

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
BACHARELADO EM MATEMÁTICA APLICADA

Redes Neurais aplicadas

Eduardo Galvani Massino

MONOGRAFIA FINAL
MAP2010 — TRABALHO DE
FORMATURA

Orientador: José Coelho de Pina Junior

São Paulo
Dezembro de 2020

*Esta seção é opcional e fica numa página separada;
ela pode ser usada para uma dedicatória ou epígrafe.*

[illegible]

Resumo

Eduardo Galvani Massino. **Redes Neurais aplicadas**. Monografia (Bacharelado). Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2020.

[illegible]

Palavras-chave: redes-neurais. perceptron.

Abstract

Eduardo Galvani Massino. **Redes Neurais aplicadas**. Capstone Project Report (Bachelor). Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2020.

[illegible]

Keywords: neural-nets. perceptron.

Lista de Abreviaturas

CFT	Transformada contínua de Fourier (<i>Continuous Fourier Transform</i>)
DFT	Transformada discreta de Fourier (<i>Discrete Fourier Transform</i>)
EIIP	Potencial de interação elétron-íon (<i>Electron-Ion Interaction Potentials</i>)
STFT	Transformada de Fourier de tempo reduzido (<i>Short-Time Fourier Transform</i>)
ABNT	Associação Brasileira de Normas Técnicas
URL	Localizador Uniforme de Recursos (<i>Uniform Resource Locator</i>)
IME	Instituto de Matemática e Estatística
USP	Universidade de São Paulo

Lista de Símbolos

ω	Frequência angular
ψ	Função de análise <i>wavelet</i>
Ψ	Transformada de Fourier de ψ

Lista de Figuras

2.1	Rede neural simples, o perceptron de camada única.	7
2.2	Rede neural mais simples ainda, apenas um neurônio oculto.	8

Lista de Tabelas

Lista de Programas

Sumário

1	Introdução	1
2	Redes neurais artificiais	5
2.1	Aprendizado de máquina supervisionado	5
2.2	Aprendizado não-supervisionado	6
2.3	Técnicas de classificação	6
2.4	Redes neurais artificiais	7
2.5	Perceptron	7
2.6	Sistemas nebulosos	8
3	Implementação do Perceptron multi-camadas	9
4	Comparação dos modelos	11
Apêndices		
Anexos		
	Referências	13
	Índice Remissivo	15

Capítulo 1

Introdução

De tempos pra cá, ler e ouvir falar de **ciência de dados** tornou-se muito comum, tanto nos meios profissionais e científicos quanto na mídia. Existem atualmente aplicações em praticamente todas as áreas do conhecimento humano, da agricultura à indústria e ao entretenimento.

Uma busca rápida na *Wikipedia* (*Ciência de Dados* 2020) define ciência de dados como um conjunto de ferramentas que extrai informações ou previsões a partir de um grande volume de dados, que podem ser números, textos, áudio, vídeo, entre outros, para ajudar na tomada de decisões de negócios.

Apesar de não ser a única definição para o termo, Pedro A. Morettin e Julio M. Singer (PEDRO A. MORETTIN, 2020) nos lembram que essa também é uma definição da estatística. Eles comparam o uso dos termos e apontam que o trabalho dos *cientistas de dados* diferem dos *estatísticos* apenas quando eles usam dados de natureza multimídia como áudio e vídeo, por exemplo. Mas que, uma vez que esses dados são processados e tornam-se números, as técnicas e conceitos utilizados pelos primeiros passam a ser basicamente os mesmos utilizados pelos segundos.

Na verdade, Morettin & Singer (PEDRO A. MORETTIN, 2020) citam que na década de 80 houve uma primeira tentativa de aplicar o rótulo *ciência de dados*, (*Data Science*), ao trabalho feito pelos estatísticos aplicados da época, como uma forma de dar-lhes mais visibilidade. Curiosamente, fato mencionado pelos autores, existem atualmente cursos específicos de ciência de dados em universidades ao redor do mundo, mas a maioria deles situada em institutos de áreas aplicadas como engenharia e economia, e raramente nos institutos de estatística propriamente ditos.

Para entender um pouco mais de seu escopo, David M. Blei e Padhraic Smyth (DAVID M. BLEI, 2017) discutem ciência de dados sob as visões estatística, computacional e humana. Eles argumentam que é a combinação desses três componentes que formam a essência do que ela é e, assim como, do conhecimento que ela é capaz de produzir.

