

Aprendizagem de máquina - exercício de implementação

O objetivo desse exercício é implementar o algoritmo KNN para classificação utilizando um dataset que contém dados sobre flores.

Tarefas

- Ler o arquivo com os dados de treinamento
- Implementar o KNN
- Treinar o KNN (podendo variar o valor de K)
- Efetuar classificação

Arquivos

O arquivo de treinamento é o **treinamento.csv**. Ele é um arquivo .csv (comma separated values) que contém informações sobre as medidas de algumas partes de flores (pétalas e sépalas) e suas respectivas classificações. As medidas são valores em centímetros e as classificações podem ser 0, 1 ou 2.

O arquivo segue o padrão:

```
sepalwidth,sepalwidth,petalwidth,petalwidth,label
5.1,3.5,1.4,0.2,0
4.9,3.0,1.4,0.2,0
4.7,3.2,1.3,0.2,1
4.6,3.1,1.5,0.2,2
...
```

A primeira linha do arquivo especifica o nome das colunas e as linhas seguintes os dados das flores. Cada coluna é separada por vírgula. Cada linha seguinte possui os dados de uma flor. Por exemplo, a primeira linha após a definição dos nomes das colunas, contém os dados de um flor que possui sépalas com tamanho 5.1 e largura 3.5, pétalas com tamanho 1.4 e largura 0.2, e é uma flor do tipo 0.

O arquivo de teste é o **teste.csv**. O arquivo segue o mesmo padrão do arquivo de treinamento, com dados de novas flores, mas sem as classificações.

O arquivo **rotulos-teste.txt** possui as classificações para cada flor do arquivo de teste. A primeira linha do arquivo rotulos-teste.txt tem a classificação da primeira flor do arquivo de teste, a segunda linha tem a classificação da segunda flor e assim sucessivamente.

Resposta

A resposta da classificação deve ser escrita no arquivo **resultado.txt**. Ela deve ter o mesmo formato do arquivo rotulos-teste.txt, ou seja, o resultado da classificação da primeira flor do arquivo de teste ficará na primeira linha, o resultado da segunda na segunda linha, etc.

Também deve ser exibido a porcentagem de acerto do classificador. Para isso, basta comparar o número de respostas no arquivo resultado.txt que são iguais ao do rotulos-teste.txt.

FAQ

1. **Em qual linguagem eu posso implementar o algoritmo?** Java ou Python.
2. **É para fazer individualmente ou em grupo?** Pode ser feito sozinho ou em dupla.
3. **É necessário acertar todas as classificações?** Não. Muitas vezes os classificadores não conseguem acertar 100% e isso pode acontecer nesse problema também.
4. **Vale ponto?** Sim. O exercício vale até 2 pontos na segunda prova.