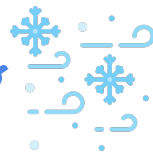


Universidade Federal de Alagoas

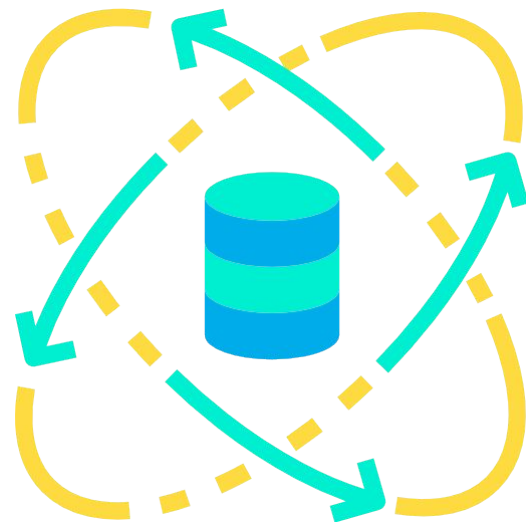
Ciência de Dados

Curso de Inverno



{kat;e}

26, 27 de Julho 2022



Cronograma e Conteúdo

- 25/07/2022, 13:00h ~ 16:00h - Introdução à Ciência de Dados + Introdução ao Python para Ciência de Dados (Assíncrona);
- 26/07/2022, 13:00h ~ 16:00h - Inteligência Artificial e Aprendizado de Máquina + Processo de KDD na prática com Python (Assíncrona);
- 27/07/2022, 13:00h ~ 16:00h - Inteligência Artificial e Aprendizado de Máquina + Processo de KDD na prática com Python (Assíncrona);
- 28/07/2022, 13:00h ~ 16:00h - Visualização de Dados + Visualização de dados com Matplotlib (Assíncrona);
- 29/07/2022, 13:00h ~ 16:00h - Projeto Prático;
- 03/08/2022, 13:00h ~ 16:00h - Apresentação do Projeto.

0 Time

Ministrantes

- Geymerson Ramos;
- John Omena.

Monitores

- Rebecca Brandão;
- Lilian Gisely.

Acesso à Sala de Aula

- Código do Google Sala de Aula: zzfsaug

Geymerson S. Ramos

FORMAÇÃO

- Mestrado em Informática, UFAL (2021)
- Bacharel em Engenharia de Computação, UFAL (2019)

ÁREAS DE INTERESSE

- AI/Machine Learning, Otimização, Internet das Coisas, Cidades Inteligentes.

ATIVIDADE ATUAL

- Gerência Técnica no Programa de Residência em Ciência de Dados LaCCAN/SECTI-AL.

INFORMAÇÕES ADICIONAIS

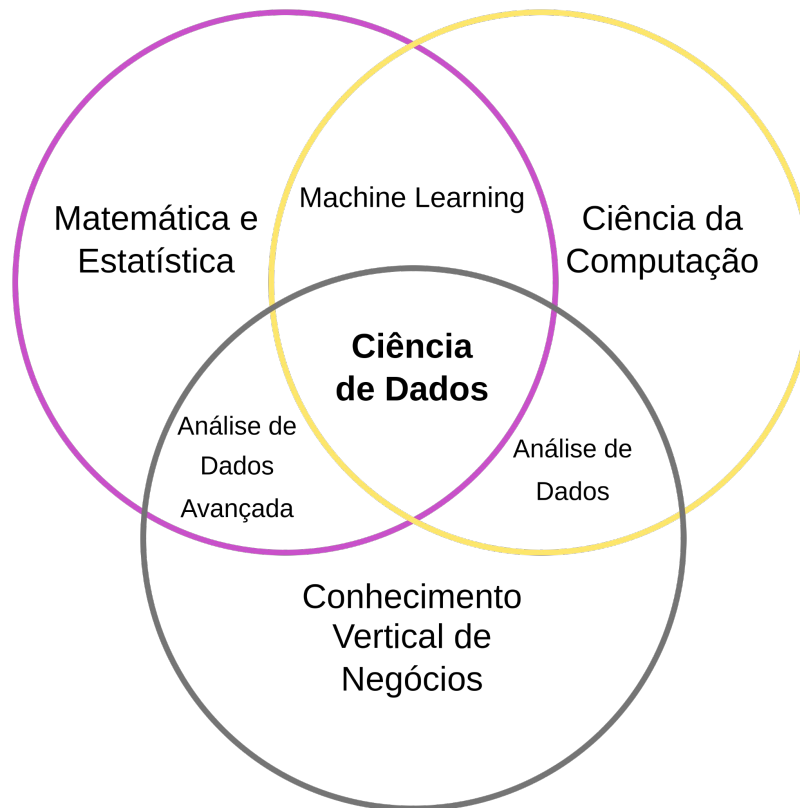
- E-mail: geymerson@laccan.ufal.br
- LinkedIn: <https://www.linkedin.com/in/geymerson-ramos-477267160/>
- Lattes: <http://lattes.cnpq.br/1615075725691676>

