel	12
4.3 AutoRegression WiSARD	12
5 Avaliação Experimental	13
5.1 Ambiente computacional	13
5.2 Coleções de dados	13
5.3 Métricas de avaliação	13
5.4 Resultados	13
5.4.1 Experimentos com dados sintéticos	13

5.4.2	Experimentos com dados do covid	13
5.4.3	Discussão	13
6	Conclusão	14
	Referências Bibliográficas	15
A	Título do Apêndice	17

Lista de Figuras

3.1	WiSARD e Discriminador. Em caso de empate, o operador máximo seleciona aleatoriamente um dos máximos encontrados.	6
3.2	Treinamento de um exemplo da classe I no seu respectivo discriminador.	7
3.3	Classificação de um exemplo da classe I no seu respectivo discriminador já treinado.	7
3.4	Exemplo de bleaching em um discriminador.	8
3.5	RAM da WiSARD na esquerda e RAM da Regression WiSARD na direita.	9
3.6	Discriminador na etapa de inferência da Regression WiSARD transformando a entrada com a função T e agregando o valor de inferência com a função média simples.	9
3.7	Transformação termômetro do número 210	10

Lista de Tabelas

Capítulo 1

Introdução

Capítulo 2

Previsão de Séries Temporais

2.1 Definição do problema

Previsão de séries temporais é uma solução adotada para muitos problemas da vida real com o objetivo de, a partir de dados observados, realizar extrapolações para o futuro. Esse tipo de abordagem é muito eficaz quando há poucos dados disponíveis dentro do escopo do problema, pois permite realizar inferências que utilizam dados históricos da própria variável, e podem ser essenciais para o auxílio à tomada de decisão.

2.2 Modelos

Existem muitos modelos que são frequentemente utilizados para tarefas de predição de séries temporais. Dentre estes, podemos citar desde modelos estatísticos como ARMA [1] e suas variações [2] até métodos com modelos aditivos como o Prophet [3]. Além desses, existem muitos estudos que fazem adaptações de Redes Neurais para esta tarefa como em [4] e até mesmo modelos misturados de Redes Neurais com ARIMA como apresentado em [5].

Os modelos utilizados para os experimentos deste trabalho foram escolhidos com base em estudos paralelos realizados com os dados do Covid-19, e serão descritos nas subseções seguintes com suas respectivas formalizações.

2.2.1 ARMA

O modelo ARMA é uma combinação dos modelos auto-regressivo (AR) e de médias móveis (MA). O modelo auto regressivo puro de ordem p , ou seja, $AR(p)$, pode ser formalizado como:

$$X_t = c + \sum_{i=1}^p \rho_i X_{t-i} + \varepsilon_t$$

onde $\rho_1 \dots \rho_p$ são os parâmetros, c é uma constante e ε_t representa um ruído branco.

O ajuste de parâmetros a ser realizado na etapa de treinamento do modelo pode ser feito utilizando o método dos mínimos quadrados ou método dos momentos através das equações de Yule-Walker.

O modelo de médias móveis (MA) é utilizado junto com o modelo auto regressivo, e é uma abordagem comum para modelagem de séries temporais univariadas. Este modelo estabelece uma dependência linear entre os valores atuais da série temporal e os valores passados, e pode ser formalizado como:

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

onde μ é a média da série, $\theta_1, \dots, \theta_q$ são os parâmetros e $\varepsilon_t, \dots, \varepsilon_{t-q}$ são os termos do ruído branco.

Cada modelo pode ser utilizado de forma independente, ou em conjunto formando o modelo ARMA. A construção desse modelo se dá pela soma do modelo auto-regressivo com o modelo de médias móveis, e pode ser formalizado como:

$$X_t = c + \varepsilon_t + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$$

Existem diversas técnicas para a otimização dos hiperparâmetros p e q para que o modelo seja melhor ajustado aos dados. Uma das formas mais aceitas para encontrar esses valores é o critério de Akaike, conforme recomendado por Peter J. Brockwell e Richard A. Davis em [6].

2.2.2 ARIMA

Alguns problemas de previsão de séries temporais são ligeiramente mais complexos, pois representam uma série temporal não estacionária. Nesse caso específico, o ajuste do modelo ARMA acaba não sendo bom, pois não acompanha a não estacionariedade dos dados. O modelo auto-regressivo integrado de médias móveis (ARIMA) se propõe a resolver o problema da estacionariedade, e pode ser estudado como uma generalização do modelo ARMA.

