

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE MATEMÁTICA
PRIMEIRA PROVA DE MÉTODOS MATEMÁTICOS - 09/10/2023

O objetivo desta prova é construir um modelo de classificação capaz de prever a espécie de um papagaio do mar. O modelo que você deve utilizar é o KNN (k nearest neighbors ou, em português, os k vizinhos mais próximos). O conjunto de dados está no arquivo `papagio.csv` e as informações sobre este conjunto podem ser lidas em `README.txt`.

- 1 Leia o arquivo `README.txt`.
- 2 Importe o arquivo `papagio.csv` para o R.
- 3 Analise o conjunto; entenda a natureza das variáveis.
- 4 Manipule o conjunto de dados; a variável `peso` deve ser numérica e a variável `especie` deve conter o nome das espécies de acordo com o arquivo `README.txt`. Para esta tarefa, utilize necessariamente expressões regulares.
- 5 Para utilizar o modelo KNN é preciso transformar as variáveis? Se sim, realize estas modificações.
- 6 Construindo o modelo. Para estimar o valor de K , construa uma validação cruzada no treinamento (cross validation) utilizando 5 pastas de igual tamanho. Pode utilizar neste momento a função pronta `knn` do pacote `class` para avaliar a validação cruzada que você contruiu.
- 7 Esboce o gráfico de linha x versus y em que $x = k$ (parâmetro do KNN) e y é a taxa de erro. Comente os resultados obtidos. Anexe este gráfico no Teams.
- 8 Após a escolha de k , avalie a acurácia do modelo. Comente os resultados obtidos. Anexe todo o arquivo `.R` da prova no Teams.