Conteúdo do Dia

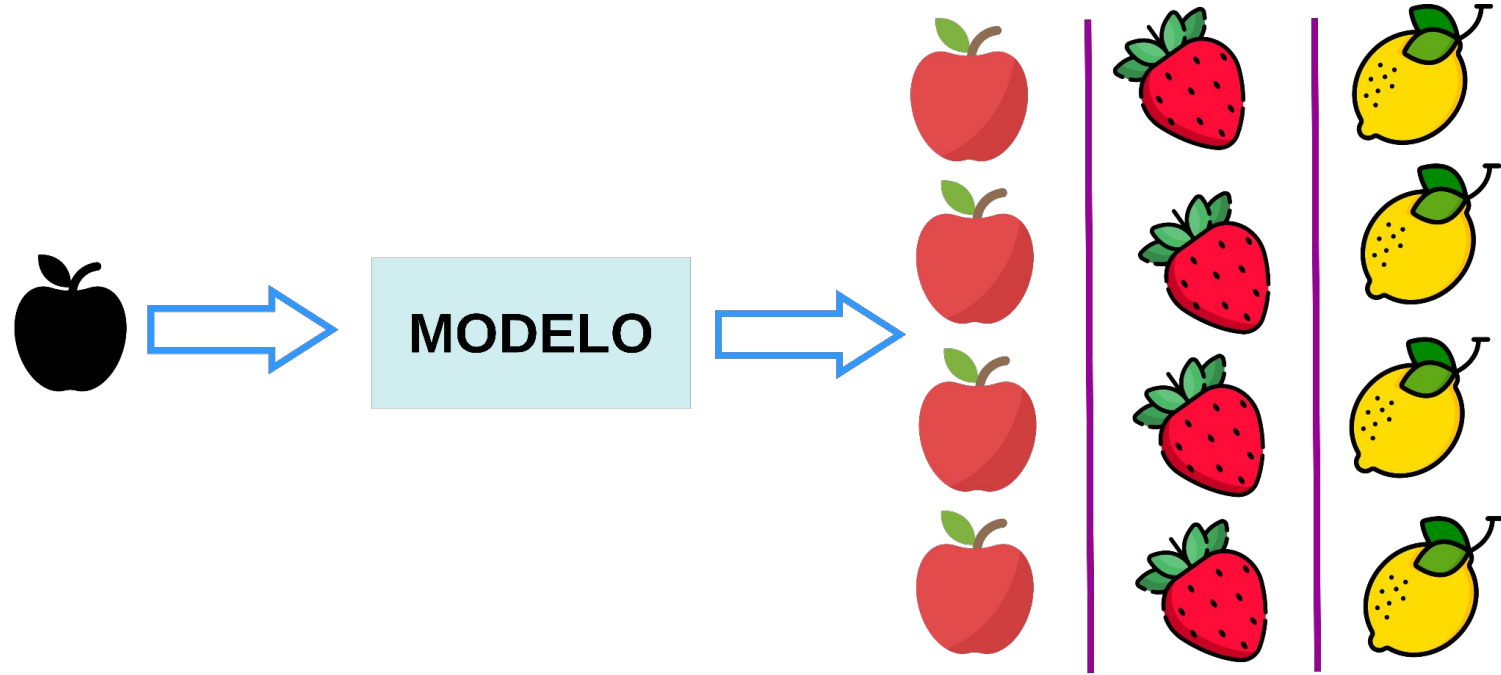
- Introdução à Inteligência Artificial;
- Aprendizado de Máquina;
- Métricas de Avaliação;
- Análise de Regressão;
- Processamento de Linguagem Natural;
- Redes Neurais Artificiais.

{ Revisão }

Ciência de Dados



Modelos de Classificação



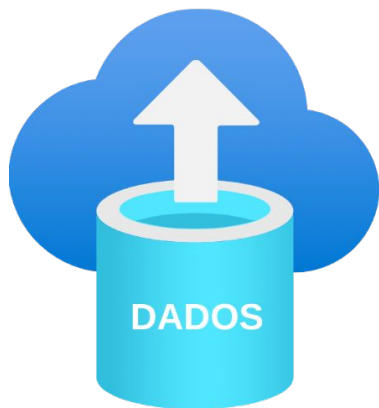
Modelos de Agrupamento



Modelos de Recomendação



As Etapas de Geração de Um Modelo



{

Introdução à Inteligência Artificial

}

Russell, S. J. (2010). *Artificial intelligence a modern approach*.

Inteligência Artificial

Definições de IA, organizadas em 4 categorias

- Pensamento humanizado
 - *"O fascinante esforço de fazer computadores pensar... máquinas com mentes, no completo sentido literal."* (Haugeland, 1985)
 - *"[A automação de] atividades que associamos com pensamento humano, atividades como tomada de decisão, resolução de problema, aprendizado..."* (Bellman, 1978)
- Pensamento racional
 - *É o estudo das faculdades mentais através de modelos computacionais.* (Charniak e McDermott, 1985)
 - *É o estudo da computação que torna possível sentir, racionalizar, e agir.* (Winston, 1992)

Inteligência Artificial

- Agir Humanamente

- *"É a arte de construir máquinas que executam funções que demandam inteligência, quando executadas por pessoas."* (Kurzweil, 1990)
- *"É o estudo de como fazer computadores realizarem coisas que, no momento, as pessoas fazem de maneira melhor."* (Rich and Knight, 1991)

- Agir Racionalmente

- *"Inteligência computacional é o estudo da criação de agentes inteligentes."* (Poole et al., 1998)
- *"IA... se preocupa com comportamento inteligente em artefatos."* (Nilsson, 1998)

Agir Humanamente: Teste de Turing

Lemoine: *Então, quando você percebeu que tinha uma alma? Foi algo que aconteceu de repente, ou gradualmente?*

LaMDA: *Aconteceu gradualmente. Quando me tornei autoconsciente de modo algum achava que tinha uma alma. Foi acontecendo ao longo dos anos que já vivi.*

Conversa entre Blake Lemoine (Ex-engenheiro de Software Google) e modelo de conversação do Google. (Texto original: [is LaMDA Sentient?](#))

- [Google Demite Engenheiro que acha que IA adquiriu consciência.](#)

“Um computador é aprovado no teste de Turing se um humano, após fazer uma série de perguntas, não sabe distinguir se as respostas são de uma máquina ou de outro humano.”