De forma análoga ao ARMA, os modelos ARIMA são geralmente denotados como $ARIMA(p, d, q)$, onde os parâmetros p , d e q são inteiros não negativos que representam a ordem do modelo auto-regressivo (AR), o grau de diferenciação da parte integrada (I) e a ordem do modelo de médias móveis (MA) respectivamente.

A parte integrada do modelo consiste na computação da diferença dos valores atuais da série temporal com os valores subsequentes da mesma. Esse procedimento é realizado d vezes, e este é o parâmetro da parte integrada do modelo.

Pode-se formalizar o molde ARIMA como:

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1 - L)^d X_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t$$

em que X_t é a série temporal de dados, t é um índice representado por um inteiro, L é o operador de defasagem, p é o hiperparâmetro do modelo auto-regressivo, q é o hiperparâmetro do modelo de médias móveis, d é o número de diferenças da parte integrada, ϕ_i são os parâmetros do modelo auto-regressivo, θ_i são os parâmetros do modelo de médias móveis e ε_t é o ruído branco ou erro aleatório.

A estimação de modelos ARIMA é realizada, geralmente, através do método de Box-Jenkins [1], que é um processo iterativo para encontrar os parâmetros que melhor ajusta o modelo aos dados observados.

2.3 Tecnologias

2.3.1 Biblioteca statsmodels

2.3.2 Biblioteca fbprophet

[3]

Capítulo 3

Redes Neurais sem Peso

O cérebro humano faz parte do sistema nervoso e é considerado o núcleo de inteligência e aprendizado de um indivíduo. Composto por células nervosas chamadas de neurônios, permite atividades como o controle das ações motoras, integração dos estímulos sensoriais e atividades neurológicas como a memória e reconhecimento de padrões.

As Redes Neurais Artificiais são baseadas em modelos matemáticos e técnicas computacionais inspiradas na estrutura neural dos organismos vivos. Geralmente são compostas por neurônios artificiais interconectados, que aplicam funções no sinal de entrada e alimentam a entrada do próximo neurônio formando uma rede, onde cada neurônio é responsável por parte do processamento da informação. Para cada conexão é atribuído um peso multiplicativo, que é um parâmetro a ser ajustado pelo algoritmo de otimização responsável pelo treinamento.

As Redes Neurais sem Peso possuem neurônios que, ao invés de aplicar funções no sinal de entrada, participam do aprendizado de forma semelhante às memórias de acesso aleatório (RAM). A analogia biológica de tal neurônio é feita com o comportamento excitatório ou inibitório do sinal de entrada da árvore dendrítica. A "força" de um sinal de entrada da árvore dendrítica depende da altura que a conexão sináptica é posicionada, assim como as RAMs decodificam um sinal de entrada binário (excitatório/inibitório) em endereços de memória [7].

As redes neurais sem peso foram inspiradas no classificador *n-tuples* [8], que também aplica a decodificação do sinal de entrada para o reconhecimento de padrões. Uma grande aplicação para esse método é o reconhecimento de caracteres. Fotomosaicos com caracteres manuscritos eram representados através de fotocélulas, que por sua vez eram utilizadas como sinal de entrada binário, ou seja, cada fotocélula podia ser estar preenchida ou não formando um padrão binário que representa o caractere e é utilizado para o treinamento do modelo.

A próxima seção explica detalhadamente o processo de codificação tal como descrito pelo parágrafo anterior e a arquitetura de Rede Neural sem Peso utilizada

neste trabalho como ponto de partida para o entendimento do modelo de RNSP para previsão de séries temporais.

3.1 WiSARD

A arquitetura do modelo WiSARD (Wilkie, Stonhan and Aleksander Recognition Device) é composta por discriminadores, que são componentes responsáveis pela identidade de uma classe em um problema de classificação supervisionado. Cada discriminador é formado por um conjunto específico de memórias de acesso aleatório (RAMs), que são responsáveis por armazenar o padrão reconhecido no exemplo de entrada. A Figura 3.1 representa a estrutura da WiSARD seguida abaixo de um de seus Discriminadores.