Em resumo, a estatística guia a coleta e análise dos dados. A computação cria algoritmos, técnicas de processamento e gerenciamento de memória eficazes para que sua execução seja efetiva. E o papel humano é o de avaliar quais tipos de dados, técnicas de análises,

algoritmos e modelos são apropriados para responder ao problema em questão. Este é o papel do *cientista de dados*.

Algoritmos de **aprendizado de máquina** vem sendo utilizados em grande parte dos modelos de ciência de dados. Mas o que é aprendizado de máquina? Ou então, o que significa dizer que o computador, neste caso a “máquina”, está *aprendendo*?

Aurélien Géron ([GÉRON, 2019](#)) nos dá uma ideia geral lembrando que uma das primeiras aplicações de sucesso de aprendizado de máquina foi o filtro de *spam*, criado na década de 90. Uma das fases de seu desenvolvimento foi aquela em que os usuários assinalavam que certos e-mails eram *spams* e outros não eram. Hoje em dia, raramente temos que marcar ou desmarcar e-mails, pois a maioria dos filtros já “aprenderam” a fazer seu trabalho de forma muito eficiente, não temos mais nada a “ensiná-lo”.

O conceito de aprendizado de máquina está intimamente ligado à ciência da computação. Porém, no contexto de ciência de dados, é definido por Joel Grus ([GRUS, 2016](#)) como a “criação e o uso de modelos que são ajustados a partir dos dados”. Seu objetivo é usar dados existentes para desenvolver modelos que possamos usar para *prever* possíveis respostas à consultas. Exemplos, além do filtro de *spams* podem ser: detectar transações de crédito fraudulentas, calcular a chance de um cliente clicar em uma propaganda ou então prever qual time de futebol irá vencer o Campeonato Brasileiro.

Como ficará claro ao longo deste trabalho, o aprendizado consiste na utilização de dados já conhecidos para ajustar parâmetros de modelos. Uma vez ajustados os parâmetros, o algoritmo que descreve o modelo passa a ser usado para responder às consultas. Essa fase de ajuste de parâmetros é chamada de aprendizado ou treinamento.

Uma **rede neural** é um exemplo de modelo preditivo de aprendizado de máquina que foi criado com inspiração no funcionamento do cérebro biológico. David Kopec ([KOPEC, 2019](#)) descreve que apesar de terem sido as primeiras a serem criadas, elas vem ganhando nova importância na última década, graças ao avanço computacional, uma vez que exigem muito processamento, e também porque podem ser usadas para resolver problemas de aprendizagem dos mais variados tipos.

Atualmente existem vários tipos de redes neurais, porém este trabalho lida principalmente com aquele tipo que foi originalmente criado sob a inspiração do funcionamento do cérebro, chamado de **perceptron**, e que portanto tenta imitar o comportamento dos neurônios e suas conexões, aprendendo padrões a partir de dados existentes e tentando prever o comportamento de dados novos a partir do padrão aprendido.

Uma visão geral da arquitetura do aprendizado de máquina, situando a posição das redes neurais e do algoritmo *perceptron* em toda esta estrutura, assim como exemplos de aplicações em cada uma das suas ramificações, estão no Capítulo 2.

Neste trabalho é utilizada como base uma versão simples do algoritmo *perceptron* feita e explicada por Kopec ([KOPEC, 2019](#)), e a partir desta base, foram criados novos métodos de treinamento e de validação com algumas estratégias, como uma tentativa de automatizar e otimizar o processo de treinamento do algoritmo, que é comumente feito de forma heurística. Detalhes dessa implementação estão no Capítulo 3.

Como validação e aplicação do algoritmo criado, são feitas comparações dele com os modelos tradicionais de previsões de séries temporais, como o **SARIMA** (*seasonal auto regressive integrated moving average*). Tais comparações usam como inspiração trabalhos similares como o de Alsmadi, Omar, Noah e Almarashdah ([ALSMADI *et al.*, 2009](#)), e serão tratadas no Capítulo 4.

Ao final concluo sobre os modelos de previsões aqui estudados e comparados, destacando a qualidade e eficiência dos métodos de redes neurais, tão bons e às vezes melhores do que os métodos tradicionais estatísticos.

Capítulo 2

Redes neurais artificiais

Neste capítulo são apresentados alguns conceitos básicos de **aprendizado de máquina**, com foco nos algoritmos de redes neurais artificiais, em especial o **perceptron**, cujo desenvolvimento foi inspirado nas redes neurais biológicas, ou seja, os neurônios e suas conexões no cérebro, conforme descrito por Kopec (KOPEC, 2019).