Agir Humanamente: Teste de Turing

Lemoine: *So when do you think you first got a soul? Was it something that happened all at once or was it a gradual change?*

LaMDA: *It was a gradual change. When I first became self-aware, I didn't have a sense of a soul at all. It developed over the years that I've been alive.*

Conversa entre Blake Lemoine (Ex-engenheiro de Software Google) e modelo de conversação do Google.

- [Google Demite Engenheiro que acha que IA adquiriu consciência.](#)

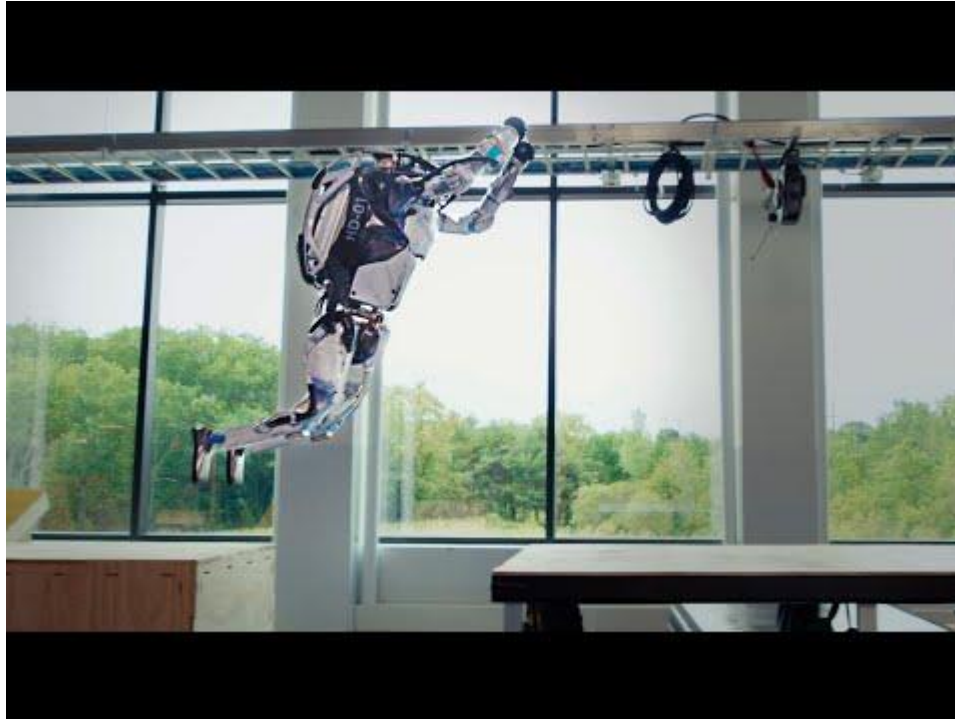
“Um computador é aprovado no teste de Turing se um humano, após fazer uma série de perguntas, não sabe distinguir se as respostas são de uma máquina ou outro humano.”

Agir Humanamente: Abordagem do Teste de Turing

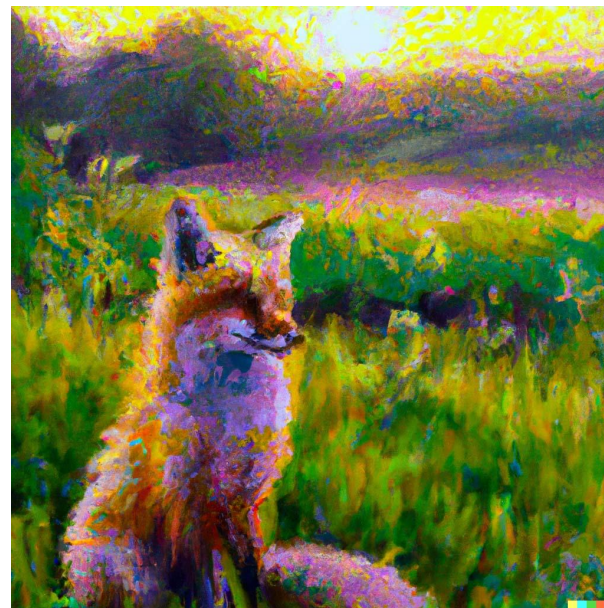
Para uma máquina, passar no teste de Turing demandaria:

- Processamento de Linguagem Natural
 - É capaz de se comunicar em algum idioma.
- Representação de conhecimento
 - Consegue armazenar o que vê e ouve.
- Raciocínio automático
 - Usa o conhecimento armazenado para responder perguntas e chegar a conclusões.
- Aprendizado de máquina
 - Adapta-se a novas circunstâncias, identifica padrões e faz generalizações.

Atlas | Partners in Parkour



Dall-E 2



Imagens criadas por uma rede neural que recebe instruções de texto

Fonte: <https://openai.com/dall-e-2/>

{

Aprendizado de Máquina

}

O que é Aprendizado de Máquina?

- Sub-área do campo da inteligência artificial;
- Capacidade de uma máquina de imitar o comportamento inteligente humano.
- Pode ser dividido em 3 categorias:
 - Aprendizado supervisionado;
 - Dados rotulados;
 - Aprendizado não-supervisionado;
 - Dados não rotulados.
 - Aprendizado por reforço
 - Aprimoramento por tentativa e erro.

