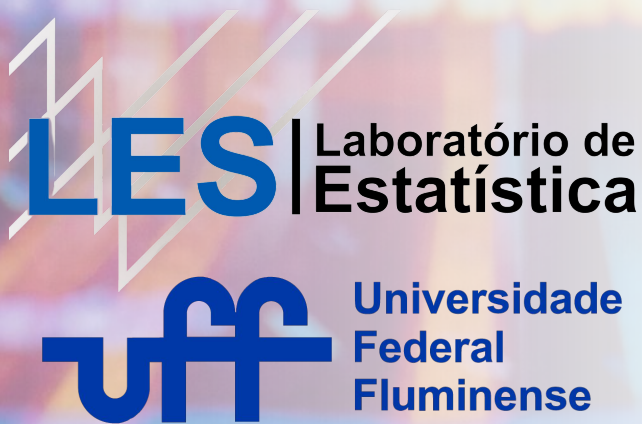


A woman's face is the central focus, overlaid with a complex digital aesthetic. Vertical lines of light in blue, green, and red pass through her face. Various snippets of code are visible, including 'elif_opd', 'mirror', '#sele', 'mirror_e', 'modifier_on', 'bpy.context', and 'print'. Binary code (0s and 1s) is scattered throughout the background, particularly on the left side.

Redes Neurais

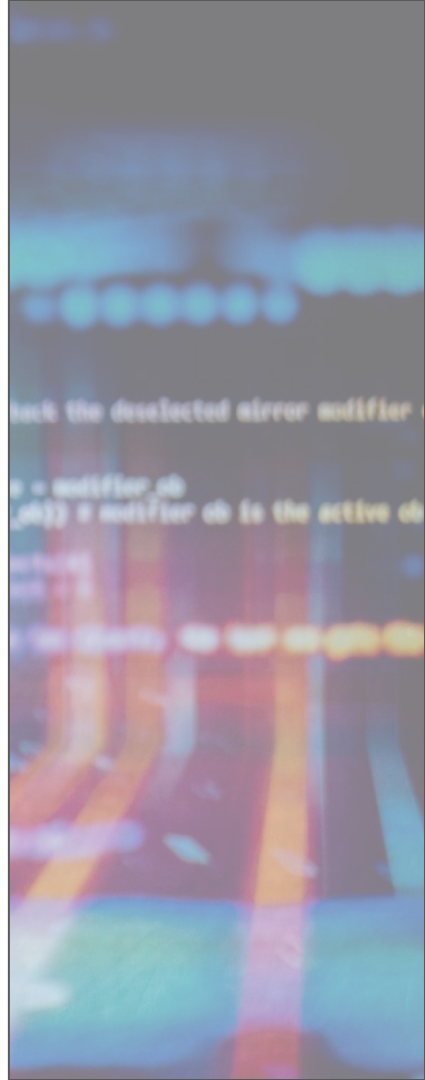
Escola de Verão



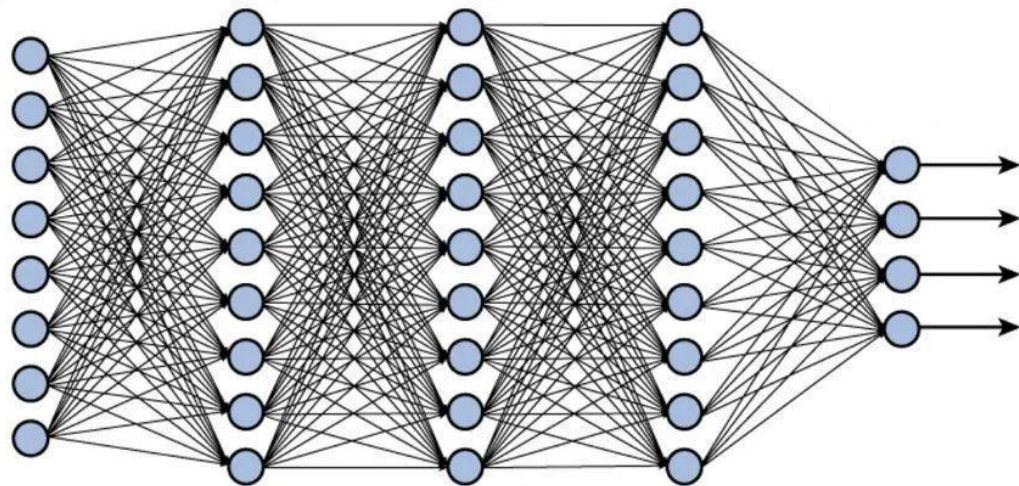
Redes Neurais

Neste curso serão respondidas algumas perguntas de forma geral e apresentadas as ferramentas necessárias para que o aluno possa se aprofundar nas respostas por conta própria após o término do curso.

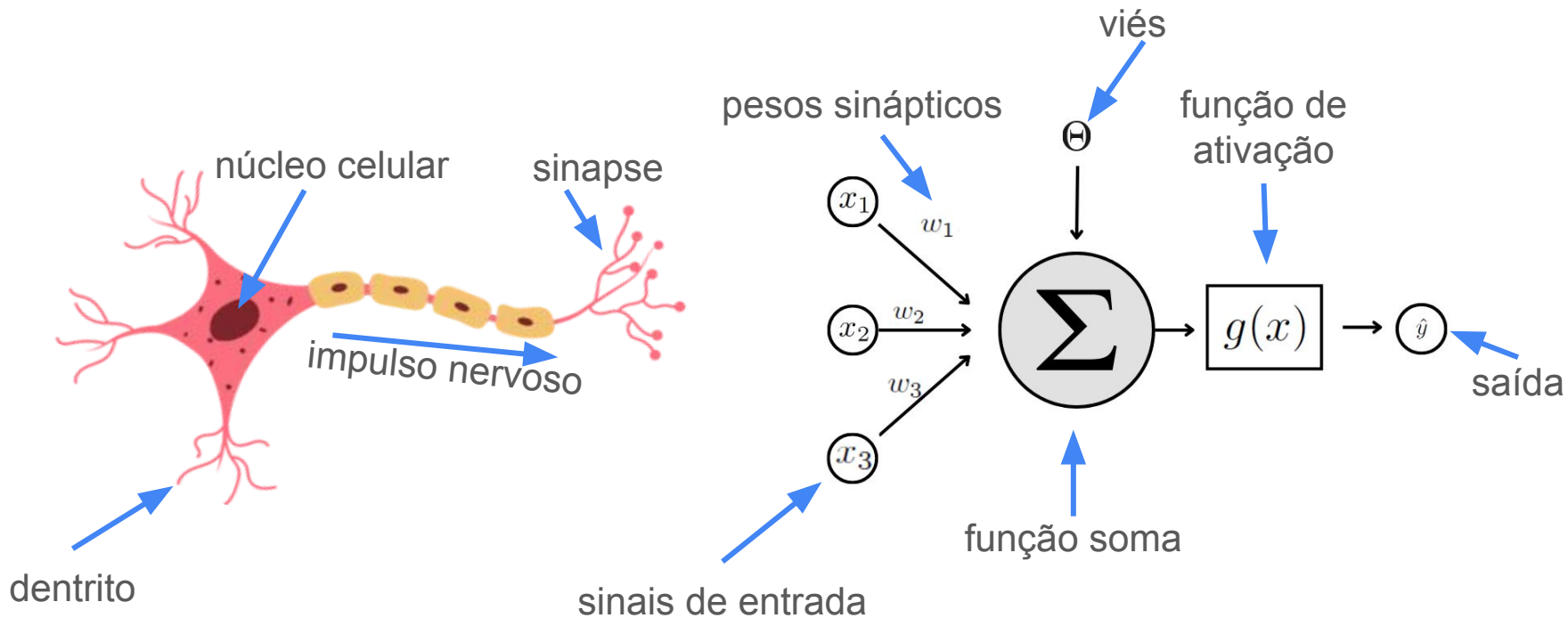
- O que são redes neurais?
- Para que elas servem?
- Quais as principais arquiteturas?
- O que significa "treinar" uma rede neural?
- Como posso treinar uma rede neural no R?