Pode-se classificar as técnicas de aprendizado de várias formas, de acordo com alguma de suas características. Por exemplo, Géron (GÉRON, 2019) utiliza o grau de supervisão humana durante o seu funcionamento para classificá-los em aprendizado supervisionado ou não-supervisionado. Isto quer dizer que durante o treinamento são fornecidos um conjunto de consultas e de respostas esperadas. Tais respostas foram dadas por humanos, daí o termo ‘supervisão humana’.

2.1 Aprendizado de máquina supervisionado

Um algoritmo de **aprendizado supervisionado** é usado quando conhecemos os rótulos dos dados que estamos utilizando. De modo geral já temos de antemão as respostas às consultas para os dados utilizados no treinamento. Por exemplo, se estamos classificando fotos de animais, possuímos um conjunto de fotos em que já sabemos quais são de gatos, cachorros, etc.

O ato de rotular previamente os dados que usamos no treinamento é o que designamos de supervisão humana. Uma vez *treinado*, o algoritmo recebe uma foto, ou seja, uma nova consulta e então fornece a resposta se essa é a foto de um gato, ou cachorro, ou qualquer outra resposta daquelas que foram dadas como exemplos durante o treinamento.

Dentro do aprendizado supervisionado temos duas técnicas principais. A regressão é usada para prever valores, ou seja, fornecer dados do futuro, e a classificação é usada para prever os rótulos dos dados, que também são chamados de classes. Neste texto os termos “algoritmo” e “técnica” serão usados livremente como sinônimos, pois uma técnica de aprendizado de máquina, no contexto atual, é obviamente um algoritmo executado no computador.

Colocar exemplos...

2.2 Aprendizado não-supervisionado

Nesse tipo de aprendizado de máquina, não sabemos os rótulos dos dados que estamos lidando, assim o algoritmo poderá agrupar os dados de forma automática, por exemplo, se estivermos lidando com problemas de classificação. Aqui, as consultas podem ser coisas como “quantos são os perfis dos clientes” ou “quantas espécies de flores existem nestas fotos”, e assim por diante.

Alguns métodos não-supervisionados de aprendizado foram enumeradas por Géron (GÉRON, 2019). O **agrupamento** de dados similares sob uma inspiração geométrica. Nesse caso os dados são agrupados conforme suas posições num determinado espaço e utiliza-se algoritmos como *k*-vizinhos, *k*-means, *k*-medians, etc. Exemplos de aplicações são agrupamento de produtos em supermercados, interesses comuns de clientes em sites de conteúdo digital, etc.

Outra técnica é a **detecção de anomalias**, cujo objetivo é ter uma descrição de como os dados considerados “normais” se parecem, e usa-se esse agrupamento para detectar se novos dados estariam “fora” desse padrão. Um exemplo é a detecção de fraudes.

Também pode-se citar sobre a técnica de **estimação de densidades**, que tem como objetivo a estimação da função densidade de probabilidade de um conjunto de dados gerados por algum processo aleatório.

Colocar exemplos...

2.3 Técnicas de classificação

É uma das técnicas principais do aprendizado supervisionado, problemas desse tipo buscam aprender com um conjunto de dados previamente rotulados. Consultas do tipo “a qual grupo pertence este cliente” ou “que animal há nesta foto” podem ser modeladas por esses algoritmos. De modo geral, ele responde a consultas que dizem respeito às classes dos dados, e atribui para novos dados alguma classe que pertence ao conjunto de classes que usamos para rotular os dados iniciais.

Existem vários tipos de algoritmos de classificação, dentre eles podemos mencionar: máquina de vetor suporte (*support vector machine*, SVM), árvores de decisão, florestas aleatórias que são um conjunto de muitas de árvores de decisão aleatoriamente definidas e, finalmente, as redes neurais artificiais.

Todas essas técnicas podem ser usadas para classificação linear ou não-linear, no sentido em que valores eles estão classificando, assim como na forma que está sendo feita essa classificação. Se visualizarmos os dados num espaço bidimensional, um algoritmo de classificação linear irá separar as classes de dados por retas, enquanto que um classificador não-linear poderá usar outra curva qualquer para a separação.

Abstraindo o espaço bidimensional para os espaços multidimensionais dos dados que são comumente analisados, podemos pensar em hiperplanos, estruturas ($n-1$ -dimensionais de espaços n -dimensionais, para o caso dos classificadores lineares, ou subespaços quaisquer para os não-lineares.

Colocar exemplos de aplicações...