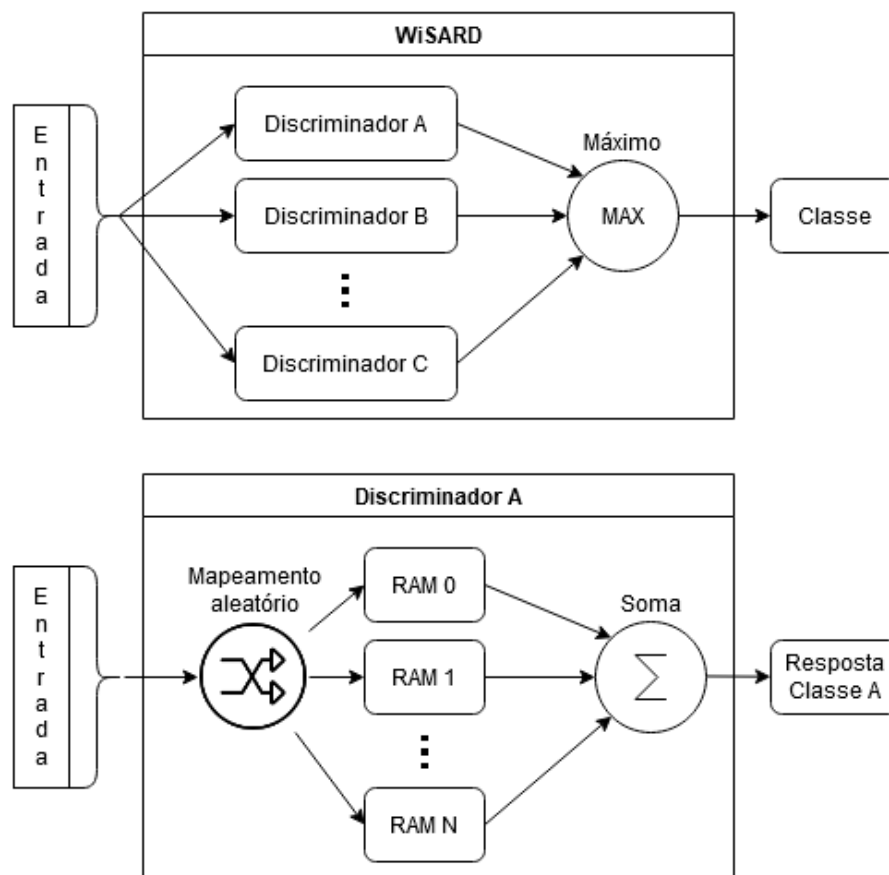


Figura 3.1: WiSARD e Discriminador. Em caso de empate, o operador máximo seleciona aleatoriamente um dos máximos encontrados.

O treinamento de um modelo WiSARD é dado pela escrita nas RAMs de cada discriminador, enquanto a classificação é dada pela leitura dessas posições de memória. A escrita se dá através de um mapeamento aleatório dos bits de entrada em um conjunto de endereços, que serão utilizados para apontar as posições de memória que

devem ser escritas. Portanto, se faz necessária a utilização de uma entrada binária para a rede.

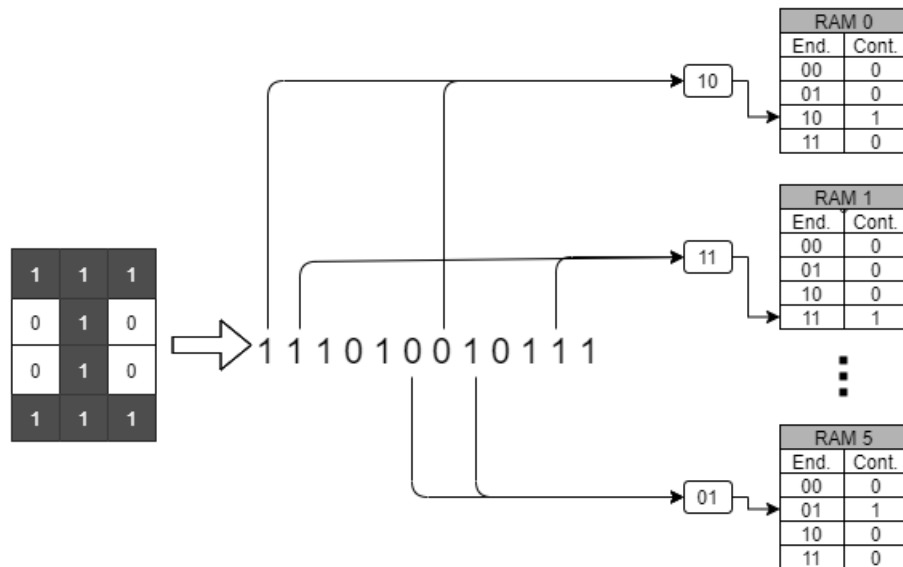


Figura 3.2: Treinamento de um exemplo da classe I no seu respectivo discriminador.

