



Universidade do Minho
Departamento de Informática
Mestrado [Integrado] em Engenharia Informática

Dados e Aprendizagem Automática
1º Ano, 1º Semestre
Ano letivo 2021/2022

Enunciado Prático nº 5
11 de novembro de 2021

Tema	Máquina de Vector de Suporte e Árvores de Decisão
Enunciado	<p>A máquina de vectores de suporte é um modelo linear para problemas de classificação e regressão, onde o algoritmo cria hiper planos que separam os casos de estudo em classes. Ele pode resolver problemas lineares e não lineares, e funcionar bem para muitos problemas práticos. Já as árvores de decisão são um modelo supervisionado em que os dados são continuamente divididos de acordo com um determinado parâmetro, e onde as “folhas” representam as decisões ou resultados estimados.</p>
Tarefas	<p>Neste enunciado, utilizaremos o famoso <i>dataset</i> de flores de íris (informação disponível em https://rb.gy/zvaoiy), como forma de classificar a espécie da íris de uma flor, dado o comprimento e a largura das sépalas e pétalas, em centímetros. Para o desenvolvimento de um modelo de classificação, foi decidido aplicar-se os modelos máquina de vectores de suporte e árvores de decisão.</p> <p>Atendendo ao problema em questão, deverão seguir os seguintes passos:</p> <p>T1. Descarregar e carregar o respetivo <i>dataset</i>, utilizando a função <code>seaborn.load_dataset('iris')</code> da biblioteca <i>seaborn</i>;</p> <p>T2. Aplicar métodos para exploração e visualização de dados;</p> <p>T3. Preparar e organizar os conjuntos de casos de estudo do <i>dataset</i> em dados de treino e teste, utilizando a função <code>sklearn.model_selection.train_test_split(..., test_size = 0.3)</code>;</p> <p>T4. Treinar um modelo de máquina de vectores de suporte (<code>sklearn.svm.SVC</code>) e um modelo de árvore de decisão como classificador (<code>sklearn.tree.DecisionTreeClassifier</code>).</p> <p><i>Nota:</i> Definir o X e o y. Atenção ao tipo dos atributos que fazem parte do X;</p> <p>T5. Obter matrizes de confusão e relatório de classificação dos modelos e efectuar a respectiva análise crítica. Avaliar a <i>accuracy</i> do modelo na previsão de ‘<i>species</i>’. Avaliar também o modelo usando a métrica <i>f1_macro</i>. Que conclusões se poderão tirar?</p> <p>T6. Aplicar a técnica de <i>gridsearchCV</i> (<code>sklearn.model_selection.GridSearchCV</code>) como forma de procurar o conjunto de hiper-parâmetros, capaz de otimizar a performance de classificação dos modelos de máquina de vectores de suporte (<i>C</i> e <i>gamma</i>) e árvore de decisão (<i>criterion</i> e <i>max_depth</i>). Qual a variação na performance do modelo subjacente a estas alterações?</p>