



## EXERCÍCIO 2

# $k$ -VIZINHOS MAIS PRÓXIMOS

### Objetivo

Neste exercício, iremos implementar um dos mais tradicionais algoritmos de Aprendizado de Máquina: o  $k$ -vizinhos Mais Próximos (KNN).

O  $k$ -vizinhos Mais Próximos é um método de aprendizado supervisionado baseado em distância. Devido a sua baixa complexidade e facilidade de implementação, é um método bastante utilizado. Mesmo sendo simples, possui capacidade de resolver problemas complexo e tem a vantagem de ser facilmente incremental.

Ao término desse exercício, espera-se que você consiga implementar o método de  $k$ -vizinhos Mais Próximos, assim como funções de distância que podem ser utilizadas pelo próprio método ou em outros problemas da área. O KNN é um algoritmo facilmente adaptável para qualquer base de dados, servindo como uma ótima primeira abordagem para solucionar diversos problemas reais com métodos inteligentes.

### O exercício

Ao longo do exercício, você deverá completar três funções:

- **normalizar**: responsável por normalizar os dados, preparando-o para o método de aprendizado;
- **distancia**: responsável por calcular a distância euclidiana entre uma amostra e um conjunto de dados, uma das principais etapas do KNN;
- **knn**: o método em si, que deve utilizar a função **distancia** para encontrar os  $k$ -vizinhos mais próximos de uma amostra-alvo, e realizar a previsão baseado na classe destes vizinhos.

Durante o exercício, também será mostrado de forma visual o resultado da aplicação do algoritmo, ensinando como é possível gerar uma visualização da superfície de decisão aprendida pelo método.

Preencha o código apenas nos espaços delimitados por comentários, normalmente iniciados por um comentário “COMPLETE O CÓDIGO AQUI” e instruções para a implementação.

As implementações devem ser o mais genéricas possíveis, funcionando para conjuntos de dados além dos fornecidos com o exercício. Não adicione comandos do tipo `print` ou `display` dentro das funções que serão completadas, apenas o código da função..

Aproveite para interagir com o *notebook*: teste valores diferentes para as amostras-alvo, rode com diferentes valores de  $k$  e entenda como o método funciona e como a escolha de parâmetros pode influenciar no resultado.

## Os casos de teste

Este exercício possui **5 casos de teste** que buscam avaliar cada uma das funções implementadas. A distribuição de tarefas avalaiadas por cada caso é feita da seguinte forma:

- **Caso de teste 1:** corrige a função `normalizar`;
- **Casos de teste 2 e 3:** corrigem a função `distancia`;
- **Casos de teste 4 e 5:** corrigem a função `knn`;

Atente-se que, na execução do `knn`, a função `distancia` é utilizada. Assim, um erro na implementação da função também pode gerar problemas para os casos de teste que avaliam o `knn`.