A classificação é realizada lendo-se o conteúdo das RAMs de cada discriminador e comparando a entrada a ser classificada com o conteúdo das RAMs de cada discriminador a fim de descobrir à qual discriminador (ou classe) o exemplo pertence.

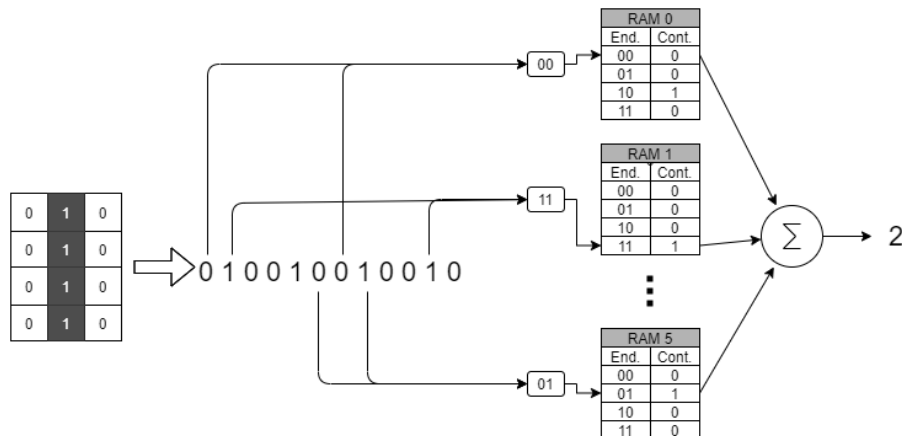


Figura 3.3: Classificação de um exemplo da classe I no seu respectivo discriminador já treinado.

Como o modelo requer uma entrada binária tanto para a etapa de treinamento quanto para a entrada de classificação, há uma dependência forte do pré-processamento dos dados a fim de obter uma representação razoável no formato binário, ou seja, transformar os dados de entrada em binário mantendo a capaci-

dade de generalização do modelo. Algumas técnicas de pré-processamento serão apresentadas na Seção 3.3.

Um problema evidente no modelo WiSARD é a possibilidade de empate entre dois ou mais discriminadores, ou seja, mais de um discriminador com a mesma quantidade (máxima) de RAMs que acessaram posições escritas no momento da classificação. Portanto, para contornar tal problema, é utilizada a técnica *bleaching*, introduzida em [9] e primeiramente utilizada como *bleaching* em [10]. A técnica consiste em utilizar as posições de RAM como contadores ao invés de bits, e aplicar um valor limite na saída de cada RAM, de forma que a resposta do discriminador seja a soma do número de RAMs que apresentam o valor acessado superior ao valor limite escolhido como hiperparâmetro.

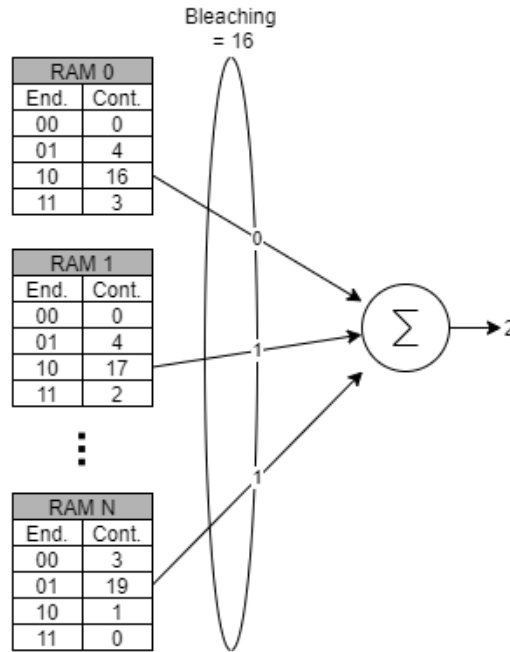


Figura 3.4: Exemplo de bleaching em um discriminador.