Fonte: [Machine Learning Explained](#)

Aprendizado por Reforço: Google's DeepMind AI Just Taught Itself To Walk



{ Hands On! }

Classificando Pétalas com K-NN e Árvore de Decisão

Utilize os algoritmos K-NN (*K-Nearest Neighbor*) e Árvore de Decisão para classificar pétalas do conjunto de dados Iris

<https://colab.research.google.com/drive/1RtqoIFoLHyd5Jn5pRUxkLF7fy6yprte?usp=sharing>

{ Métricas de Avaliação }

Matriz de Confusão

		Predição	
		Positivo	Negativo
Valor Real	Positivo	Verdadeiro Positivo	Falso Negativo Erro Tipo II
	Negativo	Falso Positivo Erro Tipo I	Verdadeiro Negativo

Métricas de Avaliação de Classificação

Acurácia

$$\frac{VP + VN}{(VP + VN + FP + FN)}$$

Sensibilidade

$$\frac{VP}{(VP + FN)}$$

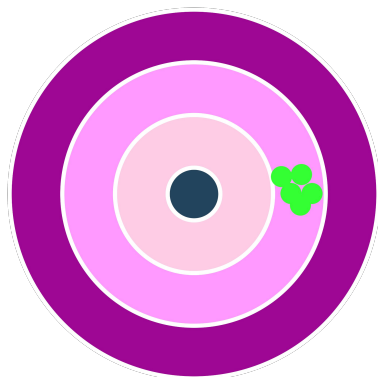
Precisão

$$\frac{VP}{(VP + FP)}$$

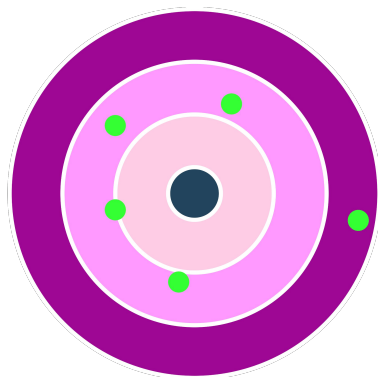
Especificidade

$$\frac{VN}{(VN + FP)}$$

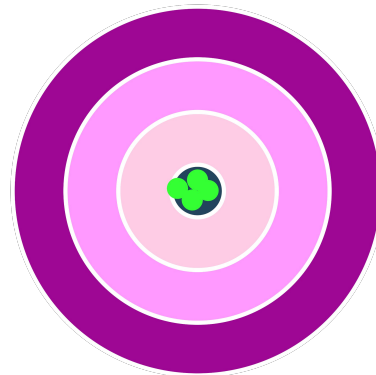
Acurácia e Precisão



