



INSTITUTO DO EMPREGO E FORMAÇÃO PROFISSIONAL, IP
DELEGAÇÃO REGIONAL DE LISBOA E VALE DO TEJO
CENTRO DE EMPREGO E FORMAÇÃO PROFISSIONAL DE SINTRA

UFCD – 10810

Fundamentos do desenvolvimento de
modelos analíticos em Python

4 – K Nearest Neighbors

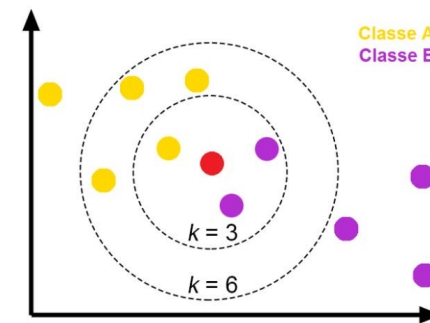
Carga horária: 25 horas
Formador: Manuel Viana



1

O que é o KNN?

- É um algoritmo de classificação, estimativa e predição.
- Olha para os vizinhos mais próximos de K para procurar uma classificação.
- O KNN assume que, coisas semelhantes existem nas proximidades, ou seja, estão próximas umas das outras.

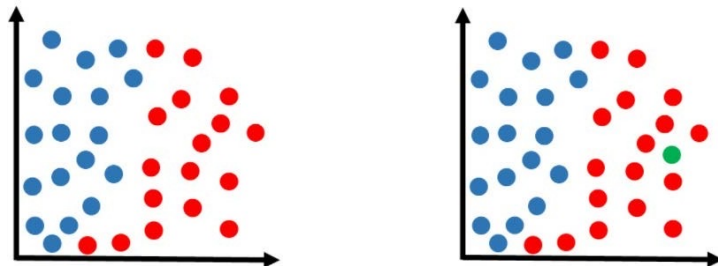


2

2

KNN – Exemplo

1. Considere-se um conjunto de dados dividido em duas classes: azul e vermelho.
2. Após recebermos uma amostra (ponto verde) que ainda não está classificada, gostaríamos de definir se ela pertence à classe azul ou à classe vermelha.

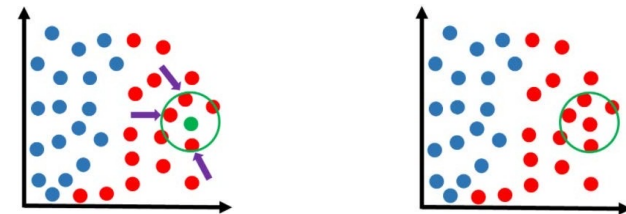


3

3

KNN – Exemplo

- Intuitivamente, podemos observar que faz mais sentido classificar a amostra como pertencendo à classe vermelha.
- No entanto, o algoritmo não possui “intuição”, visto que necessita de um cálculo matemático para poder definir a solução.
- No caso do KNN, a lógica é a seguinte:
 - Observa-se a classe dos vizinhos mais próximos, numa votação onde a maioria vence.
 - Por exemplo, vamos supor que estamos a analisar os 3 vizinhos mais próximos (com a menor distância em relação à amostra).



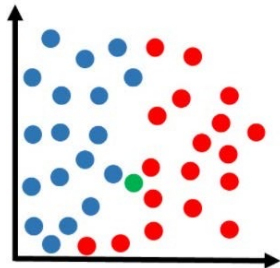
- Os 3 vizinhos mais próximos pertencem à classe vermelha. Como houve 3 votos a zero para a classe vermelha, a amostra fica classificada nesta classe.

4

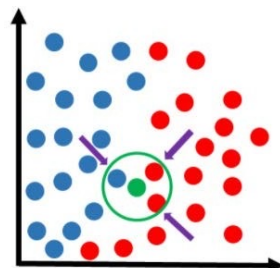
4

KNN – Exemplo

3. Agora recebemos outra amostra que queremos classificar.



4. Utilizando o mesmo método KNN com $k=3$, encontramos os 3 vizinhos mais próximos da amostra.

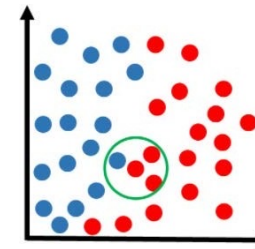


5

5

KNN – Exemplo

- Há duas amostras da classe vermelha e uma da classe azul.
- Como a votação ficou 2x1 para a classe vermelha, a amostra fica classificada nesta classe.



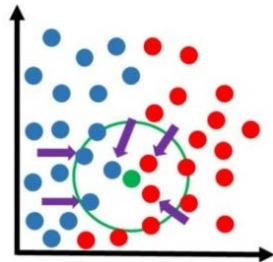
- Esta metodologia pode ser aplicada para qualquer nova amostra e estaríamos aptos a definir a sua devida classificação.
- Contudo, utilizou-se apenas o exemplo de $K=3$.
- Na prática, podemos escolher outro valor de K .

6

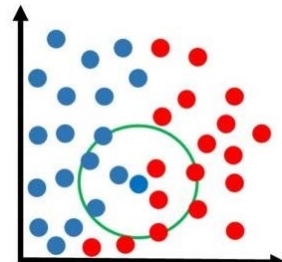
6

KNN – Exemplo

5. Vamos supor que a amostra anterior está a ser analisada com o algoritmo KNN com $K=5$.



6. Desta vez, dos 5 vizinhos mais próximos, 3 são azuis e 2 são vermelhos. Portanto a classe vencedora foi a azul. A amostra seria classificada nesta classe.

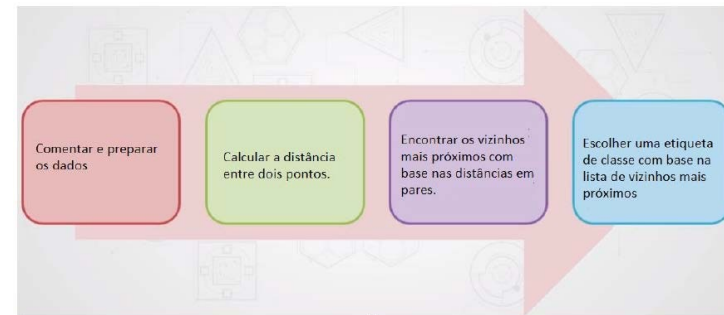


7

7

Como funciona o KNN?

1. Recebe um dado não classificado e mede a distância do novo dado em relação a cada um dos outros dados já classificados.
2. Seleciona as K menores distâncias.
3. Verifica a(s) classe(s) dos dados que tiveram as K menores distâncias e contabiliza a quantidade de vezes que cada classe apareceu.
4. Classifica o novo dado consoante a classe mais contabilizada.



8

8

Como se decide o número de vizinhos?

- O número de vizinhos (K) no KNN, é um hiper parâmetro que necessitamos de escolher quando construímos o modelo.
- Nenhum número ótimo de vizinhos se adequa a todos os tipos de conjuntos de dados. Cada conjunto de dados tem os seus próprios requisitos.
- No caso de um número pequeno de vizinhos, o ruído terá uma influência maior no resultado.
- Um número grande de vizinhos torna-o computacionalmente caro.
- Geralmente, escolhe-se um número ímpar se o número de classes for par.
- Também se pode verificar, gerando o modelo em diferentes valores de K e verificar o seu desempenho.
- Ou então, tentar o método *Elbow* (cotovelo).