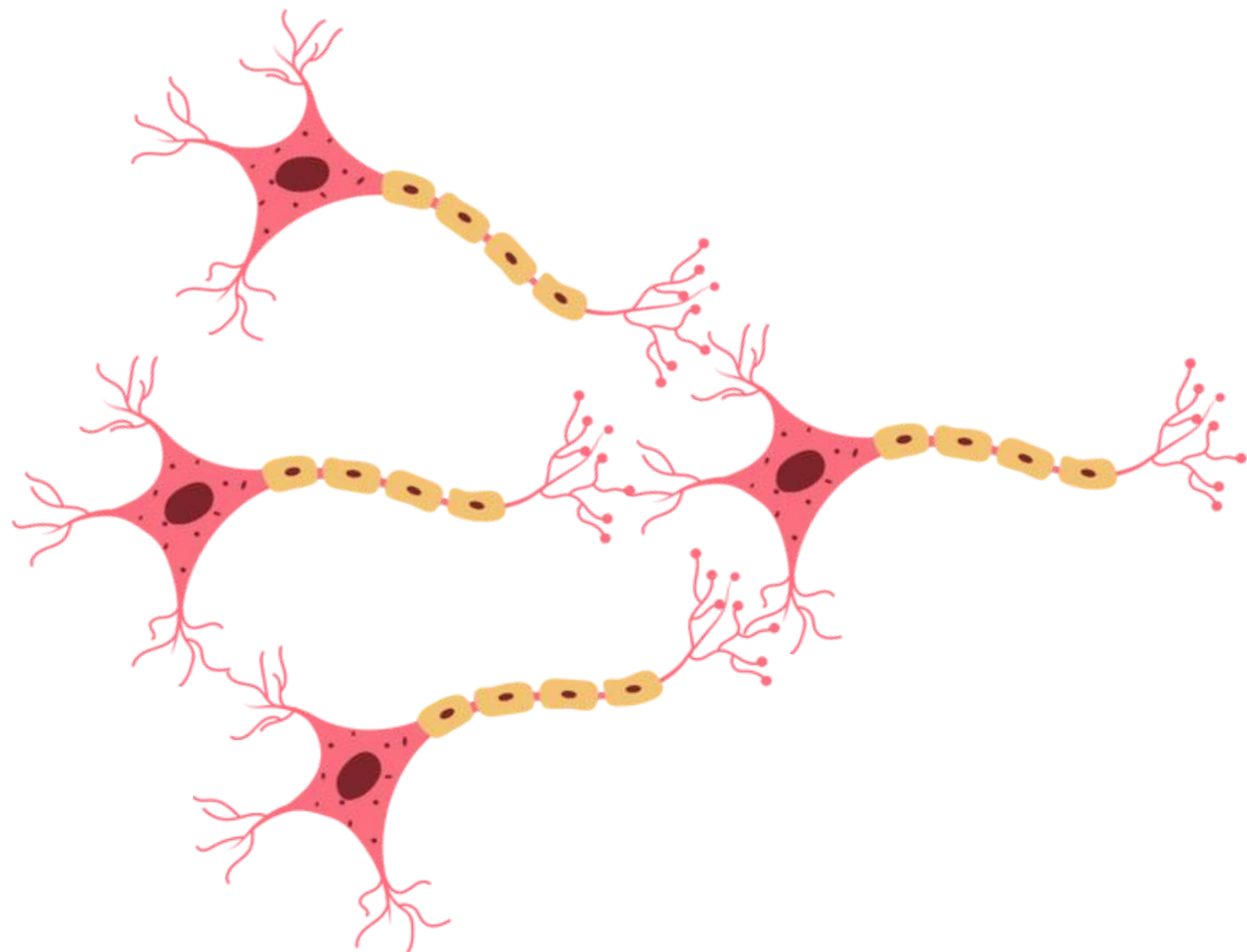


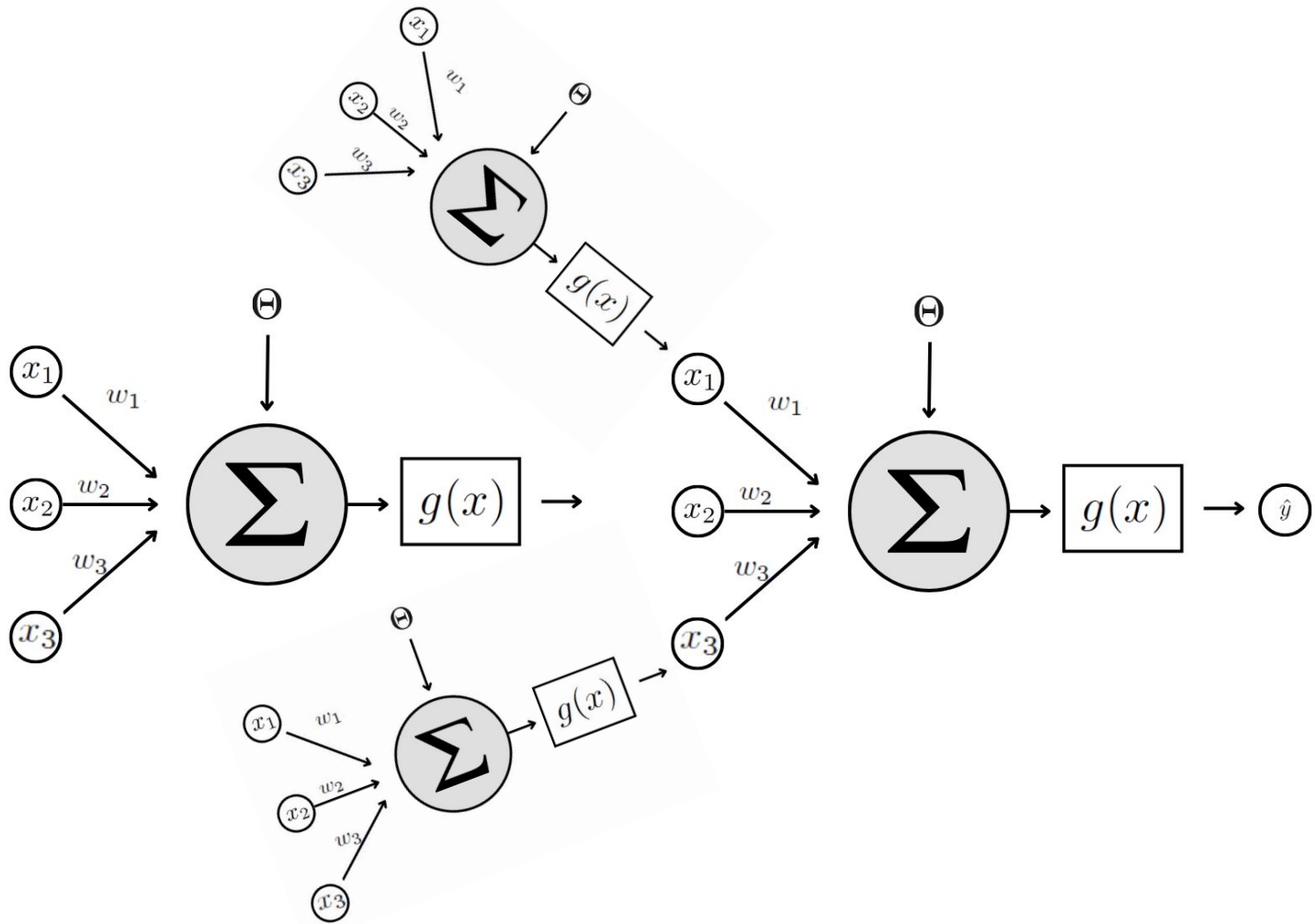
O que são redes neurais?



O que são redes neurais?









Para que elas servem?

- Previsão / Regressão
- Classificação
- Reconhecimento de Imagem
- Reconhecimento de voz
- Análise de Clusters
- Previsão de Séries Temporais

Previsão / Regressão

“Predição do peso molecular de fluidos petrolíferos por correlações empíricas e redes de neurônios artificiais”

A exatidão das correlações de peso molecular de fluidos de petróleo afeta significativamente a precisão dos cálculos de engenharia de petróleo e pode fazer projetos de processos e solução de problemas impreciso. (...) Além das correlações existentes na literatura, duas diferentes técnicas, regressão não linear e rede neural artificial (RNA), foram empregadas para modelar o peso molecular das 430 amostras de fluidos de petróleo. Verificou-se que o modelo RNA demonstrou a melhor precisão de previsão com um erro padrão relativo (RSE) de 7,2%, seguido pelo novo desenvolveram correlação de regressão não linear com RSE de 10,9%. (...) As bem conhecidas correlações de peso molecular de Lee-Kesler e Twu, para o conjunto de dados de 430 pontos de dados exibiu RSEs de 26,5 e 30,3%, respectivamente.

