# ★ Explicação melhorada sobre KNN – k-Nearest Neighbors

O KNN (*k-nearest neighbors*), ou **K vizinhos mais próximos**, é um algoritmo de aprendizado de máquina do tipo **supervisionado** e baseado em **instâncias**. Ele é utilizado principalmente para **classificação** (prever categorias) e também pode ser usado para **regressão** (prever valores numéricos).

## Como funciona:

## 1. Representação dos dados:

Cada amostra (linha de dados) é representada por um conjunto de **valores preditores** (ex.: peso, altura, idade etc.). Esses valores podem ser imaginados como **coordenadas** em um espaço n-dimensional, como se fosse um plano cartesiano com várias dimensões.

#### 2. Normalização:

Como os dados podem ter escalas diferentes (ex.: idade em anos e peso em quilos), é importante **normalizar** ou **padronizar** os valores preditores. Isso evita que variáveis com números maiores tenham mais influência no cálculo da distância.

#### Distância:

Para classificar um novo dado, o KNN calcula a **distância** entre esse novo ponto e todos os outros pontos do conjunto de treinamento. O método de distância mais comum é o **euclidiano** (como medir com uma régua), mas outros métodos como a distância de Manhattan também podem ser usados.

## 4. Escolha dos vizinhos (K):

Define-se o valor de **K**, que representa o número de vizinhos mais próximos a considerar. O algoritmo seleciona os **K dados mais próximos** com base na distância calculada.

#### 5. Classificação:

O novo dado é classificado de acordo com a classe mais frequente entre os K vizinhos. Ou seja, ele "herda" a categoria mais comum dos vizinhos próximos.

## Exemplo prático (classificação):

 Suponha que temos um conjunto de frutas com características como cor, peso e diâmetro.

- Recebemos uma nova fruta com essas características.
- O KNN irá comparar essa nova fruta com todas as frutas conhecidas, encontrar as mais parecidas (os K vizinhos) e decidir se é, por exemplo, uma maçã ou uma laranja, com base na maioria entre os vizinhos.

## **⚠** Pontos importantes:

- O KNN não aprende de forma antecipada (não tem treinamento real). Ele apenas compara no momento em que precisa prever.
- Ele pode ser lento com muitos dados, pois precisa calcular distâncias para todos os pontos.
- A escolha do valor de K é fundamental. K muito pequeno pode causar ruído; K muito grande pode gerar resultados imprecisos.