



EXERCÍCIO 1

PROCESSAMENTO E VISUALIZAÇÃO DE DADOS

Objetivo

Neste exercício, iremos trabalhar com o processamento de dados e visualização.

Ao resolver um problema utilizando Aprendizado de Máquina, a maneira como os dados são preparados pode causar até mais impacto do que o algoritmo escolhido. Dessa forma, é muito importante que os dados sejam compreendidos e devidamente tratados.

Técnicas de visualização são bastante úteis para exibir, de maneira sumarizada, características interessantes dos dados, e técnicas tradicionais de pré-processamento podem deixá-los em formatos melhores para serem consumidos por métodos de aprendizado.

Ao término deste exercício, espera-se que você entenda como é realizada uma etapa tradicional de pré-processamento na resolução de problemas de Aprendizado de Máquina. Todos os procedimentos implementados podem ser facilmente aplicados em qualquer outra base de dados.

O exercício

Ao longo do exercício, você deverá completar 6 funções:

- **trataFaltantes**: responsável pelo tratamento de amostras faltantes preenchendo-as com o valor médio dos atributos para amostras da mesma classe;
- **delDuplicatas**: responsável pela remoção de amostras duplicadas, mantendo apenas uma;
- **delInconsistencias**: responsável pela remoção de amostras inconsistentes;
- **normalizar**: responsável por realizar a normalização da base, deixando seu conteúdo com média 0 e desvio padrão 1;
- **removeOutliers**: responsável por remover os *outliers* da base utilizando o método da amplitude inter-quartil;
- **covariancia**: responsável por calcular a covariância entre amostras;

- **correlacao**: responsável por calcular a correlação entre amostras.

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 genéricas e funcionar para qualquer conjunto de dados. Na avaliação, as funções serão testadas em bases com quantidade de amostras e atributos diferentes das fornecidas com o exercício, tendo em comum apenas o nome da coluna que contém a classe das amostras.

Os casos de teste

Este exercício possui **5 casos de teste** que buscam avaliar cada uma das funções implementadas, de forma incremental. Assim, os casos iniciais corrigem apenas um subconjunto de tarefas, aumentando de acordo com o número do caso. A distribuição de tarefas é feita da seguinte forma:

- **Caso de teste 1:** corrige a tarefa 1;
- **Caso de teste 2:** corrige as tarefas 1 a 3;
- **Caso de teste 3:** corrige as tarefas 1 a 4;
- **Caso de teste 4:** corrige as tarefas 1 a 5;
- **Caso de teste 5:** corrige todas as tarefas.