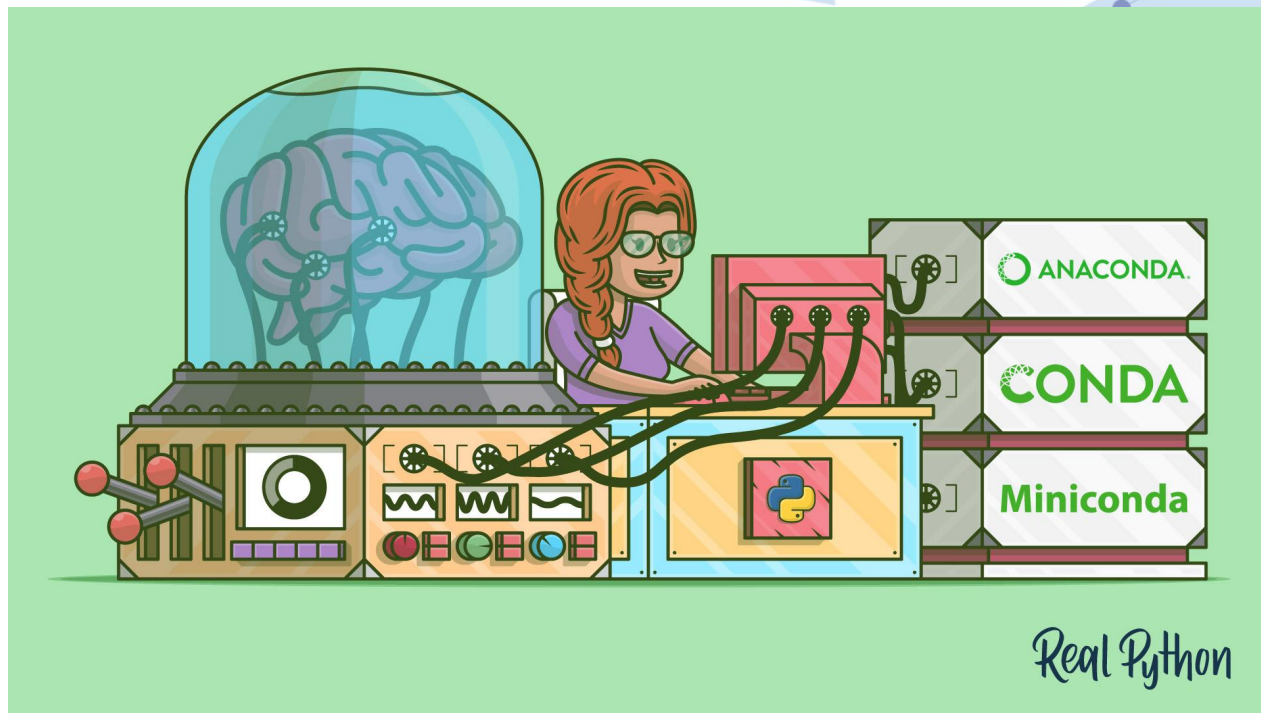


Módulo 6 - Treinamento

(aula 01)



Sumário

1. Overview do módulo
2. Ferramentas
3. Conceitos elementares
4. Feedback



Overview do módulo

Iremos conhecer melhor o nosso módulo de aulas.

Overview do módulo

(tópicos)

- Aula 01:
 - Overview do módulo
 - Ferramentas
 - Conceitos elementares
- Aula 02:
 - Subajuste
 - Sobreajuste

Overview do módulo

(tópicos)

- Aula 03:
 - Classes raras
 - Bootstrapping
 - Seleção de características
- Aula 04:
 - Grupos de treino e teste
 - Validação
 - Seleção de modelo

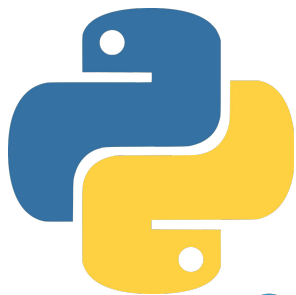


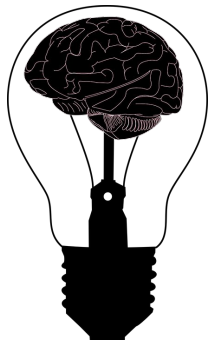
Ferramentas

Uma rápida apresentação das ferramentas que iremos utilizar neste módulo.

