

Introdução a machine learning: entenda o aprendizado supervisionado vs aprendizado não supervisionado na prática (com Python)

Aline Bitencourt

Aline Bitencourt

- ▷ Sou graduada em Sistemas de Informação (Centro Universitário UNIFAMINAS). Mestra em Modelagem Matemática e Computacional (CEFET - MG). Já trabalhei como dev JAVA. Fui lecionar, comecei em um curso de informática voltado para mulheres em situação de vulnerabilidade social. Logo após, montei e ministrei um curso de programação web para meninas do último ano do ensino médio. Retomei o mercado corporativo e encontrei a de ciência de dados e me apaixonei pela área. Já atuo há mais de dois anos como cientista de dados. Estou sempre em busca de novos aprendizados e desafios e com muita vontade de contribuir com meus conhecimentos para termos mais mulheres conosco!
- ▷ LinkedIn: <https://www.linkedin.com/in/aline-sousa-791389109>

O que é Machine Learning?

- ▷ Essa técnica abrange a ideia de máquinas com a capacidade de aprenderem sozinhas a partir de grandes volumes de dados.
- ▷ Por meio de algoritmos (estatística/matемática) e dados, identificando padrões e criando conexões para aprender a executar uma tarefa sem a ajuda humana e de forma inteligente.

Aprendizado supervisionado vs Aprendizado não supervisionado



SUPERVISIONADO				NÃO SUPERVISIONADO	
var1	var2	resposta		var1	var2
Yellow	Orange	Red		Orange	Purple
Red	Yellow	Orange		Yellow	Green
Blue	Red	Purple		Purple	
Yellow	Blue	Green		Red	Green

se vira aí máquina!

Aprendizado supervisionado vs Aprendizado não supervisionado

- No supervisionado, o modelo recebe dados de entrada e de saída para buscar relações entre eles
- No não supervisionados, o modelo recebe apenas dados de entrada para encontrar padrões.

Aprendizado supervisionado

Aprendizado supervisionado

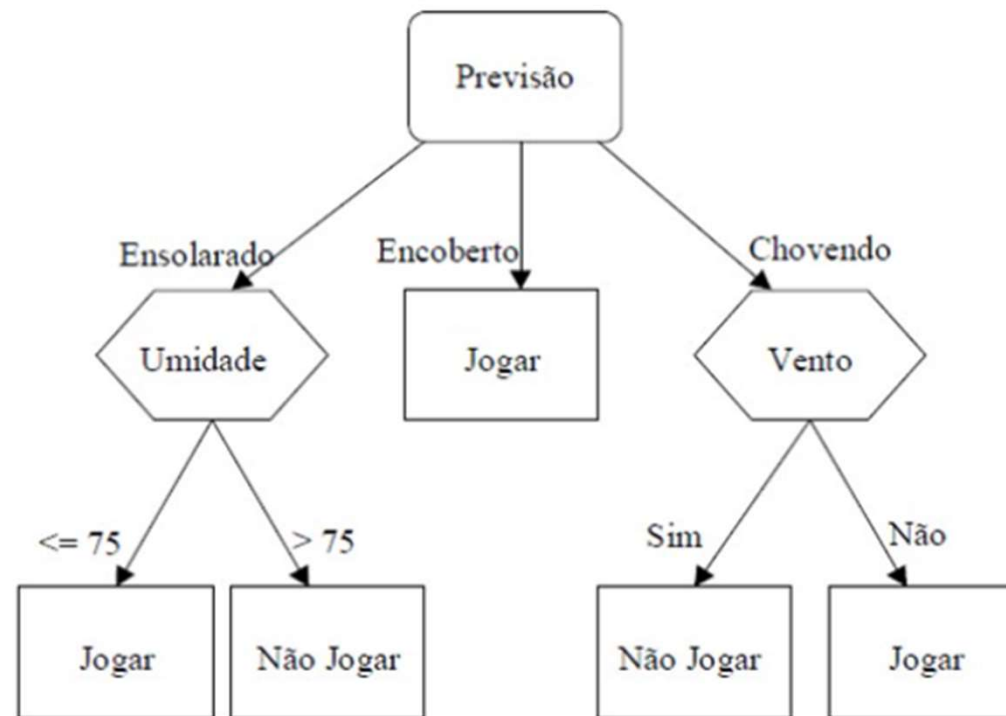
- ▷ **Classificação:** quero encontrar classes (categorias)
 - ▷ fraude e não fraude
 - ▷ spam e não spam
 - ▷ raças de cachorros

Exemplo – COMO FUNCIONA?

▷ Random Forest

Random Forest– Árvore de decisão

Exemplo clássico: Jogar tênis



Aprendizado supervisionado

- ▷ Vamos praticar!
- ▷ Problema: detecção de transações fraudulentas
 - ▷ Classificação

▷ E o código Python?

Aprendizado não supervisionado

Aprendizado não supervisionado

- ▷ Abordagem muito utilizada quando desejamos descobrir perfis ou padrões de comportamento, sem necessariamente ter um *rótulo* na análise.

Como agrupar os animais abaixo?