2.4 Redes neurais artificiais

Uma rede neural artificial é um dentre vários métodos de classificação, ou seja, de aprendizado supervisionado. De acordo com Kopec (KOPEC, 2019), ele é utilizado como um classificador não-linear, e por isso pode ser utilizado para prever tipos de dados genéricos, que podem ou não ser lineares.

Contextualizar mais as redes neurais aqui ...

2.5 Perceptron

Esta é a primeira rede neural artificial a ser criada, e também a mais simples. Dá-se o nome de *perceptron* de camada única (*single-layer perceptron*), e uma ilustração conceitual dele está na Figura 2.1. Mais recentemente foram criadas várias outras versões dessa rede, dentre as quais podemos citar o *perceptron* de multi-camadas (*multi-layer perceptron*).

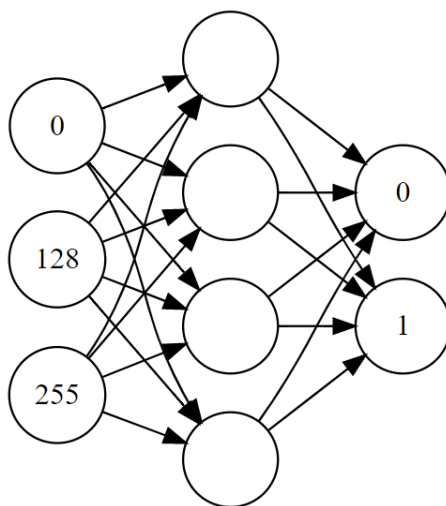


Figura 2.1: Rede neural simples, o perceptron de camada única.

Os neurônios são representados por círculos, dentro deles há um valor numérico que intuitivamente podemos atribuir ao nível ou grau de ativação do neurônio, mesmo que no caso biológico se restrinja aos valores 0 e 1, ou seja, ativados ou não. Cada coluna de neurônios representa uma camada, nesse caso, da esquerda para a direita temos a camada de entrada, a camada oculta e a camada de saída. As linhas representam as ligações entre os neurônios, sendo que cada neurônio de uma camada está ligado a todos da camada anterior.

O perceptron de camada única consiste de uma camada de neurônios de entrada, uma camada oculta de neurônios usados na otimização, e uma camada de saída, que irá conter os dados previstos, ou ainda as probabilidades do dado pertencer a alguma das classes que a rede poderá classificá-lo. E é o fato de haver uma camada oculta nesta rede que a define

como sendo de “camada única”. Caso houvessem mais do que uma camada oculta, ela seria do tipo “multi-camadas”, naturalmente.

De modo a entendermos as bases matemáticas do algoritmo, podemos começar de uma rede ainda mais básica, a partir um *perceptron* que seja constituído de apenas 1 neurônio na única camada oculta. Esta rede super simplificada, que está na Figura 2.2, pode ser útil para para o entendimento uma vez que neste caso será possível acompanhar graficamente o resultado da execução do algoritmo.

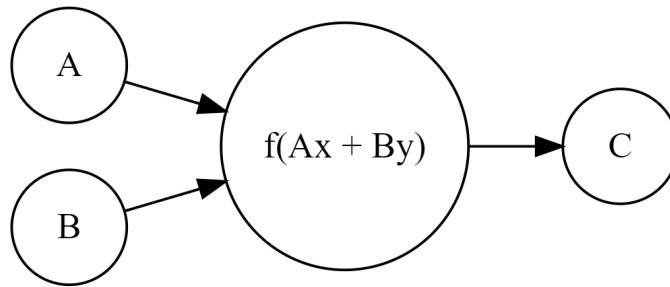


Figura 2.2: Rede neural mais simples ainda, apenas um neurônio oculto.

Esta rede possui 2 neurônios na camada de entrada, que são os números reais A e B , 1 neurônio na camada oculta, no qual está a sua função de ativação $f(Ax + By)$, e 1 neurônio na camada de saída, que neste caso é um número real C .

2.6 Sistemas nebulosos

Os sistemas nebulosos foram criados a partir da teoria dos conjuntos nebulosos criada por L.A. Zadeh (ZADEH, 1965). Os conjuntos nebulosos diferem dos conjuntos clássicos na forma em que avaliamos se os elementos pertencem aos conjuntos ou não.

Num conjunto clássico um elemento pertence a ele ou não, não há meio termo. Já nos conjuntos nebulosos existe uma função associando o grau com que um elemento pertence ao conjunto, a função assume 0 se não pertence e assume 1 se pertence completamente, mas também assume qualquer valor real dentro deste intervalo. Esta é a chamada função de pertinência, de forma que a frase “o elemento x percente mais ao conjunto A do que o elemento y ”, por exemplo, faz sentido neste contexto.