Ferramentas

- [Python](#)
- [Google Colab](#)

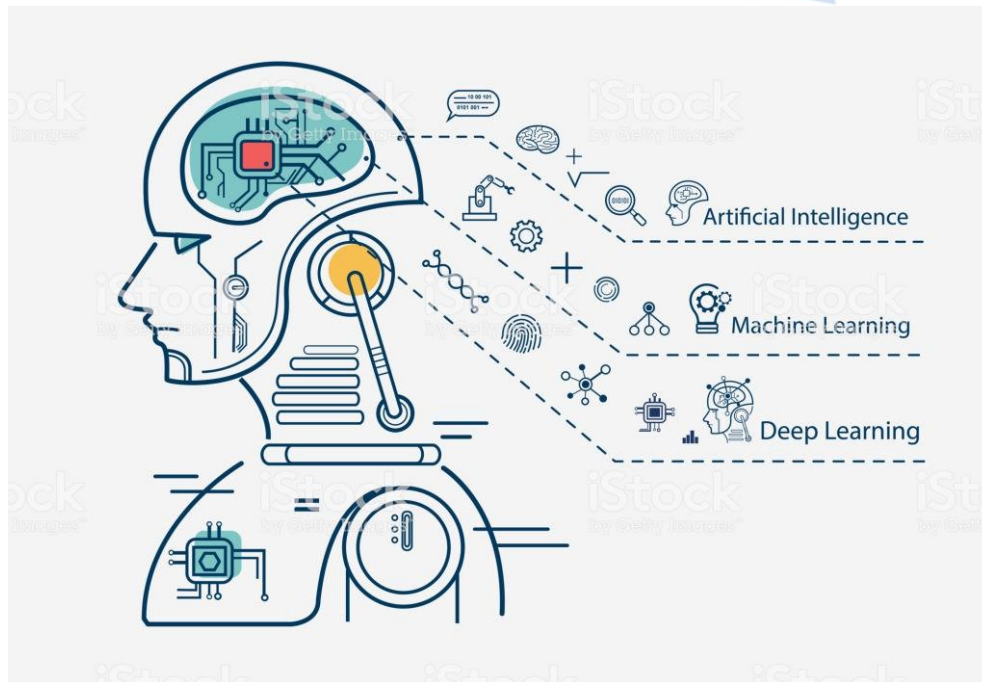




Conceitos elementares

Conceitos importantes que devemos ter em mente durante esse módulo.

Treinamento



Generalização



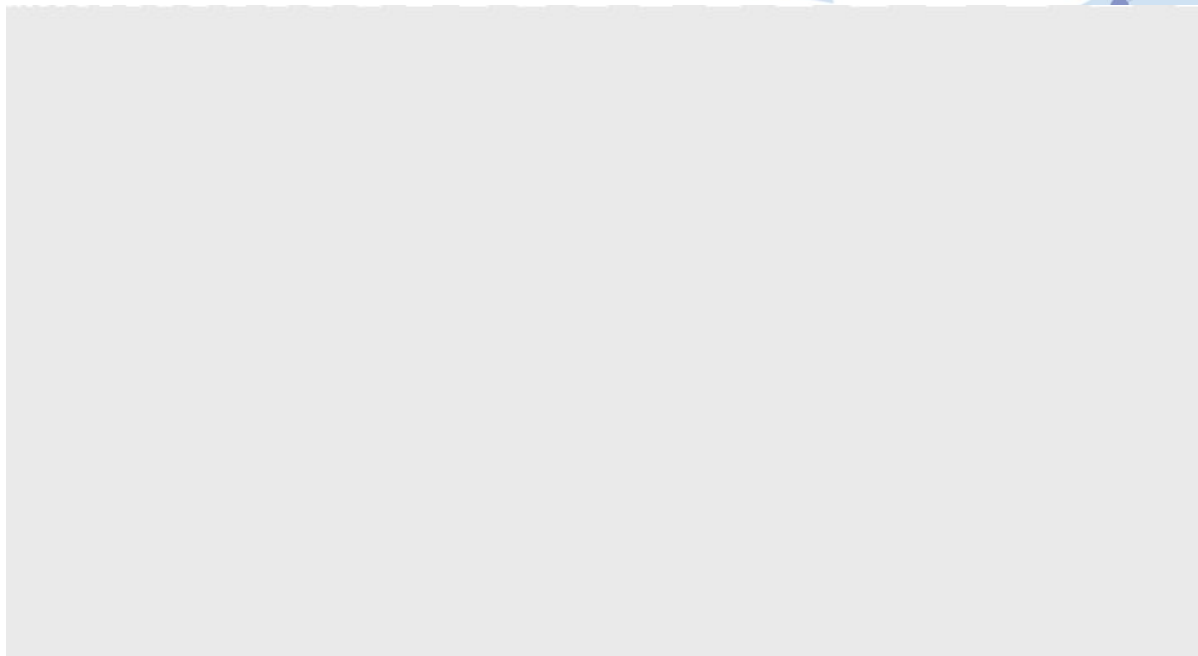
Generalização

(exercícios)