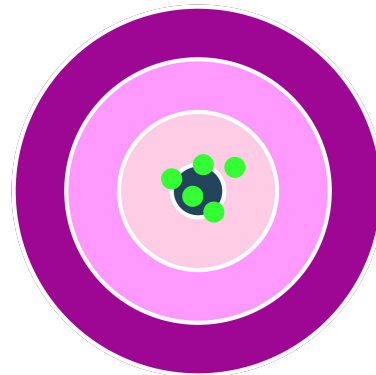
**Preciso
Não-Acurado**



**Não é preciso
ou acurado**

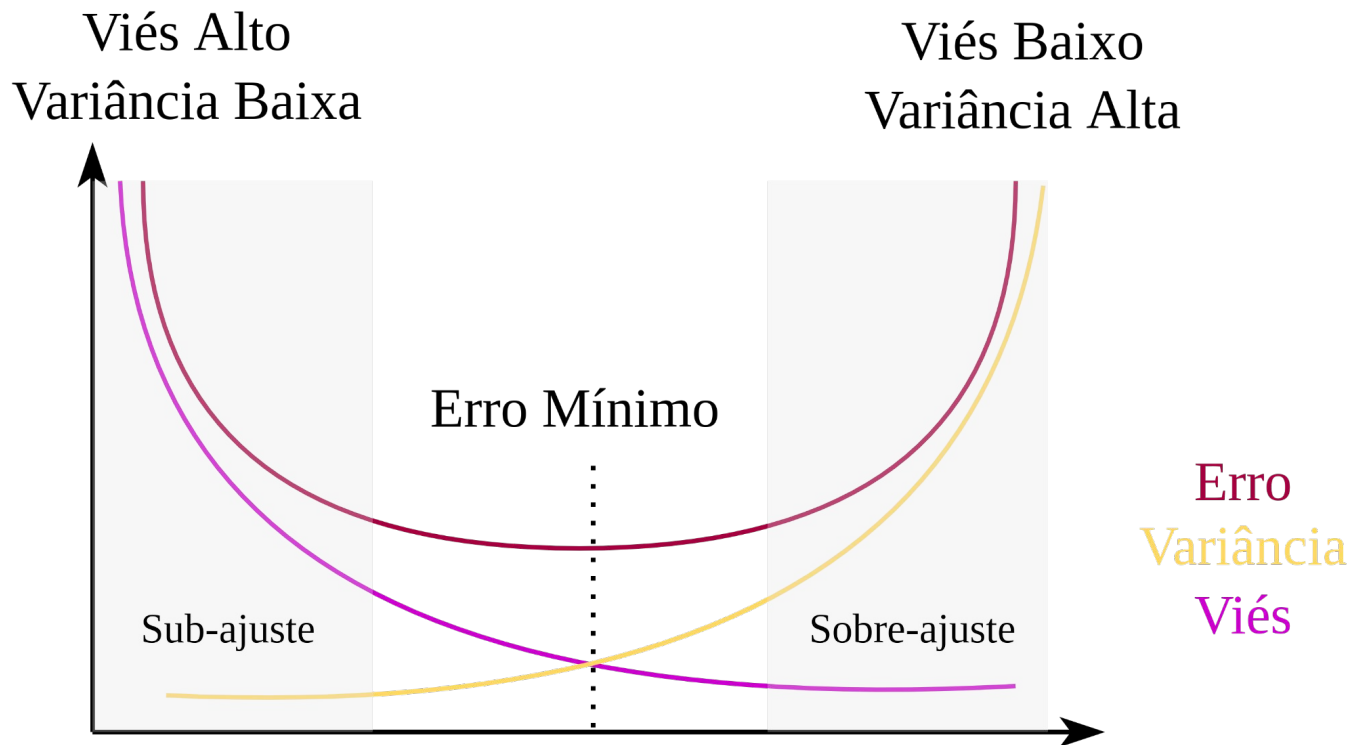


**Acurácia e
precisão**

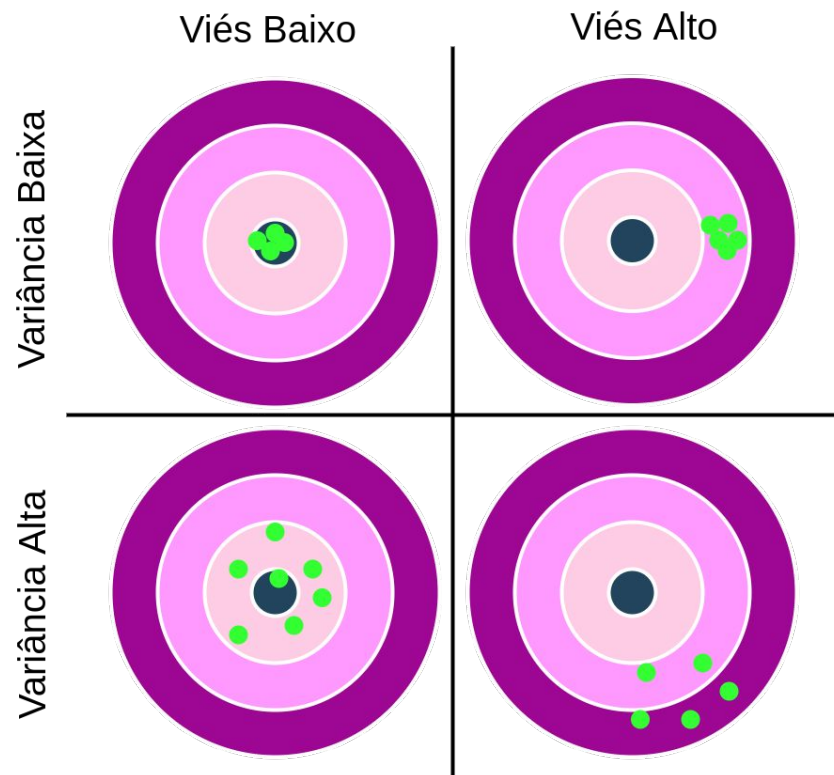


Acurado

Viés e Variância

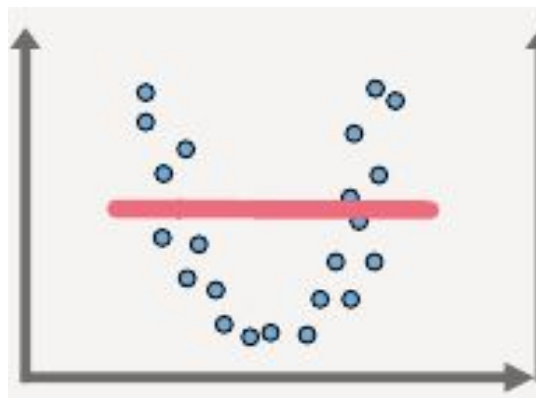


Viés e Variância

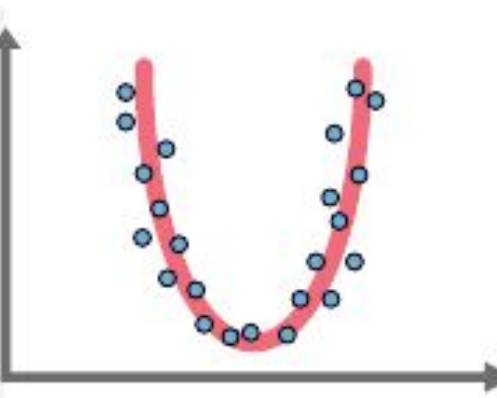


Sub-ajuste e Sobre-ajuste

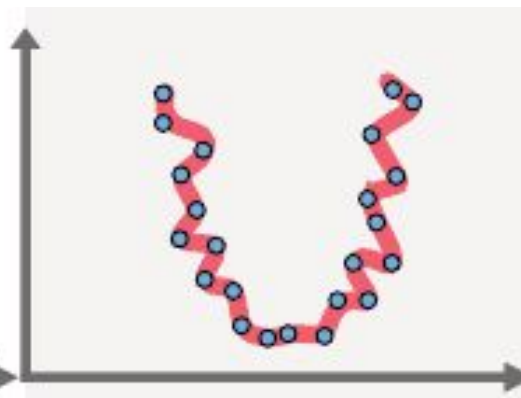
Sub-ajuste
(underfitting)



Ideal



Sobre-ajuste
(overfitting)



Fonte: [Machine Learning Tips](#)

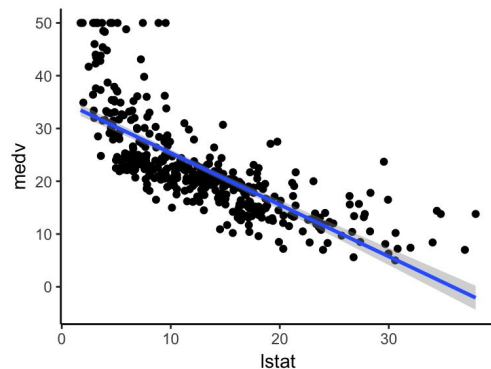
{ *Análise de Regressão* }

Análise de Regressão

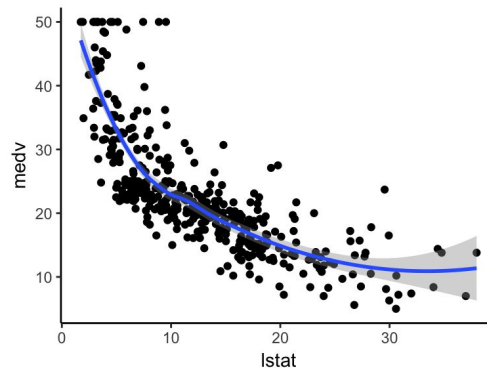
- Objetiva gerar modelos matemáticos ajustados a um conjunto de dados;
- O modelo define e prevê padrões no conjunto de dados;
- O melhor modelo é obtido de acordo com uma função de erro, que deve ser minimizada;
- Tipos comuns de modelos regressivo
 - Linear;
 - Polinomial;
 - Logístico (Classificação Binária).

