



K-Nearest Neighbors - Teoria

☰ Ciclo	Ciclo 02: Aprendizado supervisionado - Classificação
# Aula	11
🕒 Created	@January 5, 2023 4:41 PM
☑ Done	☑
☑ Ready	☑

Objetivo da Aula:

- ☐ O que é KNN?
- ☐ O processo de treinamento do KNN
- ☐ Premissas assumidas do KNN
- ☐ Resumo
- ☐ Na próxima aula

Conteúdo:

▼ 1. O que é KNN?

KNN significa K-Nearest Neighbors. Em português, ficaria algo como “K-vizinhos mais próximos”.

O KNN é um algoritmo de Machine Learning da classe de aprendizado supervisionado. Ele pode ser utilizado tanto para resolver problemas de classificação quanto problema de regressão.

Nessa ciclo, vamos estudar o KNN como algoritmo supervisionado para resolver um problema de classificação.

▼ 2. O processo de treinamento do KNN

▼ 2.1 Exemplo visual do funcionamento.

1.

▼ 2.2 Os 8 passos para treinar o KNN

Passo 01: Carregue os dados

Passo 02: Defina o número K vizinhos mais próximos.

Passo 03: Para cada exemplo sem rótulo:

Passo 04: Calcule a distância entre o ponto sem rótulo e todos os outros pontos rotulados do conjunto de dados.

Passo 05: Encontre o K pontos mais próximos.

Passo 06: Identifique os rótulos de cada um dos K vizinhos mais próximos.

Passo 07: Calcule a “moda” dos K vizinhos mais próximos.

Passo 08: Classifique o ponto sem rótulo com a moda de seus k vizinhos mais próximos.

▼ 2.3 Como calcular a distância entre 2 pontos?

▼ 2.3.1 Euclidian distance

$$d = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2}$$

▼ 2.3.2 Manhattan distance

$$d = |P_x - Q_x| + |P_y - Q_y|$$

▼ 2.3.3 Mahallanobis distance

$$d = \sqrt{\frac{(P_x - Q_x)^2 + (P_y - Q_y)^2}{\sigma_{xy}^2}}$$

▼ 2.4 Como calcular a moda de um conjunto de ponto?

$$A = \{2, 23, 4, 2, 5\}$$
$$M_o = 2$$

1. Conte quantos pontos pertencem ao rótulo A e ao rótulo B.
2. A moda é o número que aparece mais frequente.

▼ 3. Premissas assumidas do KNN

A premissa assumida pelo KNN é a seguinte:

A proximidade entre duas evidências de um fenômeno observado, pode ser representada por uma medida de distância.

Coisas similares estão próximas uma das outras, portanto, todas as suas características precisam ser numéricas.

▼ 4. Resumo

1. KNN é um algoritmo de Machine Learning com aprendizado supervisionado que resolve problemas de classificação e regressão.
2. O funcionamento do KNN é baseado no cálculo da distância entre um ponto sem rótulo e todos os seus K vizinhos mais próximos. O rótulo que possuir o maior número de vizinhos é atribuído ao ponto.
3. Para escolher o melhor valor de K para os dados, você precisa testar o algoritmo para vários valores diferentes de K e encontrar aquele que proporciona o menor erro do algoritmo.

▼ 4.1. Em outras palavras



“KNN é tipo: Me mostre os seus amigos e eu te digo quem você é”

▼ 5. Na próxima aula

K-Nearest Neighbors - Prática