3.2 Regression WiSARD

A tarefa regressão, assim como a de classificação, é uma das tarefas mais abrangentes e divulgadas na área de aprendizado de máquina. Como descrito no capítulo anterior, o modelo WiSARD é utilizado para a resolução de problemas de classificação, mas, como alguns outros modelos de aprendizado de máquina, também pode ser utilizado para problemas de regressão, necessitando apenas de algumas modificações em sua arquitetura. A principal modificação necessária para utilização da WiSARD para a tarefa de regressão está na estrutura da RAM. Essa adaptação é um aumento

de dimensionalidade dos valores armazenados em cada posição de memória, ou seja, enquanto na WiSARD cada posição de memória armazena um número inteiro (contador), na Regression WiSARD cada posição de memória armazena 2 valores: o número de acessos e o somatório do valor alvo dos exemplos que acessaram esta posição. A Figura 3.2 ilustra a diferença entre as duas arquiteturas.

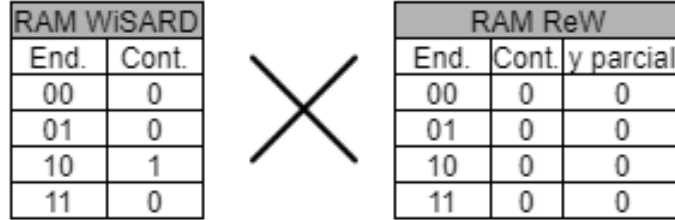


Figura 3.5: RAM da WiSARD na esquerda e RAM da Regression WiSARD na direita.

Para a etapa de treinamento essa é a única mudança necessária. Cada endereço, quando acessado, tem seu contador incrementado em 1 e seu valor incrementado do valor alvo do exemplo que acessou a posição. Já a etapa de inferência, após as RAMs já estiverem preenchidas, o acesso às posições continua sendo feito da mesma forma, porém a resposta do discriminador se torna uma função do contador e do valor das posições de memória acessadas.

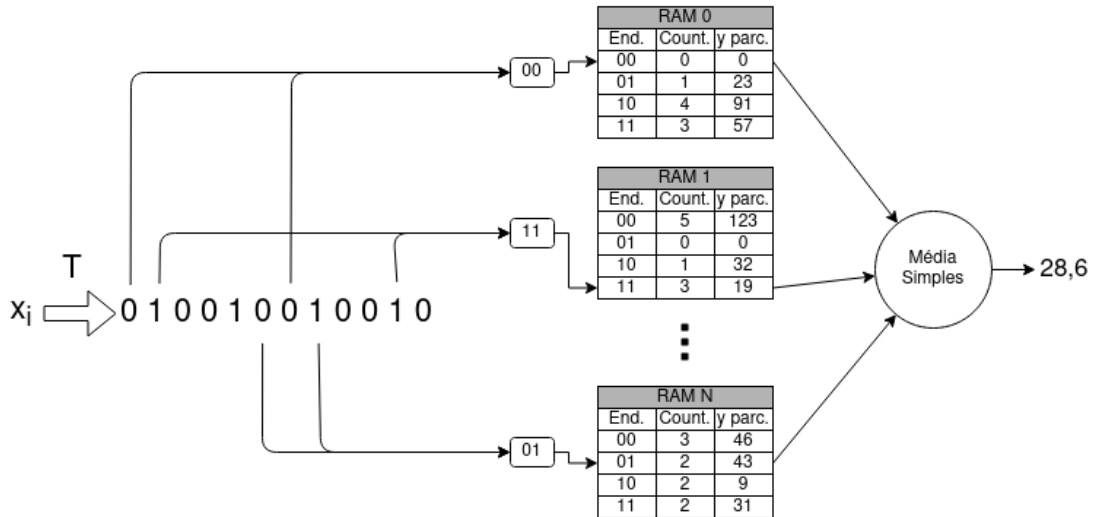


Figura 3.6: Discriminador na etapa de inferência da Regression WiSARD transformando a entrada com a função T e agregando o valor de inferência com a função média simples.

