

Classificador KNN

Agente Educacional

Sérgio M. Dias

Agenda

KNN (k-Nearest Neighbors)
Exercício

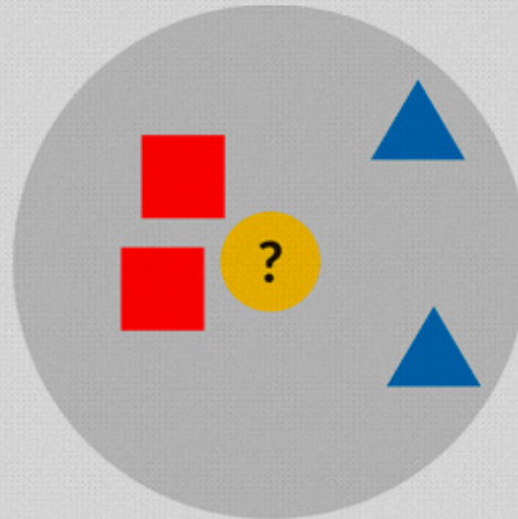


KNN (*K-Nearest Neighbors*)

É um dos algoritmos de classificação mais simples.

É usado para classificar objetos com base em exemplos de treinamento que estão mais próximos no espaço de características.

“?” está mais próximo
dos triângulos ou quadrados?



Para utilizar o **KNN** é necessário:

- Um conjunto de **exemplos de treinamento**.
- Definir uma **métrica** para calcular a distância entre os exemplos de treinamento.
- Definir o **valor de K** (o número de vizinhos mais próximos que serão considerados pelo algoritmo).

Classificar um exemplo desconhecido com o algoritmo KNN consiste em três passos:

- Calcular a **distância** entre o exemplo desconhecido e o outros exemplos do conjunto de treinamento.
- Identificar os **K** vizinhos mais próximos.
- Utilizar o **rótulo da classe** dos vizinhos mais próximos para determinar o rótulo de classe do exemplo desconhecido (votação majoritária).

Diversas métricas de distância podem ser usadas:

- Euclidiana
- Hamming
- Cosseno
- Manhattan

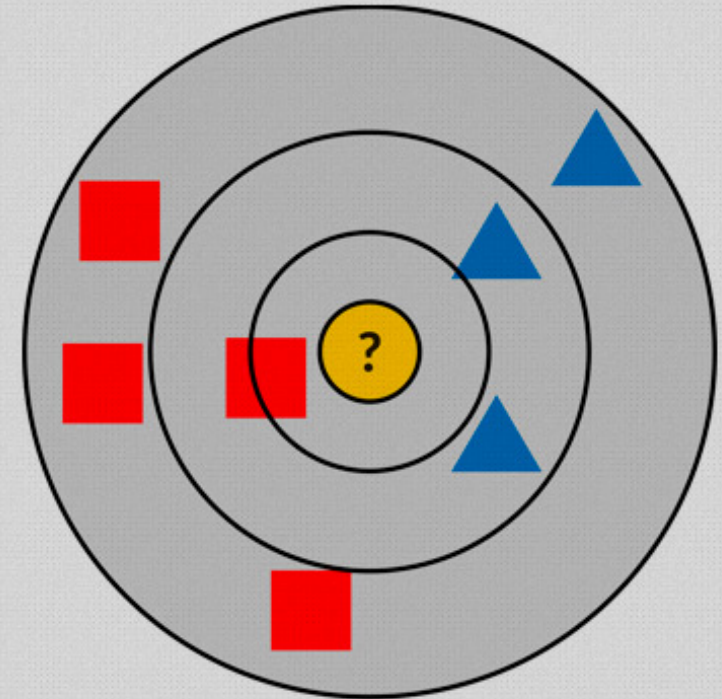
Determinando a classe do exemplo desconhecido a partir da lista de vizinhos mais próximos:

- Considera-se o **voto majoritário** entre os rótulos de classe dos **K** vizinhos mais próximos.
- Como escolher o **valor de K**?

$K=1 \rightarrow$ **QUADRADOS**

$K=3 \rightarrow$ **TRIÂNGULOS**

$K=7 \rightarrow$ **QUADRADOS**



Como escolher o valor de K ?

- Se K for muito pequeno, a classificação fica sensível a pontos de ruído.
- Se K é muito grande, a vizinhança pode incluir elementos de outras classes.
- Além disso, é importante sempre escolher um valor ímpar para K , assim se evita empates na votação.

A **precisão da classificação** utilizando o algoritmo KNN depende fortemente do modelo de dados.

Na maioria das vezes os **atributos precisam ser normalizados** para evitar que as medidas de distância sejam dominadas por um único atributo.

Exemplos

- **Altura** de uma pessoa pode variar de 1,20m a 2,10m
- **Peso** de uma pessoa pode variar de 40kg a 150kg
- O **salário** de uma pessoa podem variar de R\$ 800,00 a R\$ 20.000,00

Vantagens

- Técnica **simples** e **fácil** de implementar
- Bastante **flexível**
- Em alguns casos apresenta **bons resultados**

Desvantagens

- Classificar um exemplo desconhecido pode ser **custoso computacionalmente**, pois requer calcular a distância do exemplo desconhecido para cada exemplo de treinamento.
- A precisão da classificação pode ser **severamente degradada pela presença de ruído** ou características irrelevantes.



<Exercício *KNIME KNN*>

Obrigado!

Agente Educacional

Sérgio M. Dias

sergio.dias@serpro.gov.br | #31 6539

Demais agentes educacionais sobre o assunto

Marcelo Pita | marcelo.pita@serpro.gov.br | #81 8794

Gustavo Torres | gustavo.gamatorres@serpro.gov.br | #31 6950