

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO	Mestrado Integrado em Eng. Informática e Computação Inteligência Artificial	2018/2019 (3º Ano) 2º Sem
DOCENTES: Luís Paulo Reis, Ana Paula Rocha		

Tema do Trabalho Nº2 (Versões C1-C3)

Aprendizagem Supervisionada

Resumo

Pretende-se neste trabalho usar aprendizagem supervisionada para aprender/prever valores para o conceito em análise. Devem ser implementados/aplicados diferentes algoritmos de aprendizagem, sendo comparado o seu desempenho. Deve ser dada ênfase à análise da qualidade da aprendizagem (erros obtidos, análise da matriz de confusão, *accuracy*, ...) e tempo médio despendido para obter a solução.

O programa a desenvolver deve usar, pelo menos, 3 algoritmos de aprendizagem supervisionada (C4.5, Redes Neurais, K-Nearest Neighbor, Support Vector Machines,...) tendo por base um conjunto de dados disponibilizado para o efeito.

A aprendizagem supervisionada inclui a realização dos seguintes procedimentos. Análise do conjunto de dados de forma a verificar a eventual necessidade de pré-processamento, identificação do(s) conceito(s) a aprender (variável(is) dependente(s)) e definição do conjunto treino e conjunto de teste. Parametrização do modelo/algoritmo de aprendizagem. Avaliação da aprendizagem obtida, por medição dos resultados nos conjuntos de treino e teste.

A aplicação a desenvolver deve ter uma visualização adequada em modo de texto ou gráfico (embora isto não seja o essencial do trabalho) para mostrar a evolução da aprendizagem e permitir a interação com o utilizador na fase de teste de novos exemplos. Deve comparar os resultados obtidos com os diferentes algoritmos de aprendizagem.

Linguagem de Programação

Pode ser utilizada qualquer linguagem de programação e sistema de desenvolvimento, incluindo, a nível de linguagens, entre outras: Python, C++, Java, C#. Podem ser utilizadas bibliotecas ou ferramentas específicas para a implementação das experiências tais como o Rapid Miner, Weka ou R, entre outras. A escolha da linguagem e do ambiente de desenvolvimento a utilizar é da inteira responsabilidade dos estudantes.

Constituição dos Grupos

Os grupos devem ser compostos por 2 ou 3 estudantes. Não se aceitam grupos individuais ou compostos por 4 estudantes. Os grupos podem ser compostos por estudantes de turmas diferentes (no máximo de duas turmas diferentes) mas todos os estudantes têm de estar presentes nas sessões de acompanhamento e apresentação/demonstração do trabalho. De notar que embora a constituição de grupos compostos por estudantes de diferentes turmas seja permitida, não é aconselhada, dadas as dificuldades logísticas de realização de trabalho que pode provocar aos estudantes.

Entrega Final

Cada grupo deve submeter no Moodle dois ficheiros: um artigo em formato IEEE (relatório do trabalho) em formato PDF seguindo o modelo a disponibilizar no Moodle da disciplina, e o código implementado, devidamente comentado, incluindo um ficheiro “readme.txt” com instruções sobre como o compilar, executar e utilizar. Os estudantes devem também fazer uma demonstração da aplicação (cerca 10 min) do trabalho, na respetiva aula prática, ou em outro período a designar pelos docentes da disciplina.

Estrutura do Artigo/Relatório

Estrutura a divulgar posteriormente.

Tópicos Sugeridos

Os domínios sugeridos para aplicação de aprendizagem supervisionada são:

C1) Popularidade de notícias em plataformas de mídia social

(<http://archive.ics.uci.edu/ml/datasets/News+Popularity+in+Multiple+Social+Media+Platforms>)

C2) Análise de Drogas

(<http://archive.ics.uci.edu/ml/datasets/Drug+Review+Dataset+%28Drugs.com%29>)

C3) Reconhecimento de atividade em pessoas idosas

(<http://archive.ics.uci.edu/ml/datasets/Activity+recognition+with+healthy+older+people+using+a+batteryless+wearable+sensor>)

Estes endereços incluem o conjunto de dados para realização (treino/teste) da aprendizagem supervisionada.