

Licenciatura em Engenharia Informática – DEI/ISEP

Análise de Dados em Informática

Ficha Teórico-Prática 9

Classificação: SVM

Objetivos:

- Modelos de SVM, usando Python;
- Análise e discussão dos resultados.

1. Considere o dataset “*Iris*” que contém 50 amostras de cada uma das três espécies ***Iris setosa***, ***Iris virginica*** e ***Iris versicolor***. Foram medidos 4 atributos de cada amostra: o comprimento e a largura das sépalas e pétalas, em centímetros.
 - a. Comece por carregar o ficheiro (“*Iris.csv*”), verifique a sua dimensão e obtenha um sumário dos dados.
 - b. Usando os gráficos apropriados explore os vários atributos do conjunto de dados.
 - c. Separe o conjunto de dados em dois subconjuntos treino e teste, segundo o critério *holdout*, (90% treino/10% teste).
 - d. Aplique o algoritmo SVM com os kernels “**linear**”, “**rbf**” e “**poly**”. Comente os resultados.
 - e. Avalie os modelos para diferentes parâmetros usando **k-fold cross validation**. Verifique se a diferença de desempenho entre os modelos obtidos anteriormente é estatisticamente significativa.