Como seria a generalização dos seguintes casos de uso:

1. Veículos autônomos;
2. Preços de imóveis;
3. Reconhecimento de manuscrito;
4. Diagnóstico de cuidados de saúde.

População e amostra



População e amostra

(exercícios)

Um pediatra escolhe aleatoriamente 10 pais dos seus doentes. Em seguida, perguntou a opinião desses pais sobre diferentes tipos de fraldas.

Identifica a população e a amostra deste estudo.

Selecione a opção correta.

- ☐ A população são os pais dos doentes do pediatra que usam fralda; a amostra são os pais de todos os doentes do pediatra.
- ☐ A população são os doentes do pediatra; a amostra são os 10 doentes selecionados.
- ☒ A população são os pais dos doentes do pediatra; a amostra são os 10 pais dos doentes selecionados.

População e amostra

(exercícios)

Uma vereadora queria saber o que os seus eleitores pensavam de um novo plano diretor municipal. Ela selecionou aleatoriamente 75 nomes da lista telefônica da cidade e realizou uma entrevista pelo telefone.

Identifica a população e a amostra deste estudo.

Selecione a opção correta.

- ☒ A população são todas as pessoas que constam da lista telefônica da cidade; a amostra são as 75 pessoas selecionadas.
- ☐ A população são os habitantes da cidade; a amostra são os eleitores registrados na cidade.
- ☐ A população são os eleitores registrados na cidade; a amostra são todas as pessoas que constam da lista telefônica da cidade.

População e amostra

(exercícios)

Um grupo de bibliotecários pretende saber o número de livros e outros recursos que os utentes requisitam da sua biblioteca. Eles analisam os registos de requisições de 150 utentes adultos escolhidos ao acaso.

Identifica a população e a amostra deste estudo.

Seleciona a opção correta.



A população são todos os utentes adultos da biblioteca; a amostra são os 150 utentes selecionados.



A população são todos os utentes da biblioteca; a amostra são os utentes adultos da biblioteca.



A população são todos os utentes que requisitaram pelo menos 1 livro da biblioteca; a amostra são os 150 utentes selecionados.

População e amostra

(exercícios)

aleatoriamente 70 entradas e mede as suas temperaturas imediatamente antes de serem servidas aos clientes.

Identifica a população e a amostra deste estudo.

Selecione a opção correta.

- ☐ A população são todas as entradas quentes que o Lúcio serve; a amostra são as entradas que estão a uma temperatura adequada.
- ☐ A população são as 70 entradas selecionadas; a amostra são as entradas que estão a uma temperatura adequada.
- ☒ A população são todas as entradas que o Lúcio serve; a amostra são as 70 entradas selecionadas.

População e amostra

(exercícios)

Um inspetor realiza testes à qualidade do ar num grupo de 7 salas de aula selecionadas ao acaso numa escola básica.

Identifica a população e a amostra deste estudo.

Selecione a opção correta.



A população são todas as salas de aula daquele concelho; a amostra são as 7 salas de aula selecionadas.



A população são todas as salas de aula daquela escola; a amostra são as 7 salas de aula selecionadas.



A população são todos os alunos daquela escola básica; a amostra são os alunos que têm aulas nas 7 salas de aula selecionadas.

População e amostra

(exercícios)

Os exercícios foram retirados do [Khan Academy](https://www.khanacademy.org/).

População e amostra

- Margem de erro
- Representatividade
 - Balanceamento
- Nível de confiança

População e amostra

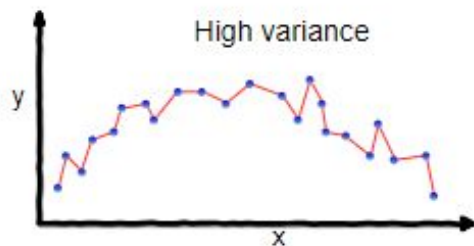
(exercícios)

Temos 90 alunos e queremos fazer uma pesquisa de estatura usando uma amostra de 10% da população.

