

O que é o algoritmo KNN?

Társila Samille

Introdução

O algoritmo KNN (k-Nearest Neighbors) é um classificador de aprendizado supervisionado não paramétrico.

Ele utiliza a proximidade para fazer classificações ou previsões sobre o agrupamento de um ponto de dados individual.

É um dos classificadores mais populares e simples usados atualmente em machine learning.

Funcionamento

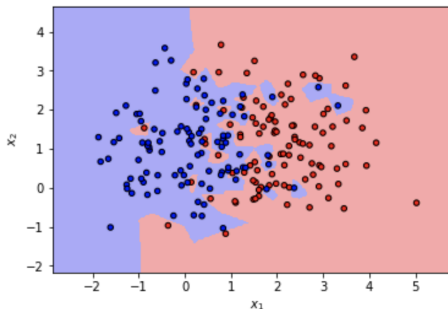
Classificação

O KNN baseia-se na suposição de que pontos semelhantes estão próximos uns dos outros.

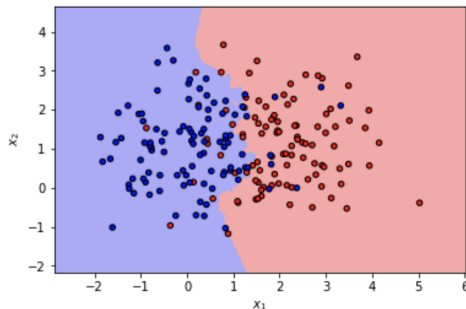
Um rótulo de classe é atribuído por meio de uma votação majoritária dos k vizinhos mais próximos.

A distância euclidiana é comumente usada para determinar a proximidade.

Funcionamento



(a) $k = 1$



(b) $k = 5$

Figura. Distribuição dos dados de treinamento e fronteira de decisão associada ao kNN para diferentes valores de k . À medida que k aumenta, a fronteira tende a ficar mais suave e menos regiões isoladas são criadas para cada classe.

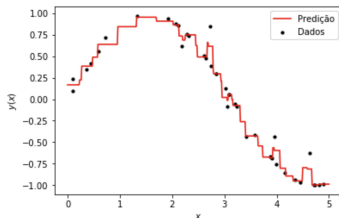
Figure 1: Classificação

Funcionamento

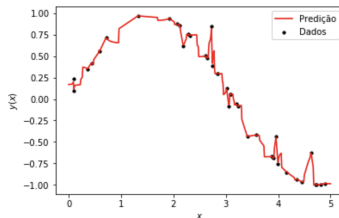
Regressão

A média dos k vizinhos mais próximos é usada para fazer uma previsão sobre uma classificação.

É semelhante à classificação, porém com valores contínuos.



(a) $k = 2$, pesos uniformes



(b) $k = 2$, pesos inversamente proporcionais à distância

Figura. Dados de treinamento e mapeamento gerado pelo kNN.

Figure 2: Regressão

Características

- Simples e fácil de entender: Ideal para iniciantes em machine learning.
- Versátil: Pode ser usado para classificação e regressão.
- Eficaz: Pode ser muito preciso em alguns casos.
- Preguiçoso: Armazena o conjunto de dados de treinamento e computa as previsões no momento da classificação.

Limitações

- Escalabilidade: Pode ser lento e ineficiente com grandes conjuntos de dados.
- Sensibilidade ao ruído: Pode ser influenciado por valores inconsistentes nos dados.
- Escolha de k : O valor de k pode afetar significativamente o desempenho do modelo.

Aplicações

- Pré-processamento de dados: Imputação de valores ausentes.
- Motores de recomendação: Previsão de conteúdo de interesse para usuários.
- Finanças: Previsão de preços de ações e outras variáveis financeiras
- Detecção de fraudes: Identificação de transações fraudulentas.
- Reconhecimento de imagem: Classificação de imagens.

Conclusão

O KNN é um algoritmo fundamental em machine learning com diversas aplicações.

Apesar de suas limitações, é uma ferramenta útil para iniciantes e para problemas específicos.