Article

Prediction of Molecular Weight of Petroleum Fluids by Empirical Correlations and Artificial Neuron Networks

Dicho Stratiev ^{1,2,*}, Sotir Sotirov ³, Evdokia Sotirova ³, Svetoslav Nenov ⁴, Rosen Dinkov ¹, Ivelina Shishkova ¹, Iliyan Venkov Kolev ^{1,2}, Dobromir Yordanov ³, Svetlin Vasilev ³, Krassimir Atanasov ^{2,3}, Stanislav Simeonov ³ and Georgi Nikolov Palichev ²

¹ LUKOIL Neftohim Burgas, 8104 Burgas, Bulgaria

² Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Georgi Bonchev 105, 1113 Sofia, Bulgaria

³ Intelligent Systems Laboratory, Department Industrial Technologies and Management, University Prof. Dr. Assen Zlatarov, 8010 Burgas, Bulgaria

⁴ Department of Mathematics, University of Chemical Technology and Metallurgy, Kliment Ohridski 8, 1756 Sofia, Bulgaria

* Correspondence: stratiev.dicho@neftochim.bg

Abstract: The exactitude of petroleum fluid molecular weight correlations affects significantly the precision of petroleum engineering calculations and can make process design and trouble-shooting inaccurate. Some of the methods in the literature to predict petroleum fluid molecular weight are used in commercial software process simulators. According to statements made in the literature, the correlations of Lee–Kesler and Twu are the most used in petroleum engineering, and the other methods do not exhibit any significant advantages over the Lee–Kesler and Twu correlations. In order to verify which of the proposed in the literature correlations are the most appropriate for petroleum fluids with molecular weight variation between 70 and 1685 g/mol, 430 data points for boiling point, specific gravity, and molecular weight of petroleum fluids and individual hydrocarbons were extracted from 17 literature sources. Besides the existing correlations in the literature, two different techniques, nonlinear regression and artificial neural network (ANN), were employed to model the molecular weight of the 430 petroleum fluid samples. It was found that the ANN model demonstrated the best accuracy of prediction with a relative standard error (RSE) of 7.2%, followed by the newly developed nonlinear regression correlation with an RSE of 10.9%. The best available molecular weight correlations in the literature were those of API (RSE = 12.4%), Goosens (RSE = 13.9%); and Riazi and Daubert (RSE = 15.2%). The well known molecular weight correlations of Lee–Kesler, and Twu, for the data set of 430 data points, exhibited RSEs of 26.5, and 30.3% respectively.

Keywords: petroleum; molecular weight; modeling; artificial neural network; nonlinear regression; empirical correlation



Citation: Stratiev, D.; Sotirov, S.; Sotirova, E.; Nenov, S.; Dinkov, R.; Shishkova, I.; Kolev, I.V.; Yordanov, D.; Vasilev, S.; Atanasov, K.; et al. Prediction of Molecular Weight of Petroleum Fluids by Empirical Correlations and Artificial Neuron Networks. *Processes* **2023**, *11*, 426. <https://doi.org/10.3390/pr11020426>

Academic Editor: Qingbang Meng

Previsão / Regressão

Article

Prediction of Molecular Weight of Petroleum Fluids by

Empirical Correlations and Artificial Neural Networks

Table 2. Statistical analysis of studied methods to predict petroleum fluid molecular weight for the whole range of studied molecular weights.

	Standard Error	Rel. St. Error	Sum of Squared Errors	%AAD	SRE
ANN	23.0	7.2	1.6	4.3	−126.8
New empirical correlation (this work)	34.8	10.9	3.4	6.4	695.5
Hosseiniifar (2021) [10]	84.4	26.6	8.6	9.4	−1851.7
Goosens (1996) [9]	44.1	13.9	4.5	7.6	1970.2
Riazi and Daubert $MW \leq 700$ g/mol (2005) [18]	48.3	15.2	3.9	6.3	358.2
API (2011) [23]	39.3	12.4	5.1	8.6	−999.1
Liñan (2011) [23]	54.4	17.1	10.2	11.3	−2973.1
Twu (1984) [20]	96.4	30.3	20.6	18.7	7840.0
Lee–Kesler (1976) [14]	84.2	26.5	7.1	9.7	345.9
Riazi and Daubert ≤ 300 g/mol (2005) [18]	117.3	36.9	12.4	12.5	4155.7
Lemus et al. (2016) [2]	103.0	32.4	14.0	10.8	864.9
Soreide (1989) [12]	84.1	26.5	12.2	9.4	−598

“Previsão de peso molecular de petróleo”

A exactidão da previsão de peso molecular de petróleo por meio de modelos empíricos e de redes neurais artificiais (ANN) foi avaliada. Os resultados mostraram que a ANN apresentou o menor erro padrão (23,0) e o menor erro relativo (7,2%). A regressão não linear também apresentou bons resultados, com erro padrão de 34,8 e erro relativo de 10,9%. As correlações empíricas existentes na literatura, como a de Riazi e Daubert (2005) e a de API (2011), apresentaram erros padrão mais altos, variando de 39,3 a 117,3, e erros relativos de 12,4% a 36,9%.

Conhecidas correlações de peso molecular de Lee–Kesler e Twu, para o conjunto de dados de 430 pontos de dados exibiu RSEs de 26,5 e 30,3%, respectivamente.

Prediction of Molecular Weight of Petroleum Fluids by Empirical Correlations and Artificial Neural Networks. *Processes* **2023**, *11*, 426.
<https://doi.org/10.3390/pr11020426>
 Academic Editor: Qingbang Meng

correlations in the literature were those of API (RSE = 12.4%), Goosens (RSE = 13.9%), and Riazi and Daubert (RSE = 15.2%). The well known molecular weight correlations of Lee–Kesler, and Twu, for the data set of 430 data points, exhibited RSEs of 26.5, and 30.3% respectively.

Keywords: petroleum; molecular weight; modeling; artificial neural network; nonlinear regression; empirical correlation

Ishkova ¹,
 Leonov ³

Bonchev 105,
 University Prof. Dr.
 Kirilski 8,

significantly the
 bubble-shooting
 molecular weight are
 the literature,
 and the other
 correlations. In order
 for petroleum
 fluids for boiling
 point carbons were
 two different
 models to model the
 demonstrated
 by the newly
 molecular weight

Classificação

“Aplicação de modelos de Regressão Logística Regularizada e Rede Neural Artificial para Classificação de Ozônio em todo o Condado de El Paso, Texas, Estados Unidos”

Este artigo foca na previsão de ozônio na atmosfera usando uma abordagem de aprendizado de máquina. Nós utilizamos conjuntos de dados de poluentes atmosféricos e variáveis meteorológicas da região de El Paso para classificar os níveis de ozônio como altos ou baixos. Os algoritmos LR e ANN são empregados para treinar os conjuntos de dados. (...) Com base em métricas como precisão, taxa de erro, perda de log e tempo de previsão, o modelo ANN se mostra mais rápido e mais adequado para a classificação de ozônio na área de El Paso, Texas.