O exemplo da Figura 3.2 mostra o discriminador utilizando uma média simples como função dos valores da posição de memória acessada. É bem intuitivo e coerente

pensar que essa função pode variar de acordo com o problema e ser tratada como um hiperparâmetro do modelo. Alguns exemplos de funções utilizadas são a média simples, mediana, média harmônica e média geométrica [11].

Além da função de agregação, é fato a permanência da necessidade de transformação da entrada para valores binários na Regression WiSARD. Para tal, existem diferentes técnicas de binarização que serão apresentadas na Seção 3.3.

3.3 Representação da entrada

Tanto a WiSARD quanto a *Regression* WiSARD possuem como requisito a representação binária da entrada. Como grande parte dos problemas do mundo real não são representados de forma binária, então é imprescindível a utilização de uma técnica de binarização que minimize a perda de informação para o modelo. Existem diversas técnicas já desenvolvidas que possuem vantagens e desvantagens quando utilizadas como pré-processamento para o treinamento de RNSPs, como o Limiar, a Transformação Termômetro, Filtro de *Marr-Hildreth*, Filtro Laplaciano, entre outros. Um estudo comparativo de tais métodos pode ser encontrado em [12].

Para o escopo desse trabalho, será utilizada a transformação Termômetro, que possui as características necessárias para garantir um bom desempenho da rede. A transformação recebe 3 parâmetros: tamanho (*sz*), valor mínimo (*min*) e valor máximo (*max*). O tamanho é a quantidade de bits que é utilizada para representar 1 número real, enquanto o valor mínimo e máximo representam o menor e o maior valor real possível assumido respectivamente. O algoritmo da transformação termômetro consiste na divisão do espaço entre o mínimo e o máximo em *sz* pedaços de mesmo tamanho, atribuindo um valor limite para cada uma das divisões. Em seguida, cada pedaço é preenchido por um bit 1 ou 0 dependendo se o valor que está sendo transformado está acima ou abaixo do valor limite. A Figura 3.3 ilustra um número inteiro antes e após a transformação termômetro.

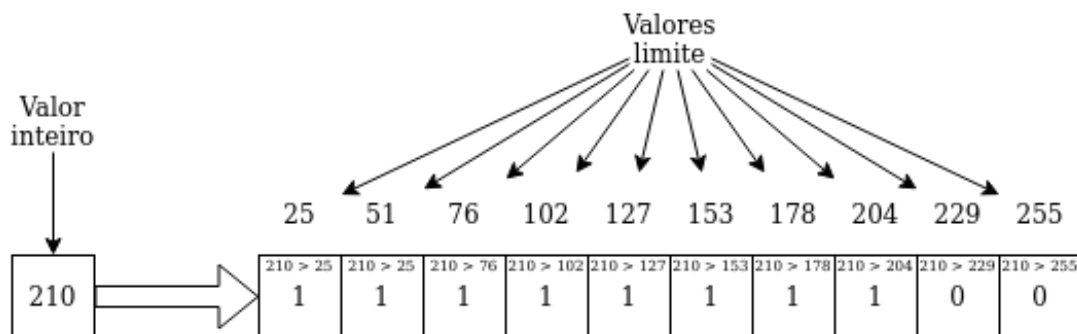


Figura 3.7: Transformação termômetro do número 210

É importante evidenciar que as técnicas utilizadas acima são fundamentais. A transformação dos números em binário modificando sua base de 10 para 2 tem um problema crucial. Ao mudar a base, os números perdem a propriedade de proximidade na representação, o que dificulta o processo de aprendizado da rede. Por exemplo, os números 11 e 12 possuem uma representação bem diferente quando transformados da base 10 para 2, quando na verdade, deveriam ter representações próximas, pois são números próximos. A transformação termômetro garante essa propriedade.

3.4 Biblioteca wisardpkg

Quase todos os métodos descritos nas seções anteriores, incluindo as RNSPs WiSARD e Regression WiSARD, são implementados na biblioteca wisardpkg.

Capítulo 4

Autoregressão com WiSARD

Como previsto no capítulo de redes neurais sem peso, o modelo Regression WiSARD é utilizado para resolver problemas de regressão, porém não é possível realizar previsão de séries temporais por depender de características que não são função do tempo. É possível, entretanto, tratar um problema de previsão de séries temporais como um problema de regressão aplicando técnicas como a janela deslizante e médias móveis conforme detalhado nas seções seguintes.