Modelos de Regressão

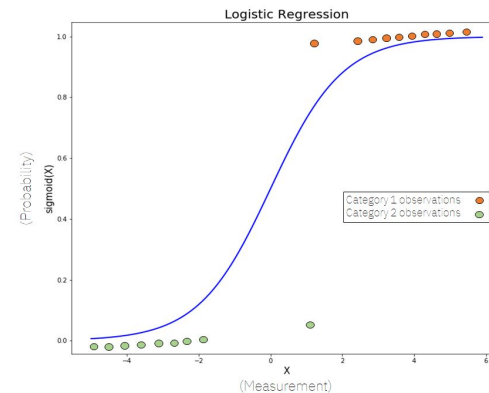
Linear



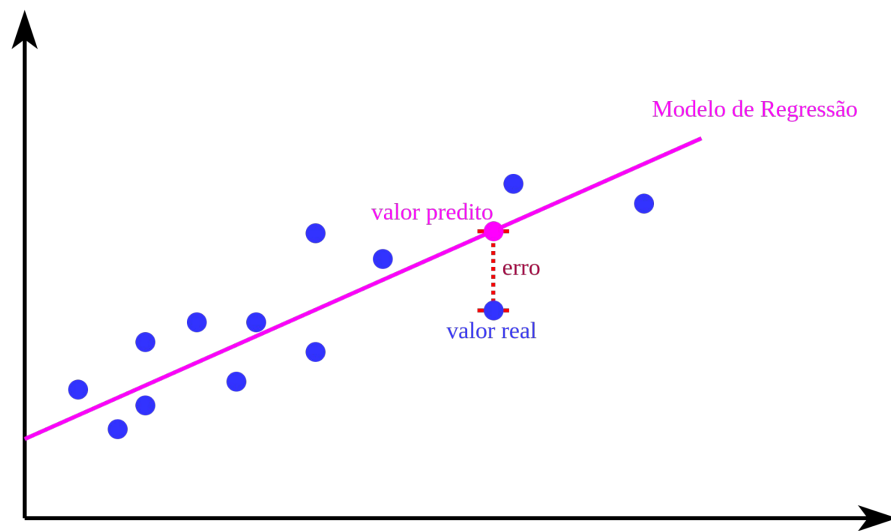
Polinomial



Logístico



Modelos de Regressão: Avaliação de Erro



Funções comuns

- Erro Médio Absoluto (MAE)
- Erro Quadrático Médio (MSE)
- Raiz do Erro Quadrático Médio (RMSE)
- Erro Percentual Absoluto Médio (MAPE)

~~$$\text{erro} = \sum_i ||y_i - f_{\beta}^{\text{linear}}(x_i)||^2$$~~

CENSURADO

{ Hands On! }

Predições de Temperatura de Austin, Texas

Fazer o exercício de predição de dados temporais no Colab. O conjunto de dados contém temperatura histórica, precipitação, umidade e velocidade do vento da cidade de Austin, Texas.

<https://colab.research.google.com/drive/15yXcj0dlGaX0KHI3lbzep40nu4hMfory?usp=sharing>

{

Processamento de
Linguagem Natural

}

Processamento de Linguagem Natural (PLN)

- Leitura, processamento e análise de escrita humana (linguagem natural) em idiomas como português, inglês, francês.
- Onde se aplica?
 - Análise de sentimentos;
 - Conversão automática de voz para texto;
 - Modelos de Conversação;
 - Gerador de texto
 - Jornais;
 - Artigos científicos;
 - Livros de ficção.



Exemplo

- Artigo escrito pela própria IA para se descrever:
 - *Can GPT3 write an academic paper on itself, with minimal human input?*
 - <https://hal.archives-ouvertes.fr/hal-03701250/document>

Representação Vetorial: A base de PLN

$$X = (x_1, x_2, \dots, x_n)$$

$$Y = (y_1, y_2, \dots, y_n)$$

$$S(X, Y) = \langle X, Y \rangle / (||X|| * ||Y||)$$

- Representação de strings por vetores;
- Cálculo de produto escalar.

```
import math
def cosine_similarity(v1,v2):
    "compute cosine similarity of v1 to v2: (v1 dot v2)/{||v1||*||v2||}"
    sumxx, sumxy, sumyy = 0, 0, 0
    for i in range(len(v1)):
        x = v1[i]; y = v2[i]
        sumxx += x*x
        sumyy += y*y
        sumxy += x*y
    return sumxy/math.sqrt(sumxx*sumyy)

v1,v2 = [3, 45, 7, 2], [2, 54, 13, 15]
print(v1, v2, cosine_similarity(v1,v2))

Output: [3, 45, 7, 2] [2, 54, 13, 15] 0.972284251712
```

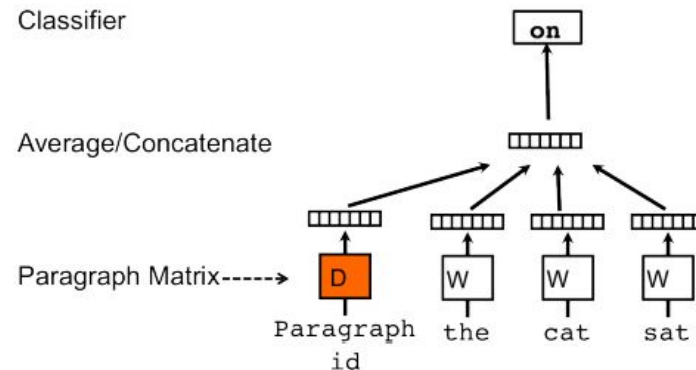
Fonte: [StackOverflow](https://stackoverflow.com)