Application of Regularized Logistic Regression and Artificial Neural Network model for Ozone Classification across El Paso County, Texas, United States

Callistus Obunadike, Adekunle Adefabi, Somtobe Olisah, David Abimbola, Kunle Oloyede

Department of Computer Science, Austin Peay State University, Clarksville, USA

Email: callistusobunadike@gmail.com, aadefabi@my.apsu.edu, solisah@my.apsu.edu, dabimbola@my.apsu.edu, koloyede@my.apsu.edu

How to cite this paper: Obunadike, C., Adefabi, A., Olisah, S., Abimbola, D. and Oloyede, K. (2023) Application of Regularized Logistic Regression and Artificial Neural Network model for Ozone Classification across El Paso County, Texas, United States. *Journal of Data Analysis and Information Processing*, 11, 217-239. <https://doi.org/10.4236/jdaip.2023.113012>

Received: April 25, 2023

Accepted: July 8, 2023

Published: July 11, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This paper focuses on ozone prediction in the atmosphere using a machine learning approach. We utilize air pollutant and meteorological variable datasets from the El Paso area to classify ozone levels as high or low. The LR and ANN algorithms are employed to train the datasets. The models demonstrate a remarkably high classification accuracy of 89.3% in predicting ozone levels on a given day. Evaluation metrics reveal that both the ANN and LR models exhibit accuracies of 89.3% and 88.4%, respectively. Additionally, the AUC values for both models are comparable, with the ANN achieving 95.4% and the LR obtaining 95.2%. The lower the cross-entropy loss (log loss), the higher the model's accuracy or performance. Our ANN model yields a log loss of 3.74, while the LR model shows a log loss of 6.03. The prediction time for the ANN model is approximately 0.00 seconds, whereas the LR model takes 0.02 seconds. Our odds ratio analysis indicates that features such as “Solar radiation”, “Std. Dev. Wind Direction”, “outdoor temperature”, “dew point temperature”, and “PM10” contribute to high ozone levels in El Paso, Texas. Based on metrics such as accuracy, error rate, log loss, and prediction time, the ANN model proves to be faster and more suitable for ozone classification in the El Paso, Texas area.

Classificação

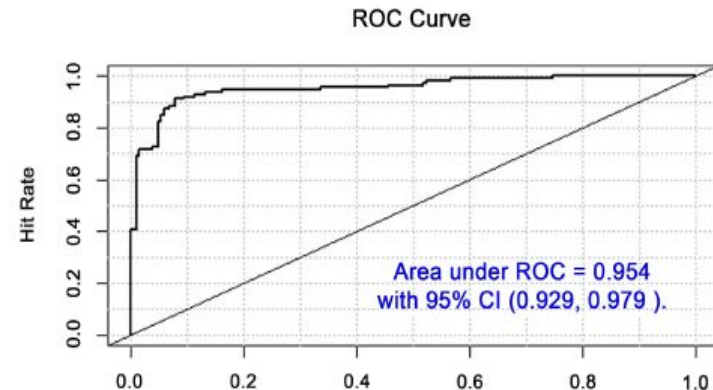
Application of Regularized Logistic Regression and Artificial Neural Network model for Ozone

Table 9. Results of the ANN model.

Miss-classification rate	MSE	cvAUC	SE	CI	Confidence
0.107	0.0833	0.954	0.0126	0.929, 0.979	0.95

Table 10. Confusion matrix and other statistical prediction parameters for ANN.

Confusion Matrix and Statistics	
Accuracy	0.893
95%CI	(0.854, 0.925)
Sensitivity/Recall	0.802
Specificity (True Negative Rate/TNR)	0.957
Pos Pred Value/Precision	0.929
F1 Score	0.861
Prediction Time	0.00 secs
Binary Cross Entropy/Log Loss	3.74



a machine
table data-
the LR and
monstrate
zone levels
LR models
the AUC
95.4% and
, the high-
log loss of
me for the
takes 0.02
olar radi-
point tem-
so, Texas.
ction time,
ssification

Previsão de Séries Temporais

"Comparação da performance dos modelos LSTM e ARIMA para a Previsão de Curto Prazo dos Preços do Bitcoin"

Esta pesquisa avalia a previsão dos preços do Bitcoin utilizando os modelos autoregressivos integrados de médias móveis (ARIMA) e memória de curto e longo prazo (LSTM). Preveremos o preço do Bitcoin para o dia seguinte utilizando o método de previsão estática, com e sem a reestimação do modelo de previsão a cada etapa. Consideramos duas amostras de treinamento e teste diferentes para a validação cruzada dos resultados da previsão. Na primeira amostra de treinamento, o ARIMA supera o LSTM, mas na segunda amostra de treinamento, o LSTM supera o ARIMA. Além disso, nos dois períodos de previsão de amostras de teste, o LSTM com reestimação do modelo a cada etapa supera o ARIMA. Comparando o LSTM com o ARIMA, as previsões foram muito mais próximas dos preços históricos reais. Ao contrário do ARIMA, que só podia acompanhar a tendência dos preços do Bitcoin, o modelo LSTM foi capaz de prever tanto a direção quanto o valor durante o período de tempo especificado. Esta pesquisa mostra a capacidade persistente do LSTM para previsão de preços do Bitcoin, apesar da sofisticação do ARIMA.



Comparative Performance of LSTM and ARIMA for the Short-Term Prediction of Bitcoin Prices

Navmeen Latif¹, Joseph Durai Selvam², Manohar Kapse³, Vinod Sharma⁴ and Vaishali Mahajan⁵

Abstract

This research assesses the prediction of Bitcoin prices using the autoregressive integrated moving average (ARIMA) and long-short-term memory (LSTM) models. We forecast the price of Bitcoin for the following day using the static forecast method, with and without re-estimating the forecast model at each step. We take two different training and test samples into consideration for the cross-validation of forecast findings. In the first training sample, ARIMA outperforms LSTM, but in the second training sample, LSTM exceeds ARIMA. Additionally, in the two test-sample forecast periods, LSTM with model re-estimation at each step surpasses ARIMA. Comparing LSTM to ARIMA, the forecasts were much closer to the actual historical prices. As opposed to ARIMA, which could only track the trend of Bitcoin prices, the LSTM model was able to predict both the direction and the value during the specified time period. This research exhibits LSTM's persistent capacity for fluctuating Bitcoin price prediction despite the sophistication of ARIMA.



Comparative Performance of LSTM and ARIMA for the Short-Term Prediction of Bitcoin Prices

Latif¹, Joseph Durai Selvam², Manohar Kapse³, Vinod Sharma⁴ and Vaishali

LSTM - Actual Vs Predicted

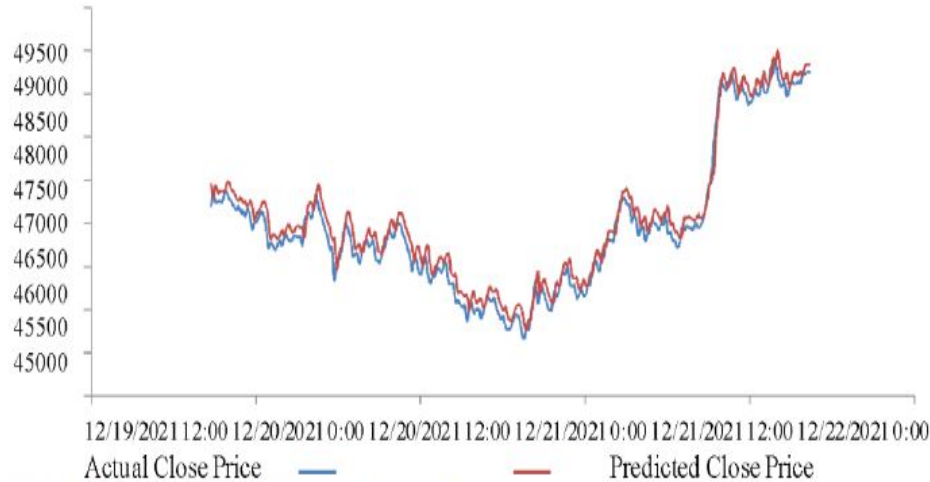


FIGURE 16: Actual vs. Predicted Prices of LSTM

TABLE 2

Performance Measurement Metrics

	MAE	MAPE	RMSE	Accuracy
ARIMA (3,1,3)	837.77	1.79%	940.40	98.21%
RNN-LSTM	126.97	0.27%	151.95	99.73%

Bitcoin, apesar da sophistication do ARIMA.

Bitcoin prices, the LSTM model was able to predict both the direction and the value during the specified time period. This research exhibits LSTM's persistent capacity for fluctuating Bitcoin price prediction despite the sophistication of ARIMA.

Processamento de Linguagem Natural

Sigmae, Alfenas, v.13, n.1, p.79-90. 2024.

XVI Encontro Mineiro de Estatística - MGEST, Juiz de Fora, MG.