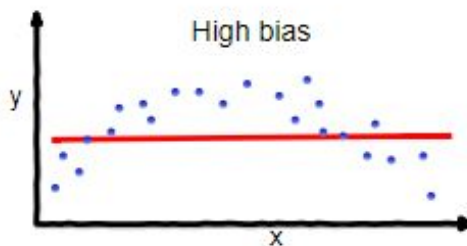
SEXO	POPULAÇÃO	10%	AMOSTRA
M	54	5,4	5
F	36	3,6	4
TOTAL	90	9,0	9

Retirado da [lista de exercícios do Centro Paula Souza](#).

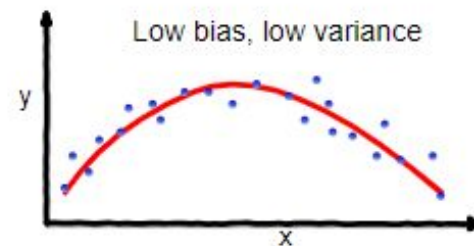
Viés e Variância



overfitting



underfitting



Good balance

Viés e Variância

- Viés de algoritmo: referente ao proprietário do código;
- Viés de amostra / Viés latente: os dados usados no treinamento não representam a população;
- Viés de preconceito: quando o conteúdo dos dados de treinamento é influenciado por estereótipos ou preconceitos provenientes da população;
- Viés de medida: resulta de uma medição incorreta;
- Viés de interação: quando o modelo é dependente de indivíduos que interagem com ele, podendo ser enviesado.

Viés e Variância

(exercícios)

INTELIGÊNCIA ARTIFICIAL >

O inconsistente e perigoso 'radar gay'

Estudo polêmico cria algoritmo que supostamente identifica gays através de inteligência artificial



PATRICIA R BLANCO 
12 SEP 2017 - 16:52 BRT

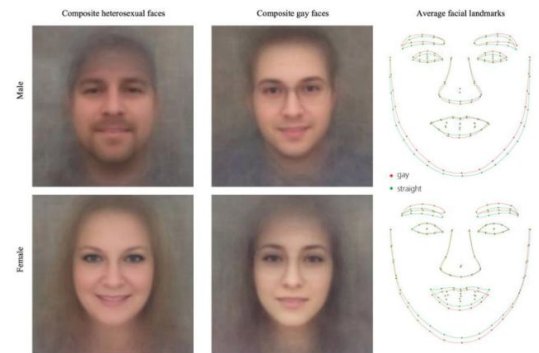


Imagem do estudo, que supostamente mostra traços faciais homossexuais e heterossexuais

- Dataset de apenas pessoas brancas;
- Possíveis problemas.

Fonte: [El País](#).

Viés e Variância

(exercícios)

Inteligência artificial: Os 'eventos estranhos' que fizeram tecnologia pensar que tartaruga era uma arma

Linda Geddes
BBC Future

🕒 9 fevereiro 2019



Compartilhar



Fonte: [BBC](#).

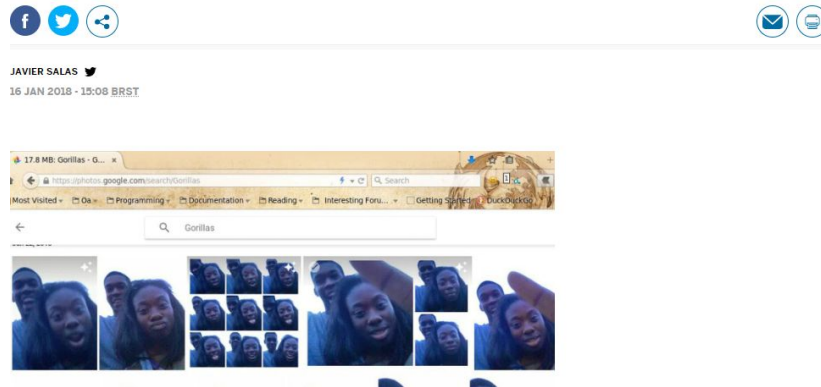
- Tartaruga reconhecido como um fuzil;
- Sinfonia como comandos de voz;
- Outros reconhecimentos errados.

Viés e Variância

(exercícios)

Google conserta seu algoritmo “racista” apagando os gorilas

Google Photos confundia pessoas negras com macacos. Este patch mostra a opacidade dos algoritmos



Fonte: [El País](#).

- Reconhecia pessoas negras como gorilas;
- “Solução”: apagar “gorilas”, “macacos” e “chimpanzés” do buscador.