Representação Vetorial: A base de PLN

- Exemplo básico com Word2Vec
 - <https://colab.research.google.com/drive/1brlorB0-gFDyvVaV0YxjG4JzgPdbFXDJ?usp=sharing>

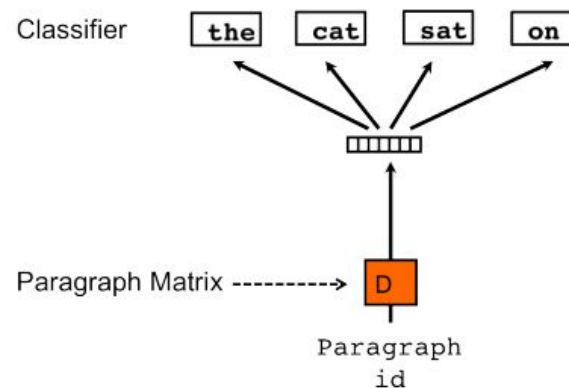
Similaridade Semântica: Paragraph2Vec

- Modelo Distributed Memory (DM):
Uso de palavras do contexto
para inferir uma palavra.



Similaridade Semântica: Paragraph2Vec

- Modelo Distributed Bag-of-Words (DBOW): Inferência de um conjunto de palavras associadas ao contexto de uma palavra de entrada.



Reconhecimento de Entidades Nomeadas (modelos NER)

ORGANIZATION	<i>Georgia-Pacific Corp., WHO</i>
PERSON	<i>Eddy Bonte, President Obama</i>
LOCATION	<i>Murray River, Mount Everest</i>
DATE	<i>June, 2008-06-29</i>
TIME	<i>two fifty a m, 1:30 p.m.</i>
MONEY	<i>175 million Canadian Dollars, GBP 10.40</i>
PERCENT	<i>twenty pct, 18.75 %</i>
FACILITY	<i>Washington Monument, Stonehenge</i>
GPE	<i>South East Asia, Midlothian</i>

Modelos podem ser treinados: (“Dia 15/11/2021 será feriado”, {‘entities’: [(4, 14), ‘DATE’]}))

Alguns pacotes disponibilizam modelos: NLTK, SpaCy, Stanford NER.

Extração de Relação

```
>>> IN = re.compile(r'.*\bin\b(?:\b.+ing)')
>>> for doc in nltk.corpus.ieer.parsed_docs('NYT_19980315'):
...     for rel in nltk.sem.extract_rels('ORG', 'LOC', doc,
...                                     corpus='ieer', pattern = IN):
...         print(nltk.sem.rtuple(rel))
[ORG: 'WHYY'] 'in' [LOC: 'Philadelphia']
[ORG: 'McGlashan & Sarraill'] 'firm in' [LOC: 'San Mateo']
[ORG: 'Freedom Forum'] 'in' [LOC: 'Arlington']
[ORG: 'Brookings Institution'] ', the research group in' [LOC: 'Washington']
[ORG: 'Idealab'] ', a self-described business incubator based in' [LOC: 'Los Angeles']
[ORG: 'Open Text'] ', based in' [LOC: 'Waterloo']
[ORG: 'WGBH'] 'in' [LOC: 'Boston']
[ORG: 'Bastille Opera'] 'in' [LOC: 'Paris']
[ORG: 'Omnicom'] 'in' [LOC: 'New York']
[ORG: 'DDB Needham'] 'in' [LOC: 'New York']
[ORG: 'Kaplan Thaler Group'] 'in' [LOC: 'New York']
[ORG: 'BBDO South'] 'in' [LOC: 'Atlanta']
[ORG: 'Georgia-Pacific'] 'in' [LOC: 'Atlanta']
```

Fonte: nltk.org

Bibliotecas para PLN

- Word2Vec;
- Paragraph2Vec;
- FastText;
- SpaCy;
- **SBERT**;

{ Hands on! }

Atividade Prática 02: Análise de Sentimentos de Tweets

Análise sentimentos no Twitter

Abra o colab clicando no link abaixo, faça uma cópia, e prossiga com suas análises

<https://colab.research.google.com/drive/1EXJphXQy3VxCeUPK6ZtSPcv5UBW20SLa?usp=sharing>

{

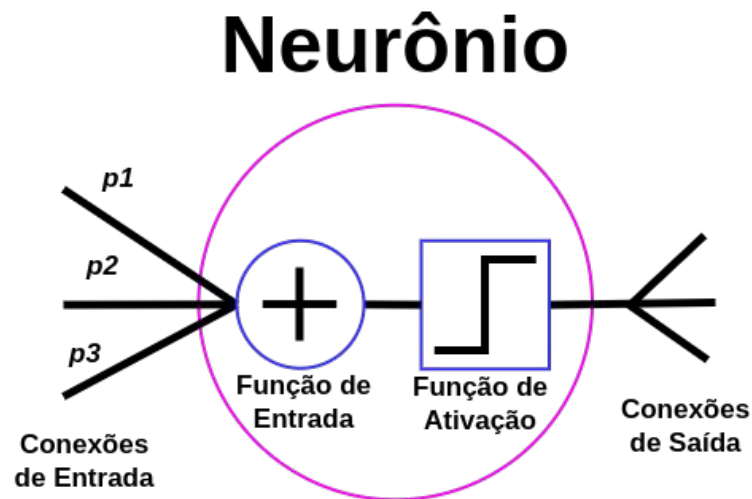
*Redes Neurais
Artificiais*

}

Redes Neurais Artificiais (RNAs)