ISSN: 2317-0840

Comparação de métodos de vetorização de documentos: um estudo de caso com dados textuais

Resumo: *A explosão de informações digitais nas últimas décadas trouxe um enorme volume de dados em forma de texto. O interesse em extrair conhecimento desta vasta quantidade de dados originou a Mineração de Texto. Um dos desafios nesta área é transformar um banco de textos em uma base de dados numérica. Esse processo, chamado de vetorização de documentos, é fundamental para a automatização da extração de informação. O objetivo deste trabalho é comparar o desempenho de quatro métodos de vetorização de documentos quando utilizados para fins de classificação. Os métodos de vetorização comparados foram: BoW, TF-IDF e as duas arquiteturas diferentes do doc2vec, CBOW e skip. Os métodos de classificação aplicados foram: Regressão Logística, Árvore de Classificação, Floresta Aleatória, XGBoost e Perceptron. A base de dados foi a base pública The Women's E-Commerce Clothing Reviews, composta por 10 atributos, entre os quais 3 deles foram considerados neste trabalho: o texto de avaliação do item, o título da avaliação e uma variável categórica que indica se o cliente recomenda ou não o produto. Uma amostra aleatória balanceada de 8.000 documentos, 4.000 documentos com recomendação positiva e 4.000 com recomendação negativa, foi sorteada e dividida em treino (70%) e teste (30%). A medida de comparação de desempenho foi a área embaixo da curva ROC (AUC). Quando comparados os métodos de vetorização de documentos, as duas arquiteturas do doc2vec apresentaram resultados superiores às demais em todos os métodos de classificação testados.*

Palavras-chave: *Mineração de Texto; Doc2Vec; TF-IDF; Métodos de Classificação.*

Figure 1: A text preprocessing example.

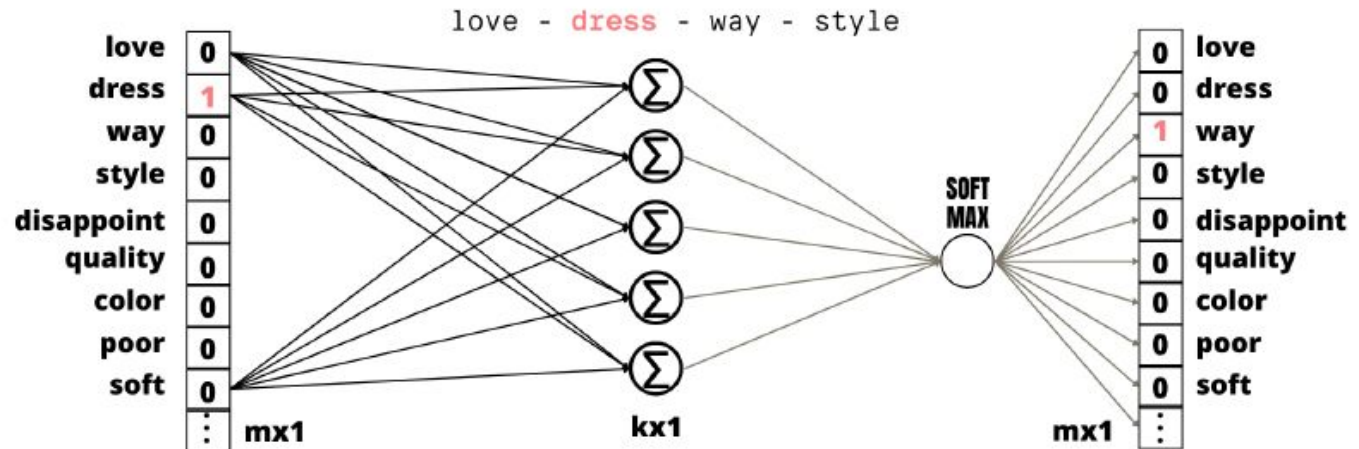
Processamento Linguagem Natural

word2vec

1. I love this dress! There are so many ways to style it! Love, love, love!
2. Disappointed in the quality of the dress. Love the style and the colors, but the quality is poor.
3. Soft, comfortable and stylish. The color is just as pictured online.
4. I just tried on this dress in the store and i loved the off the shoulder design. My favorite dress!



1. love - dress - way - style - love - love - love
2. disappoint - quality - dress - love - style - color - quality - poor
3. soft - comfort - style - color - picture - online
4. try - dress - store - love - shoulder - desing - favorit - dress



Processamento de Linguagem Natural

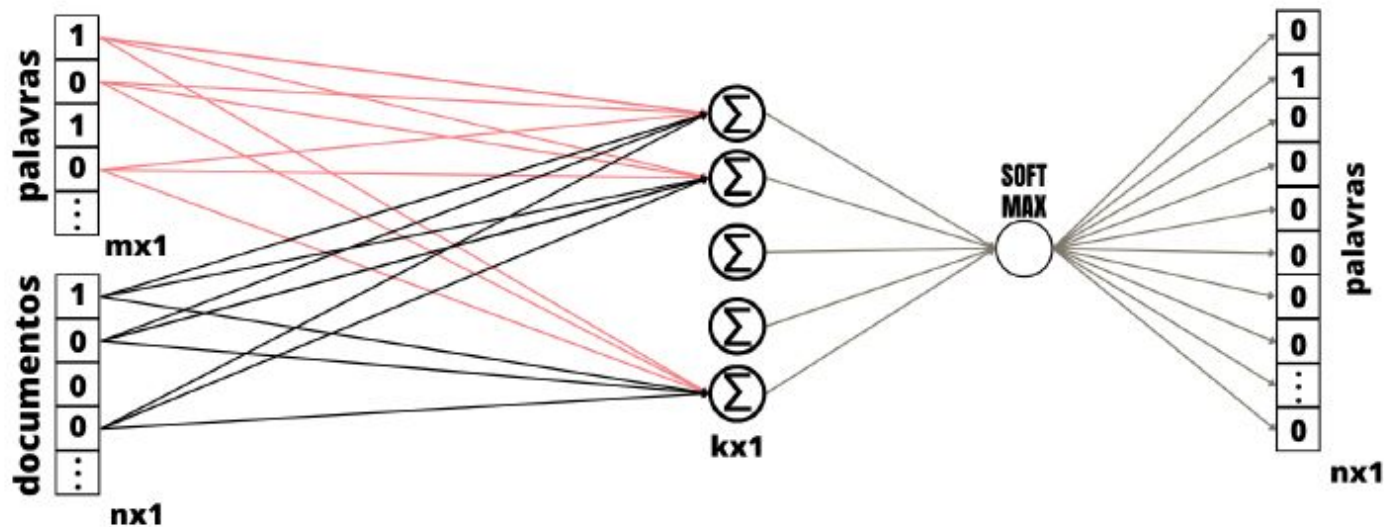
<https://contexto.me/>

word2vec



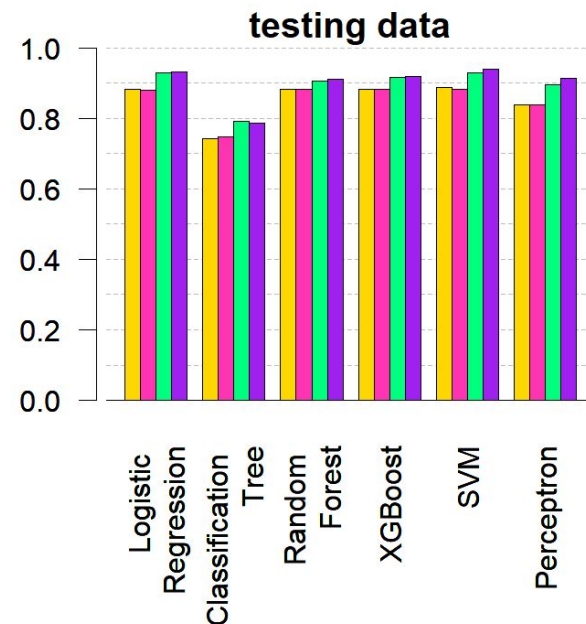
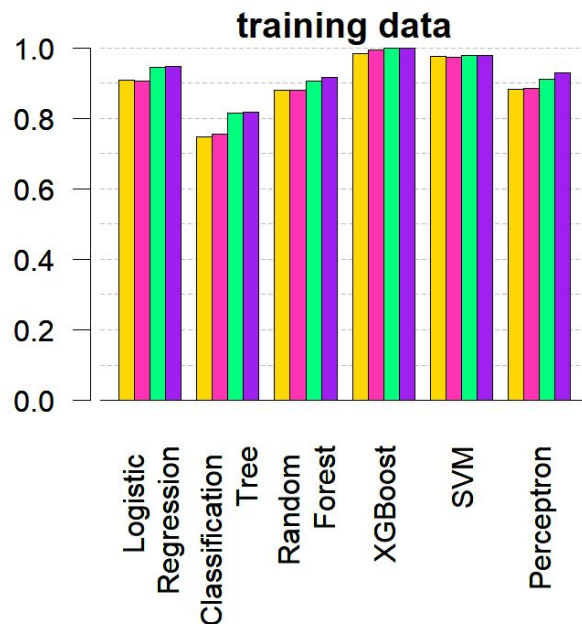
Processamento de Linguagem Natural

doc2vec



Processamento de Linguagem Natural

BoW
TF-IDF
Doc2Vec-CBOW
Doc2Vec-Skip-gram



Quais as principais arquiteturas?