4.1 Janelas deslizantes

O método de janelas deslizantes é essencial para a Regression WiSARD ser capaz de realizar previsões de séries temporais. Isso ocorre porque é a técnica que permite transformar o problema temporal em um problema de regressão supervisionado.

O método consiste em tornar cada amostra dependente das N amostras anteriores, e faz isso alocando uma janela de N amostras no início da série temporal e deslocando até o final de amostra em amostra para formar a matriz de características, conforme ilustra a Figure ??.

4.2 Média móvel

4.3 AutoRegression WiSARD

Capítulo 5

Avaliação Experimental

5.1 Ambiente computacional

5.2 Coleções de dados

5.3 Métricas de avaliação

5.4 Resultados

5.4.1 Experimentos com dados sintéticos

5.4.2 Experimentos com dados do covid

5.4.3 Discussão

Capítulo 6

Conclusão

Referências Bibliográficas

- [1] BOX, G. E. P., JENKINS, G. M. *Time Series Analysis: Forecasting and Control*. 3rd ed. USA, Prentice Hall PTR, 1994. ISBN: 0130607746.
- [2] ALZAHRANI, S. I., ALJAMAAN, I. A., AL-FAKIH, E. A. “Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions”, *Journal of Infection and Public Health*, v. 13, n. 7, pp. 914–919, 2020. ISSN: 1876-0341. doi: <https://doi.org/10.1016/j.jiph.2020.06.001>. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1876034120304937>>.
- [3] TAYLOR, S. J., LETHAM, B. “Forecasting at scale”, *PeerJ Preprints*, v. 5, pp. e3190v2, set. 2017. ISSN: 2167-9843. doi: 10.7287/peerj.preprints.3190v2. Disponível em: <<https://doi.org/10.7287/peerj.preprints.3190v2>>.
- [4] ZHANG, G., QI, M. “Neural network forecasting for seasonal and trend time series”, *European Journal of Operational Research*, v. 160, n. 2, pp. 501–514, 2005. ISSN: 0377-2217. doi: <https://doi.org/10.1016/j.ejor.2003.08.037>. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0377221703005484>>. Decision Support Systems in the Internet Age.
- [5] ZHANG, G. “Time series forecasting using a hybrid ARIMA and neural network model”, *Neurocomputing*, v. 50, pp. 159–175, 2003. ISSN: 0925-2312. doi: [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0). Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0925231201007020>>.
- [6] BROCKWELL, P. J., DAVIS, R. A. *Time Series: Theory and Methods*. Berlin, Heidelberg, Springer-Verlag, 1986. ISBN: 0387964061.
- [7] ALEKSANDER, I., DE GREGORIO, M., FRANÇA, F., et al. “A brief introduction to Weightless Neural Systems”. 01 2009.

- [8] BLEDSOE, W. W., BROWNING, I. “Pattern Recognition and Reading by Machine”. In: *Papers Presented at the December 1-3, 1959, Eastern Joint IRE-AIEE-ACM Computer Conference*, IRE-AIEE-ACM '59 (Eastern), p. 225–232, New York, NY, USA, 1959. Association for Computing Machinery. ISBN: 9781450378680. doi: 10.1145/1460299.1460326. Disponível em: <<https://doi.org/10.1145/1460299.1460326>>.
- [9] GRIECO, B., LIMA, P., DE GREGORIO, M., et al. “Producing pattern examples from “mental” images”, *Neurocomputing*, v. 73, pp. 1057–1064, 03 2010. doi: 10.1016/j.neucom.2009.11.015.
- [10] FRANÇA, F., DE GREGORIO, M., LIMA, P., et al. “Advances in Weightless Neural Systems”. 04 2014. doi: 10.13140/2.1.2688.6403.
- [11] LUSQUINO FILHO, L., DE OLIVEIRA, L. F., FILHO, A., et al. “Extending the Weightless WiSARD Classifier for Regression”, *Neurocomputing*, 03 2020. doi: 10.1016/j.neucom.2019.12.134.
- [12] KAPPAUN, A., CAMARGO, K., RANGEL, F., et al. “Evaluating Binary Encoding Techniques for WiSARD”. pp. 103–108, 10 2016. doi: 10.1109/BRACIS.2016.029.

Apêndice A

Título do Apêndice