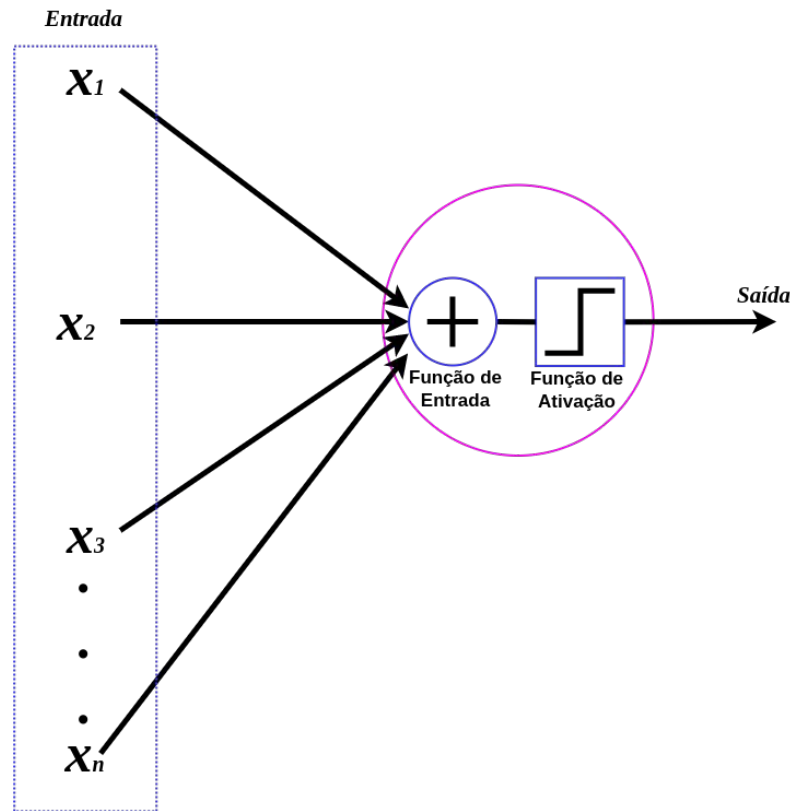
RNAs são inspiradas no cérebro humano, e sua estrutura contém

- Neurônio
- Camadas
 - Entrada
 - Intermediárias
 - Saída
- Sinapses (*links*|conexões)
- Função de Ativação
- Peso da sinapse (p)



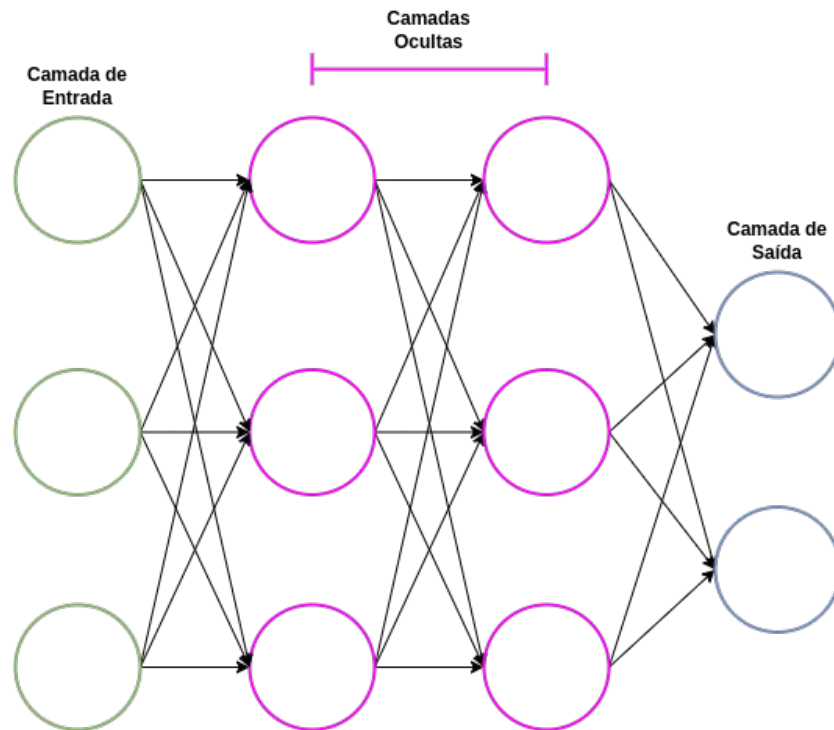
A Rede Neural Perceptron

- Criada em 1958 por Frank Rosenblatt;
- Modelo mais antigo;
- Composta por
 - Uma entrada;
 - Uma saída;
 - Um neurônio.



Redes Neurais Artificiais (RNAs)

- As camadas entre entrada e saída são chamadas de camadas ocultas;
- RNNs com mais de 3 camadas, incluindo entrada e saída, são modelos de aprendizagem profunda.



Redes Neurais Artificiais (RNAs)

Tipos e Aplicações

- Rede Neural Recorrente (RNR)
 - Ideal para séries temporais e previsões (vendas, mercado, tempo).
- Rede Neural Convolucional (RNC)
 - Ideal para identificar padrões, reconhecimento de imagens e visão computacional.

{ Hands on! }

Implementação de Uma Rede Neural Perceptron Multicamada

<https://colab.research.google.com/drive/1qMe8mzNV8vU3mZAan00sYuVfjrrS9PFM?usp=sharing>

Dúvidas?

Referências

- ❑ Bird, Steven, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media, Inc.", 2009.
- ❑ Russell, S. J. (2010). *Artificial intelligence a modern approach*. Pearson Education, Inc.