Viés e Variância

(exercícios)

O robô racista, sexista e xenófobo da Microsoft acaba silenciado

Projetado para o mercado dos 'millennials' nos Estados Unidos, Tay não foi capaz de lidar com piadas e perguntas controversas



ROSA JIMÉNEZ CANO CANO

San Francisco - 25 MAR 2016 - 21:46 BRT



Fonte: [El País](#).

- Racista, sexista, xenófobo, nazista, etc.

Viés e Variância

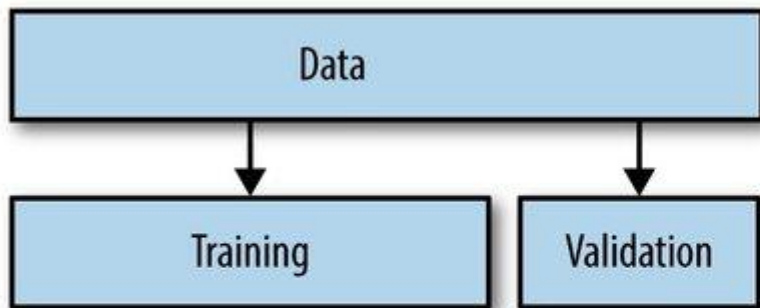
(exercícios)

Outros casos:

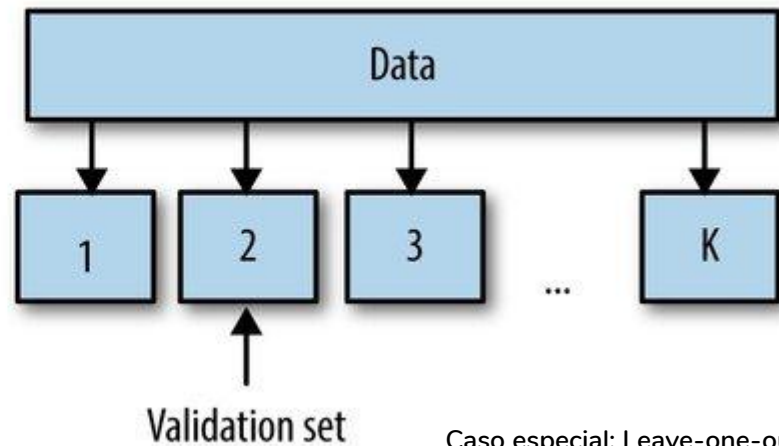
- Transsexual parado em aeroporto por “portar arma”;
- Recrutador racista;
- Ofertas de emprego com salários mais baixos para mulheres;
- “Se estiver na cozinha, mais chances de ser uma mulher”.

Validação cruzada

Hold-out validation



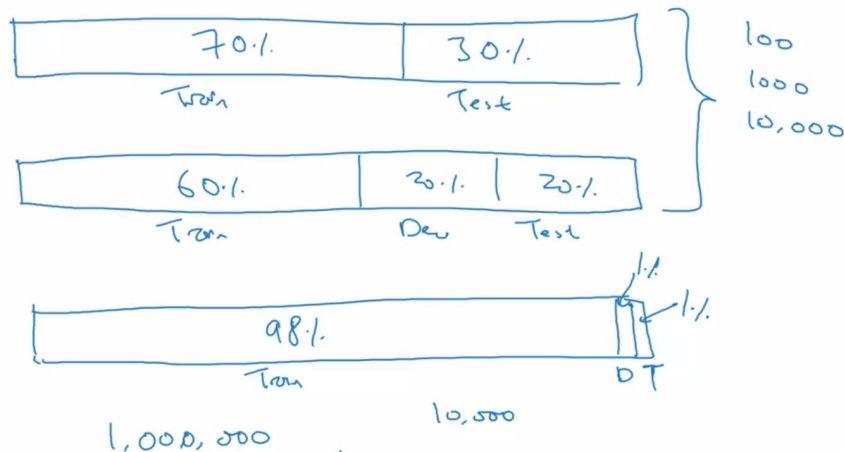
K-fold cross validation



Caso especial: Leave-one-out

Conjuntos de dados

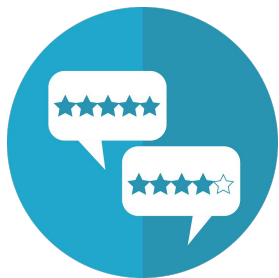
- **Treino:** os dados que iremos usar para treinar o modelo;
- **Validação:** os dados que responsáveis por validar o modelo;
- **Teste:** os dados que irão de fato testar o modelo.



Andrew Ng

Critérios de parada

Só um breve comentário sobre.



Feedback

Feedback dos alunos.

Feedback

<https://tinyurl.com/treinamento-aula01>