KNN
k-nearest neighbors (KNN)

INFORMAÇÃO,

TECNOLOGIA

& INOVAÇÃO

KNN

- O algoritmo dos k-vizinhos mais próximos (KNN) é um algoritmo de aprendizado de máquina supervisionado simples e fácil de implementar que pode ser usado para resolver problemas de classificação e regressão.

- assume que coisas semelhantes estão próximas umas das outras.

KNN

- O algoritmo KNN usa "similaridade de instâncias" para predizer o valor de um atributo de novos exemplos.

- Um valor é associado ao atributo de predição com base na similaridade com as instâncias do conjunto de treinamento.

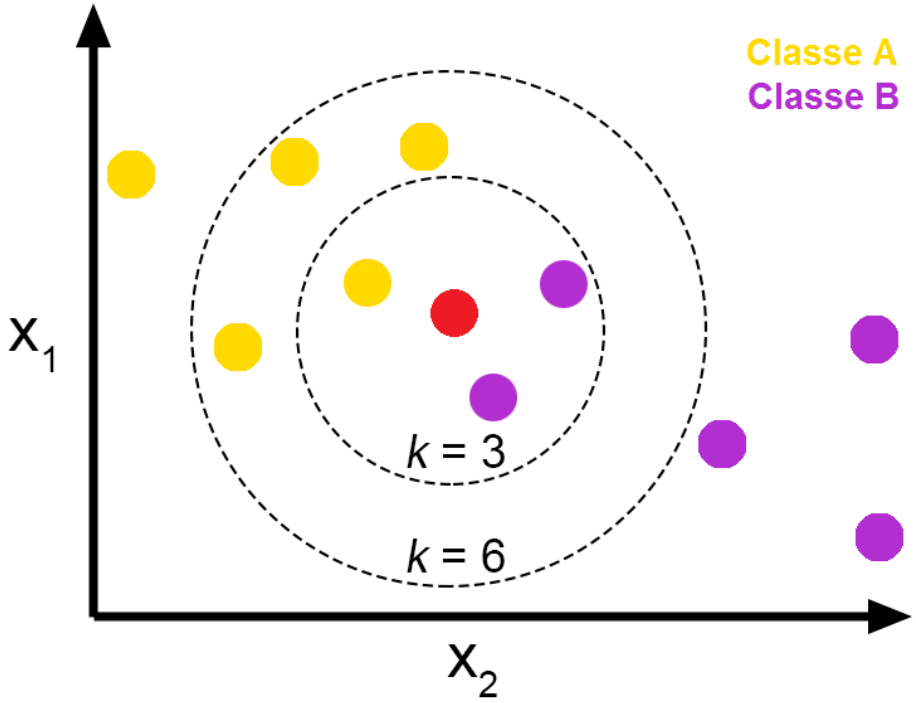
KNN

- Dados: uma Base de dados de *m* instâncias previamente rotuladas.
- Importante: para problemas de regressão o rótulo é um valor contínuo. Para problemas de classificação o rótulo é um valor categórico.

- Dado um exemplo de teste X = (x1,...,xn), cujo rótulo não é fornecido

- Calcula-se a distância de X a cada uma das instâncias.
- Pega-se as k instâncias mais próximas (similares) de X.
- X é rotulado com a média (regressão) ou a moda (classificação) das k-instâncias mais próximas a X.

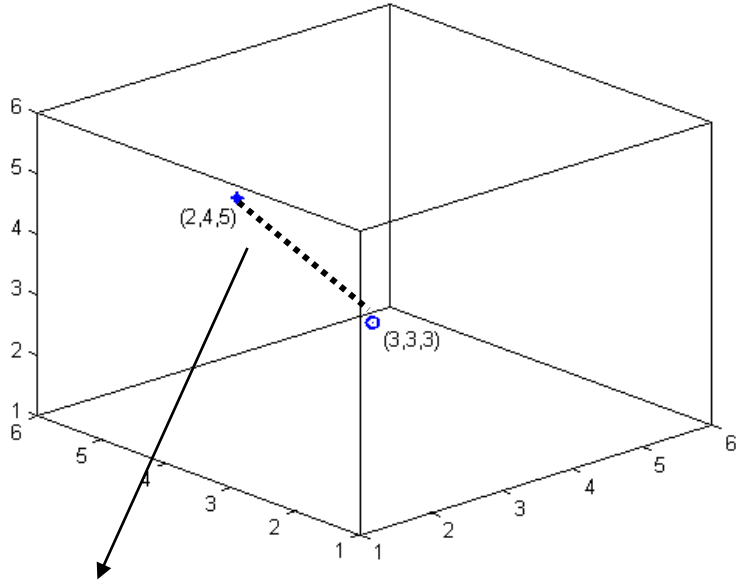
KNN



Fonte da Figura: https://towardsdatascience.com/knn-k-nearest-neighbors-1-a4707b24bd1d

KNN

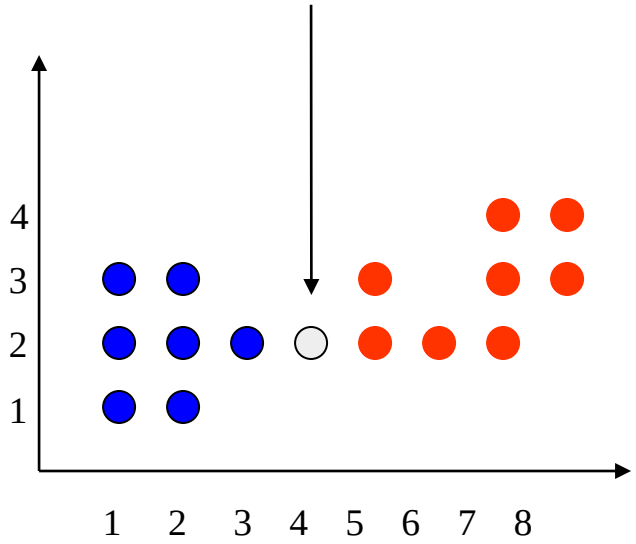
- Uma medida de proximidade bastante utilizada é a distância Euclidiana:



$$d(x,y) = \sqrt{(2-3)^2 + (4-3)^2 + (5-3)^2} = \sqrt{6} = 2.44$$

KNN - EXEMPLO

A qual classe pertence este ponto?
Azul ou vermelho?



Calcule para os seguintes valores de k:

k=1 não se pode afirmar

k=3 vermelho – 5,2 - 5,3

k=5 vermelho – 5,2 - 5,3 - 6,2

k=7 azul – 3,2 - 2,3 - 2,2 - 2,1

A classificação pode mudar de acordo com a escolha de *k*.
***k* pode ser escolhido por validação cruzada!**

KNN

Considerações

- Performance
  - Não constrói um modelo de aprendizado.
  - Processo de classificação de um exemplo é lento.
  - Utiliza todos os dados para fazer a predição.

- Sensível a ruídos
  - KNN faz predição baseando-se em informações locais ao exemplo a ser classificado.
  - Arvores de decisão, regressão logística e redes neurais encontram modelo global que se leva em conta todo o banco de dados de treinamento.