



Data Science do Zero

O curso completo para você aprender a trabalhar com essa tecnologia revolucionária e alavancar sua carreira.

Aprenda técnicas de Análise de Dados, Machine Learning e muito mais através de aulas 100% práticas partindo do nível zero até o avançado.

Quero aprender



Data Science do ZERO

Capítulo 06 - Machine Learning
K Nearest Neighbors – KNN

O que é K Nearest Neighbors - KNN

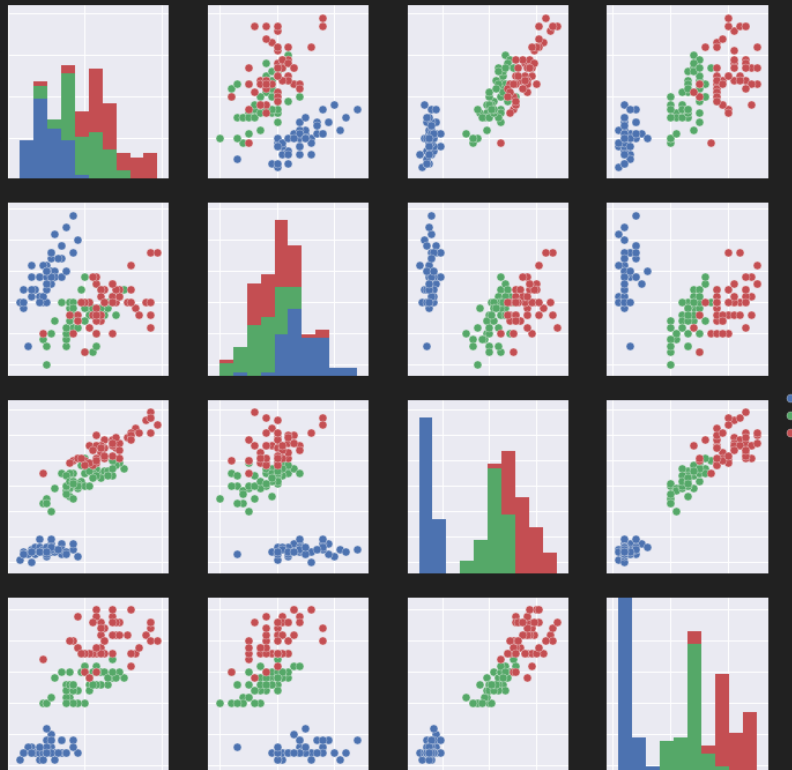
Vizinhos próximos

- Mais simples algoritmo de Machine Learning
- Método baseado em cálculo de distância.
- Pode ser utilizado para classificação quanto para regressão.
- Objetivo do KNN é determinar o rótulo de classificação de uma amostra baseado nas amostras vizinhas advindas de um conjunto de treinamento.



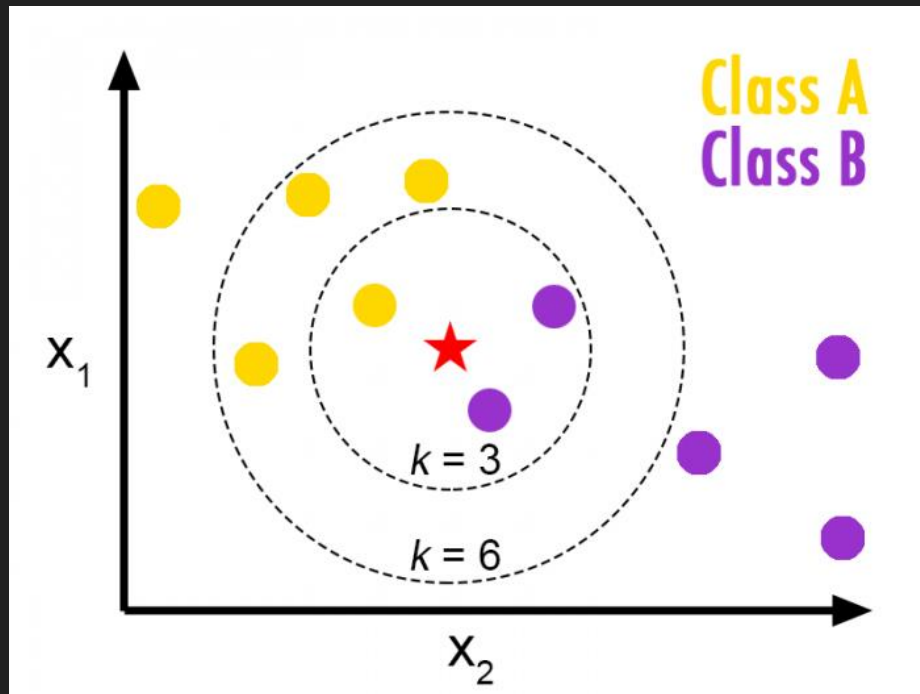
Conceitos sobre o KNN

- Hipótese é que dados semelhantes tendem a estar concentrados no mesmo espaço de dispersão.
- Considera a proximidade dos dados para realizar as previsões.
- Pode trabalhar com diferentes medidas de distância.
 - Mais utilizada é a distância Euclidiana



Conceitos sobre o KNN

- O parâmetro K especifica quantos vizinhos serão comparados.
- Quanto maior o valor de K, maior será o número de amostras a serem comparadas.