- Redes Neurais Artificiais (RNA / ANN)
- Redes Neurais Recorrentes (RNR / RNN)
- Redes Neurais Convolucionais (RNC / CNN)
- Mapas auto-organizáveis (SOM)
- Long Short Term Memory (LSTM)

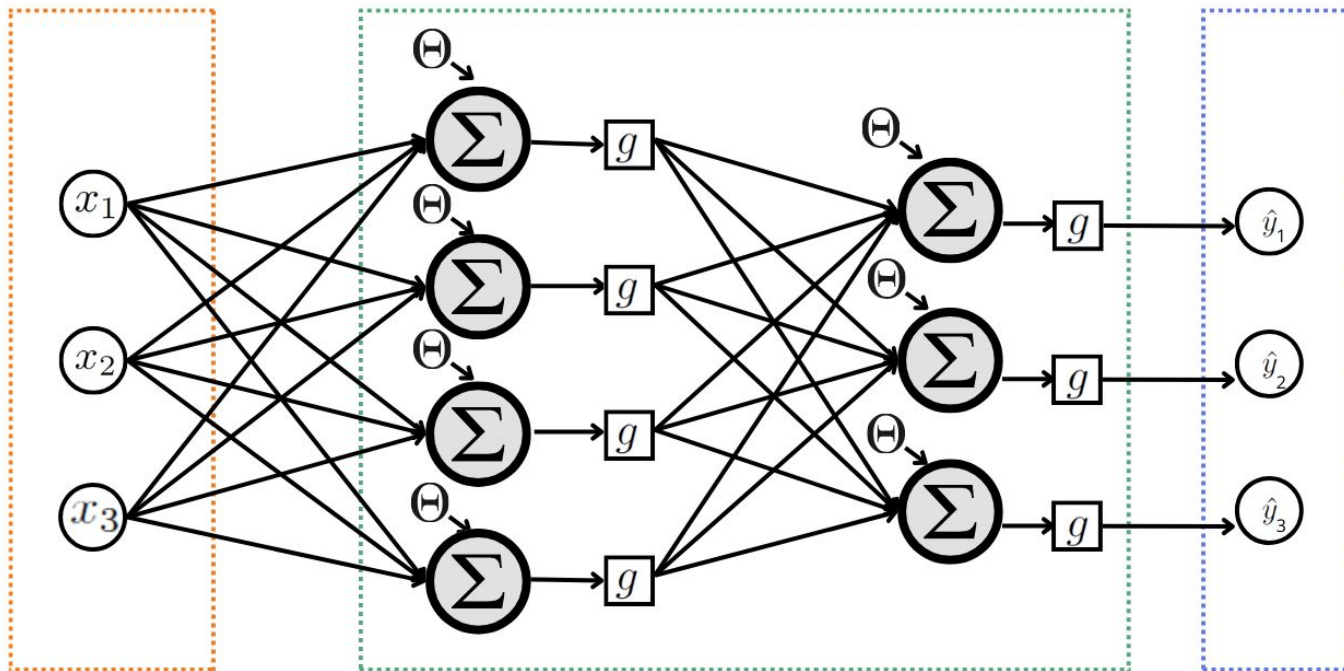


Redes Neurais Artificiais

Camada de
entrada

Camadas
ocultas

Camada de
saída

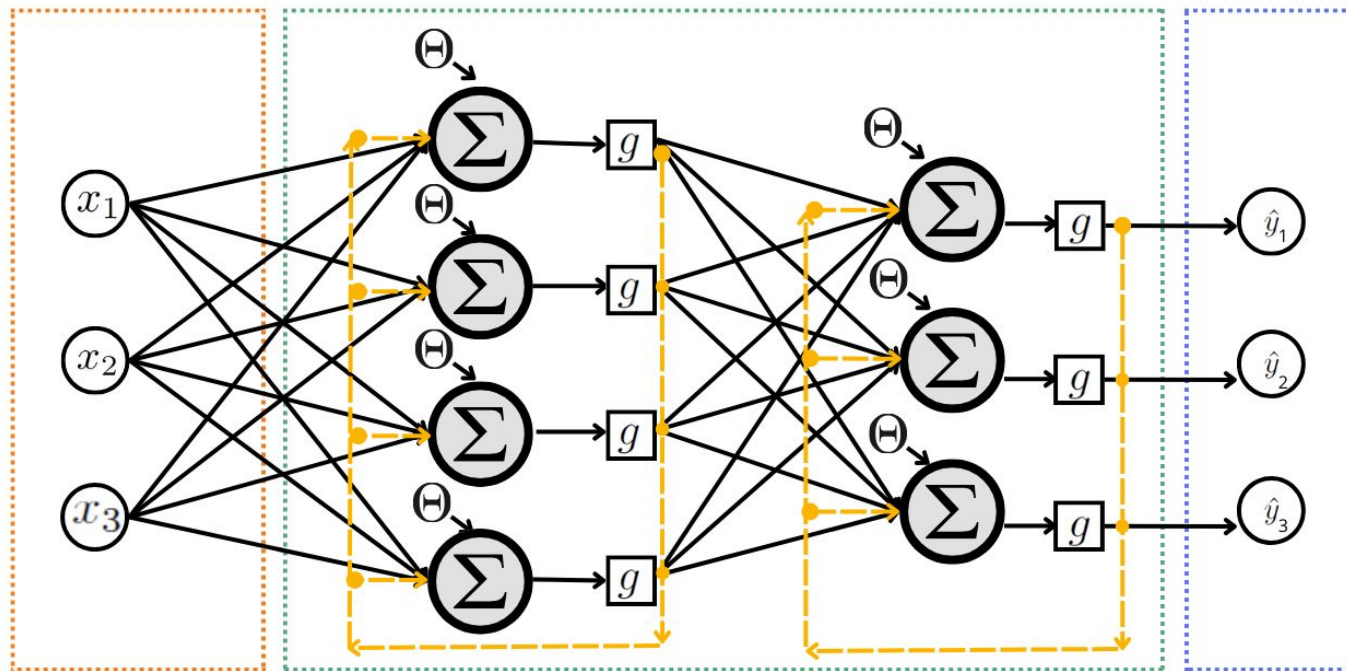


Redes Neurais Recorrentes

Camada de
entrada

Camadas
ocultas

Camada de
saída



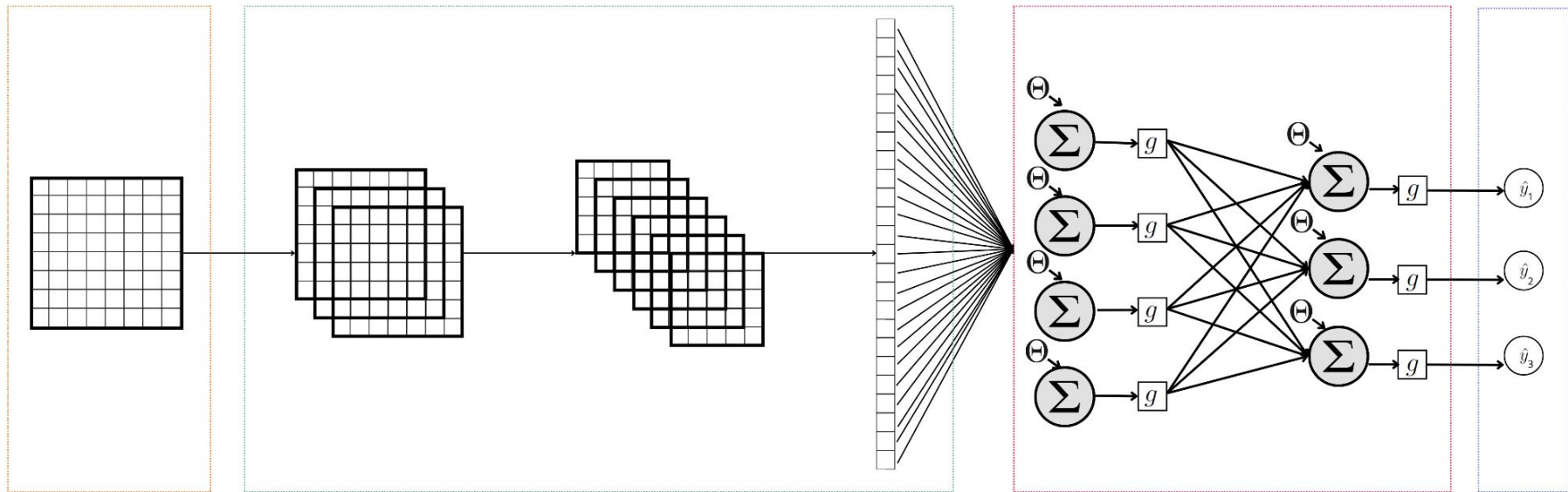