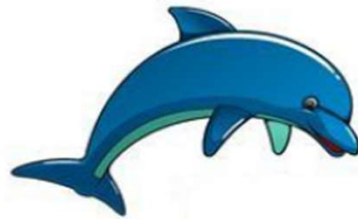


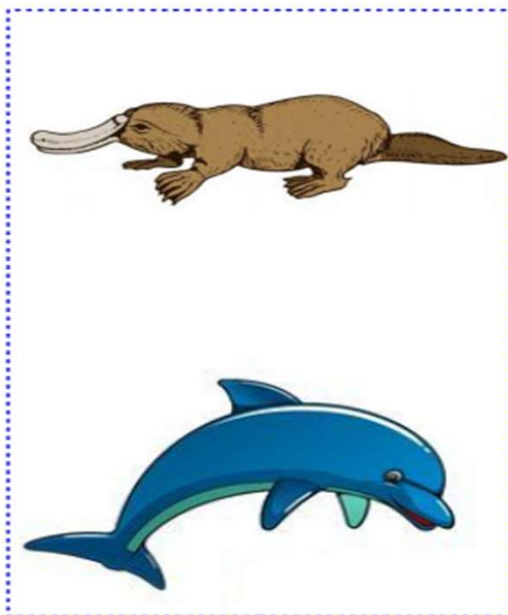
Fonte: http://www.modcs.org/wp-content/uploads/2015/01/Clustering_k_means_Agglomerative.pdf

Com bico

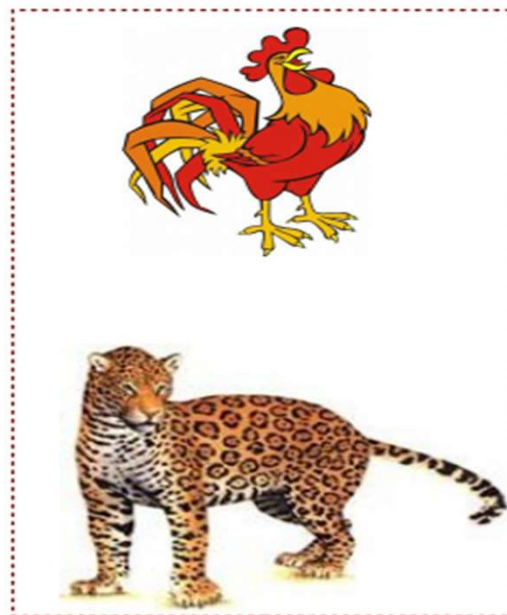


Sem bico





Aquático



Terrestre

Aprendizado não supervisionado

- ▷ Objetivo: separar os elementos em grupos
- ▷ Considera as características dos elementos
- ▷ No mesmo grupo teremos elementos similares
- ▷ Mas os grupos são diferentes um do outro

Aprendizado não supervisionado

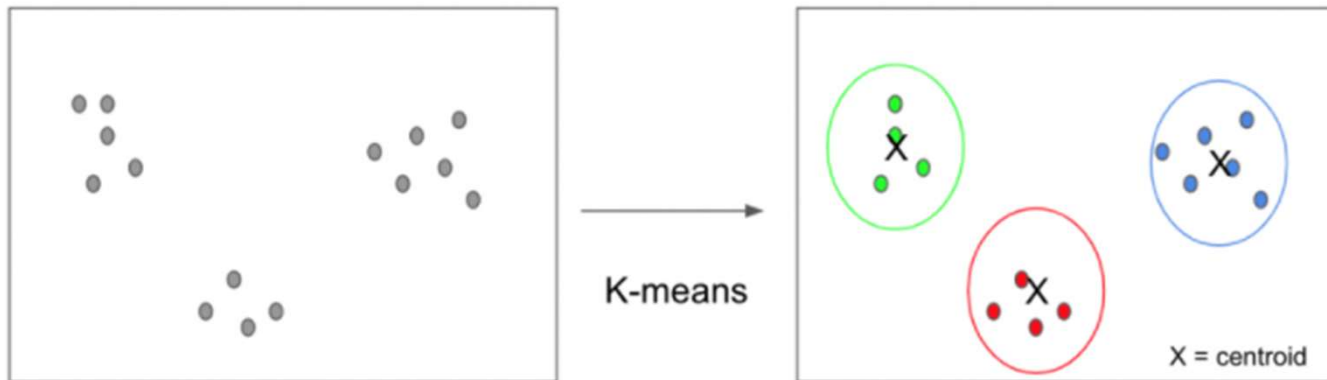
- ▷ Exemplos de Aplicações
 - ▷ Treinamento empresarial - agrupar funcionários por necessidade de treinamento
 - ▷ Marketing – criar grupos de clientes para divulgação de produtos de interesse

Kmeans

- ▷ Agrupa dados que possuem características similares entre si em um *cluster* (grupo).
- ▷ É necessário dizer ao algoritmo em quantos clusters desejo que seja feita a divisão dos dados - **K**

Kmeans – COMO FUNCIONA?

- ▷ Dado o valor de K , selecione k centróides (de maneira aleatória), sendo um para cada grupo
- ▷ Como o kmeans avalia a similaridade dos elementos através de medida de distância, ao termos os centróides, iremos associar cada elemento ao centróide mais próximo.



Fonte: <https://datascience.eu/pt/aprendizado-de-maquina/k-significa-clustering-no-aprendizado-de-maquinas/>

Aprendizado não supervisionado

- ▷ Vamos praticar!
- ▷ **Problema:** Entender alguns comportamentos dos clientes ao realizar compras em um E-commerce.
- ▷ Vamos agrupar clientes semelhantes, baseado nos dados de renda e no volume de gastos.
 - Assim, o marketing pode divulgar produtos para os grupos de clientes conforme perfil identificado

Dicas de aprendizado - gratuitos

1. [Microsoft Power BI para Data Science - Data Science Academy](#) (curso)
2. [minerandodados.com.br](#) (site muito bom - tem canal no youtube)
3. [Kaggle](#)
Com uma comunidade composta por mais de três milhões de usuários, o Kaggle é uma plataforma na qual os participantes aprendem muito ao participar das competições promovidas no site.

▷ Perguntas?