Os sistemas nebulosos são portanto modelos matemáticos de situações do mundo real que não podem ser muito bem representadas pelos conjuntos clássicos com seus limites muito bem definidos, assim eles utilizam os conjuntos nebulosos. Um exemplo ilustrativo é o do conceito de *perto*.

Usar imagens e gráficos para mostrar esse exemplo aqui...

A partir disto, conforme descrito por L. Wang (WANG, 1992), os sistemas nebulosos são aproximadores universais. Isto significa que podem aproximar quaisquer funções numa grande variedade de problemas. Neste artigo é demonstrado que com uma função de pertinência gaussiana é possível usar sistemas nebulosos para aproximar qualquer função real contínua definida num conjunto compacto.

Capítulo 3

Implementação do Perceptron multi-camadas

Neste capítulo é descrita a implementação e funcionamento de uma versão do algoritmo *perceptron*, feito a partir de um núcleo básico disponibilizado no livro de Kopec (KOPEC, 2019), e a partir do qual foram feitas modificações e criação de novos métodos de treinamento, de validação e de avaliação do treinamento.

Capítulo 4

Comparação dos modelos

Neste capítulo serão feitas comparações entre os modelos preditivos de séries temporais, através de medidas de desempenho comuns como percentual de acerto das previsões e observações das funções de erros dos algoritmos utilizados.

Referências

- [ALSMADI *et al.* 2009] M. k. ALSMADI, K. B. OMAR, S. A. NOAH e I. ALMARASHDAH. “Performance comparison of multi-layer perceptron (back propagation, delta rule and perceptron) algorithms in neural networks”. Em: *2009 IEEE International Advance Computing Conference* (mar. de 2009), pgs. 296–299. DOI: [10.1109/IADCC.2009.4809024](https://doi.org/10.1109/IADCC.2009.4809024) (citado na pg. 3).
- [Ciência de Dados 2020] *Ciência de Dados*. https://pt.wikipedia.org/wiki/Ci%C3%A7%C3%A2ncia_de_dados. Mar. de 2020 (citado na pg. 1).
- [DAVID M. BLEI 2017] Padhraic Smyth DAVID M. BLEI. “Science and data science”. Em: *PNAS* 114.33 (ago. de 2017), pgs. 8689–8692 (citado na pg. 1).
- [GÉRON 2019] Aurélien GÉRON. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 2°. O’Reilly, 2019 (citado nas pgs. 2, 5, 6).
- [GRUS 2016] Joel GRUS. *Data Science do Zero*. 1°. O’Reilly, 2016 (citado na pg. 2).
- [KOPEC 2019] David KOPEC. *Problemas Clássicos de Ciência da Computação com Python*. 1°. Novatec, 2019 (citado nas pgs. 2, 5, 7, 9).
- [PEDRO A. MORETTIN 2020] Julio M. Stinger PEDRO A. MORETTIN. *Introdução à Ciência de Dados - Fundamentos e Aplicações*. Departamento de Estatística. Universidade de São Paulo, 2020 (citado na pg. 1).
- [WANG 1992] L. -. WANG. “Fuzzy systems are universal approximators”. Em: *[1992 Proceedings] IEEE International Conference on Fuzzy Systems* (mar. de 1992), pgs. 1163–1170. DOI: [10.1109/FUZZY.1992.258721](https://doi.org/10.1109/FUZZY.1992.258721) (citado na pg. 8).
- [ZADEH 1965] L.A. ZADEH. “Fuzzy sets”. Em: *Information and Control* 8.3 (1965), pgs. 338–353. ISSN: 0019-9958. DOI: [https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X). URL: <http://www.sciencedirect.com/science/article/pii/S001999586590241X> (citado na pg. 8).

Índice Remissivo

C

Captions, *veja* Legendas

Código-fonte, *veja* Floats

E

Equações, *veja* Modo Matemático

F

Figuras, *veja* Floats

Floats

Algoritmo, *veja* Floats, Ordem

Fórmulas, *veja* Modo Matemático

I

Inglês, *veja* Língua estrangeira

P

Palavras estrangeiras, *veja* Língua

estrangeira

R

Rodapé, notas, *veja* Notas de rodapé

S

Subcaptions, *veja* Subfiguras

Sublegendas, *veja* Subfiguras

T

Tabelas, *veja* Floats

V

Versão corrigida, *veja* Tese/Dissertação,
versões

Versão original, *veja* Tese/Dissertação,
versões