Redes Neurais Convolucionais

Camada de
entrada

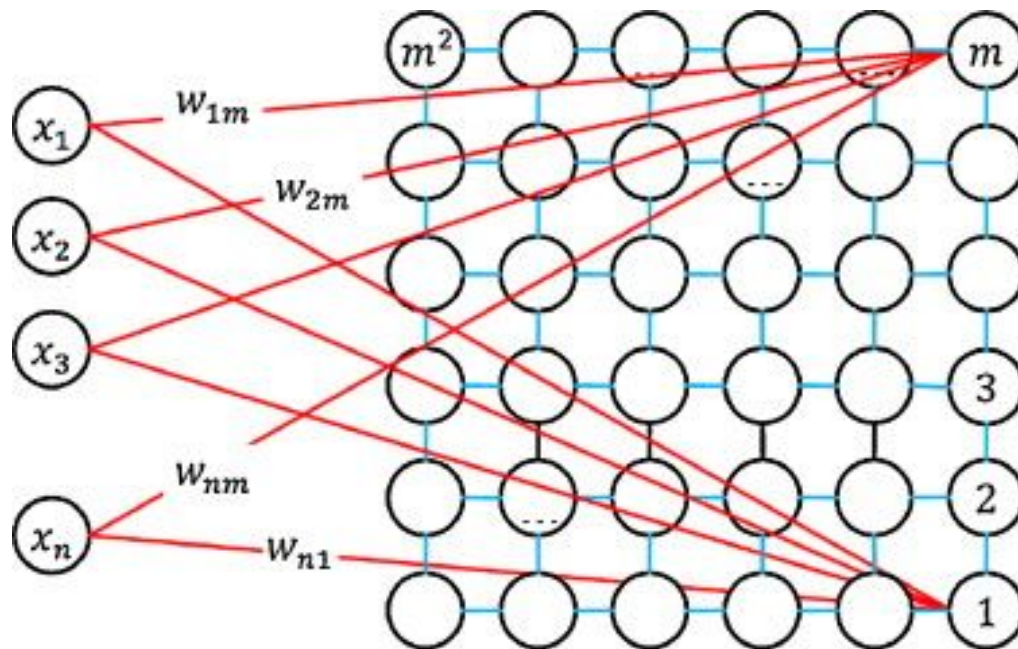
Camadas de Aprendizado Profundo

Camadas Totalmente
Conectadas

Camada de
saída

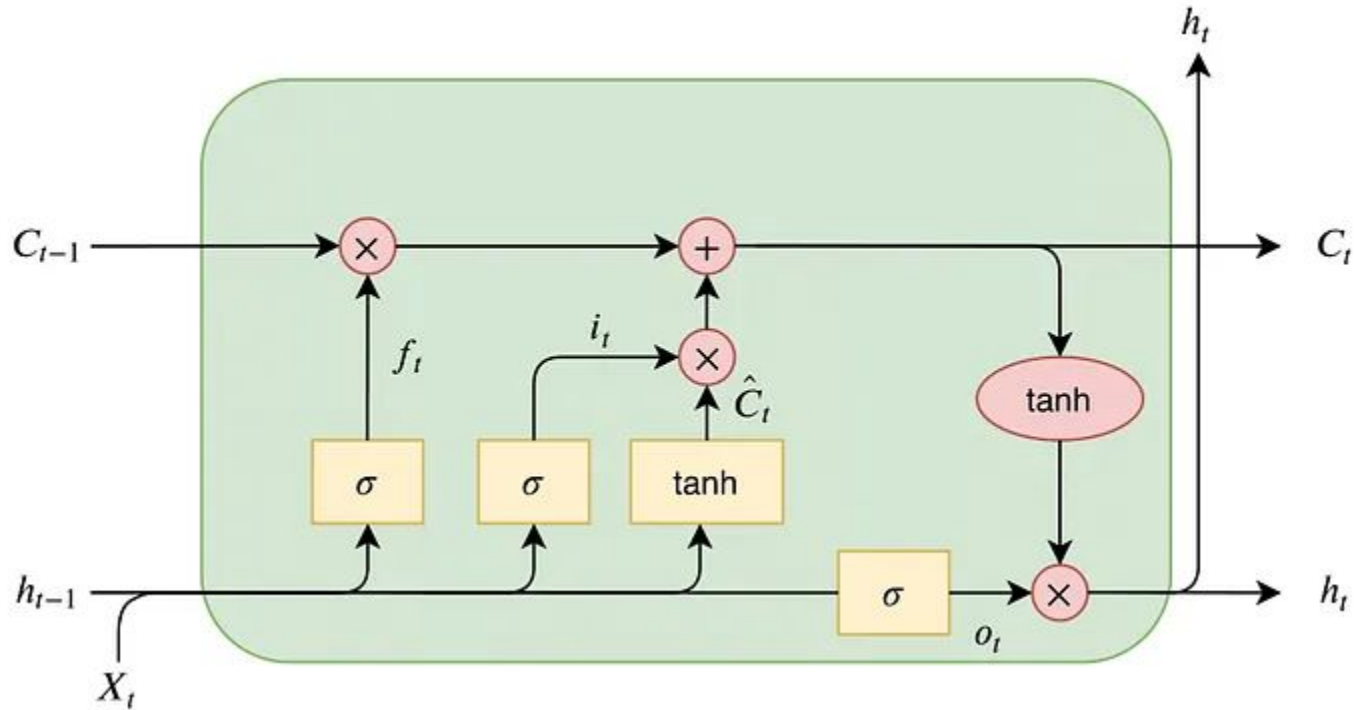


Mapas auto-organizáveis



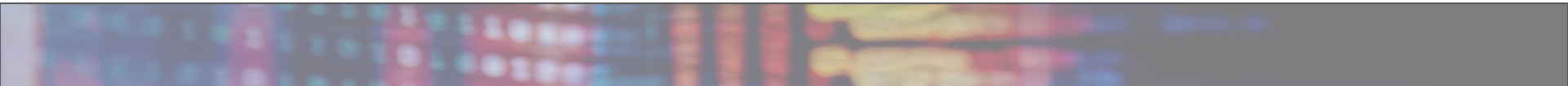
Ghaseminezhad, M. H., and Ali Karami. "A novel self-organizing map (SOM) neural network for discrete groups of data clustering." *Applied soft computing* 11.4 (2011): 3771-3778.