Parâmetros e otimização

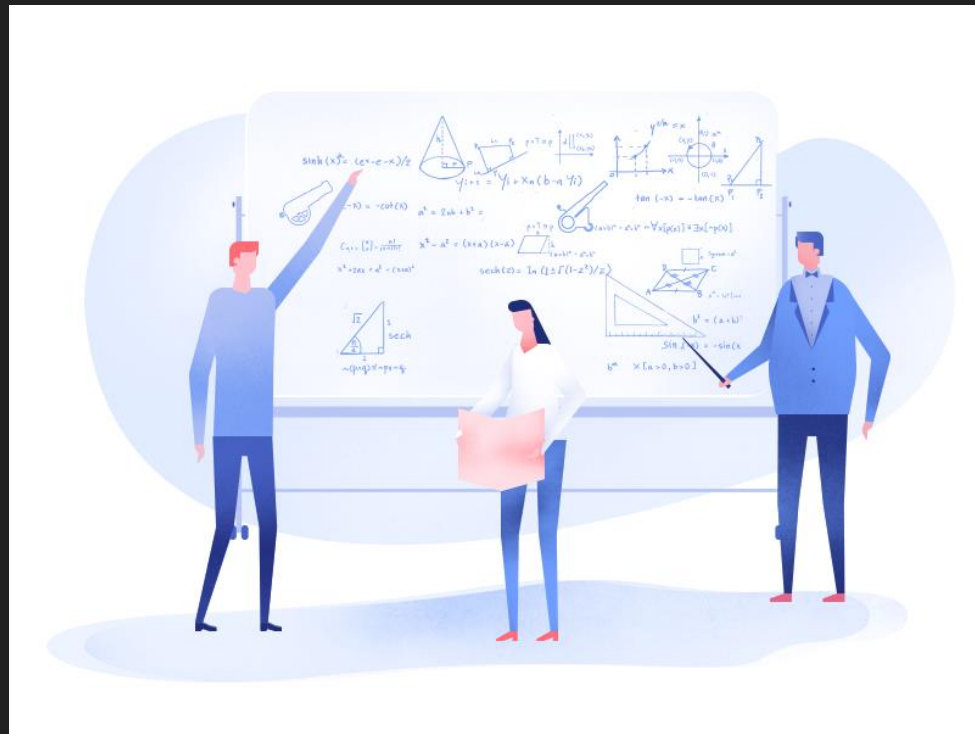
- A escolha do parâmetro K é de vital importância para o melhor resultado do modelo.
- O algoritmo se beneficia da possibilidade de utilização de diferentes métricas de distância:
 - Euclidiana
 - Manhattan
 - Minkowski
 - etc.



Algoritmo K-NN

Pseudocódigo do funcionamento do algoritmo.

```
1 inicialização:  
2   Preparar conjunto de dados de entrada e saída  
3   Informar o valor de  $k$ ;  
4 para cada nova amostra faça  
5   Calcular distância para todas as amostras  
6   Determinar o conjunto das  $k$ 's distâncias mais próximas  
7   O rótulo com mais representantes no conjunto dos  $k$ 's  
8   vizinhos será o escolhido  
9 fim para  
10 retornar: conjunto de rótulos de classificação
```



Entendendo o KNN

Dado o seguinte Dataset..

Altura (centímetros)	Peso (quilograma)	Tamanho Camisa
158	58	P
158	59	P
158	63	P
160	59	M
160	60	M
163	60	M
163	61	M



Entendendo o KNN

Dado o seguinte Dataset..

Altura (centímetros)	Peso (kilograma)	Tamanho Camisa
158	58	P
158	59	P
158	63	P
160	59	M
160	60	M
163	60	M
163	61	M

Nova amostra para fazer
predição com os valores
Altura 161 e Peso 61

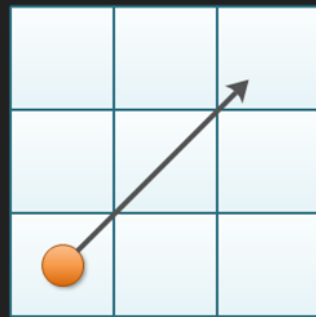


Entendendo o KNN

Dado o seguinte Dataset..

Altura (centímetros)	Peso (quilograma)	Tamanho Camisa
158	58	P
158	59	P
158	63	P
160	59	M
160	60	M
163	60	M
163	61	M

Nova amostra para fazer
predição com os valores Altura
161 e Peso 61

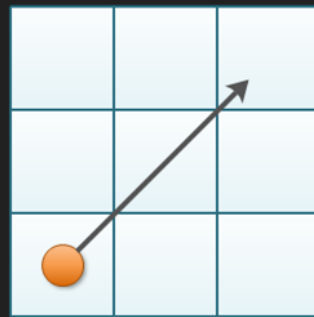


$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Entendendo o KNN

Dado o seguinte Dataset..

Altura (centimetros)	Peso (kilograma)	Tamanho Camisa	Distância
158	58	P	4.242640687119285
158	59	P	3.605551275463989
158	63	P	3.605551275463989
160	59	M	2.23606797749979
160	60	M	1.4142135623730951
163	60	M	2.23606797749979
163	61	M	2.0



$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

$$= \text{SQRT}((161 - 158)^2 + (61 - 58)^2)$$

Entendendo o KNN

Dado o seguinte Dataset..

Altura (centímetros)	Peso (kilograma)	Tamanho Camisa	Distância	
158	58	P	4.242640687119285	
158	59	P	3.605551275463989	
158	63	P	3.605551275463989	5
160	59	M	2.23606797749979	4
160	60	M	1.4142135623730951	1
163	60	M	2.23606797749979	3
163	61	M	2.0	2

Para K = 5

Entendendo o KNN

Dado o seguinte Dataset..

Altura (centimetros)	Peso (kilograma)	Tamanho Camisa	Distância	
158	58	P	4.242640687119285	
158	59	P	3.605551275463989	
158	63	P	3.605551275463989	
160	59	M	2.23606797749979	
160	60	M	1.4142135623730951	1
163	60	M	2.23606797749979	3
163	61	M	2.0	2

Para K = 3

Hands on!