Long Short Term Memory (LSTM)

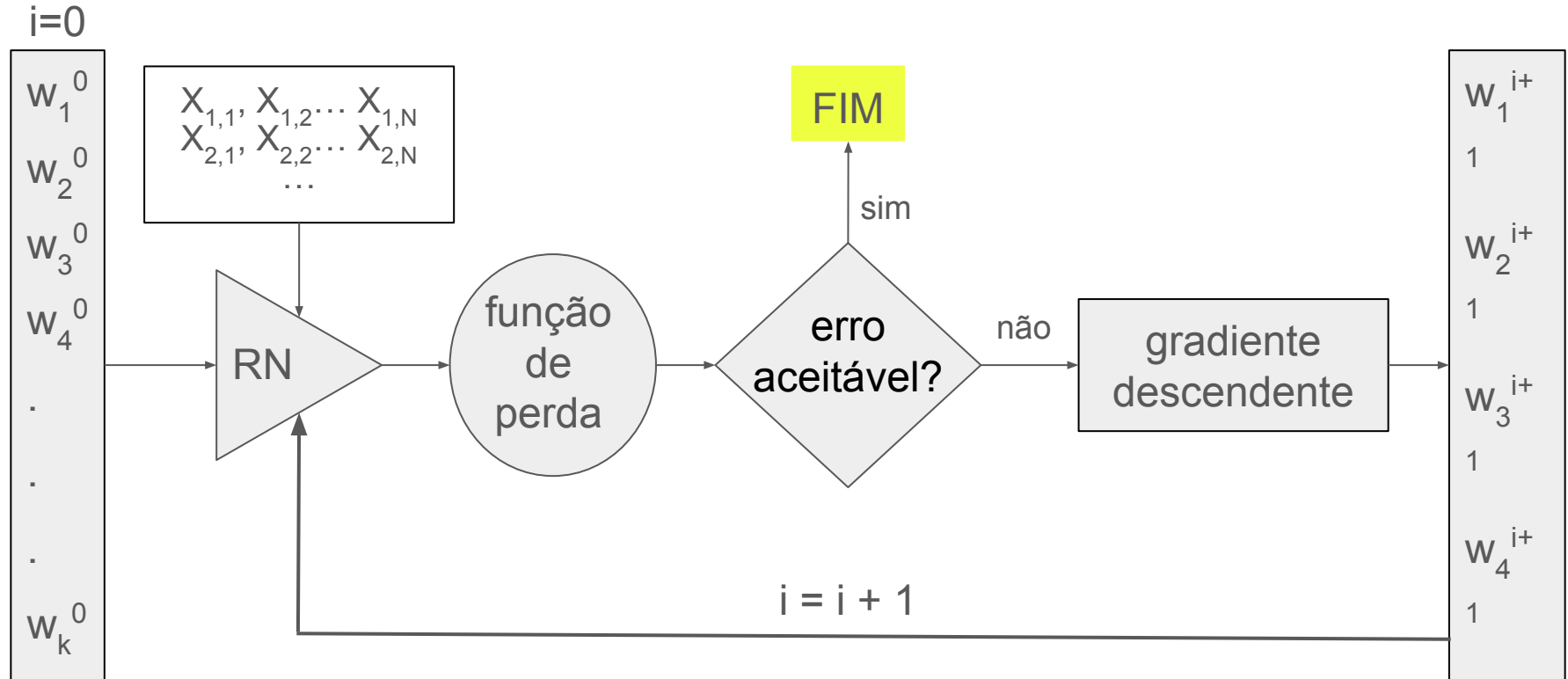


O que significa "treinar" uma rede neural?

- Ajustar os pesos;
- Apresenta-se à rede um conjunto de exemplos de entrada e as saídas correspondentes;
- Processo feito de forma iterativa (gradiente descendente);
- Objetivo: minimizar a diferença entre as saídas previstas e as saídas reais (função de perda);

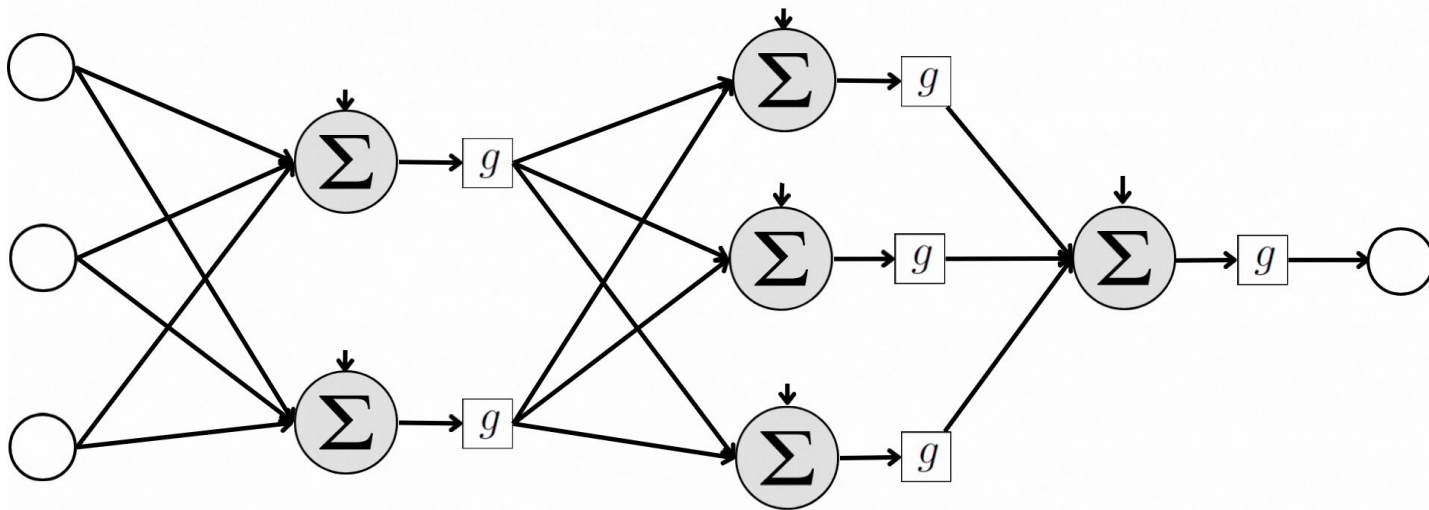


O que significa "treinar" uma rede neural?



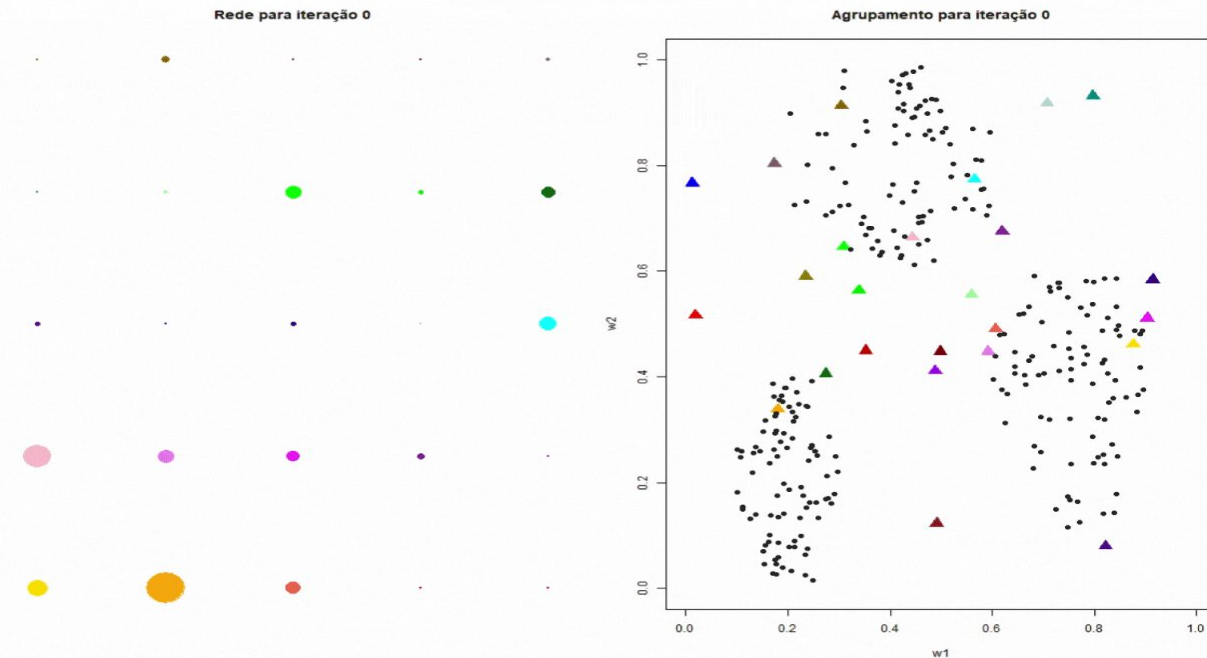
O que significa "treinar" uma rede neural?

Exemplo 1: retropropagação



O que significa "treinar" uma rede neural?

Exemplo 1: som



Como posso treinar uma rede neural no R?

https://github.com/jessicakubrusly/Escola_de_Verao_2024_LES_Neural_Neural_Neural