

## Aprendizagem Automática II

### Trabalho prático – ano letivo 2020/2021

---

O trabalho prático a realizar nesta UC terá duas opções distintas, descritas abaixo.

#### Opção 1:

Escolha de um projeto que envolve análise de dados num dado contexto de aplicação, usando ferramentas Python já desenvolvidas aqui na Universidade do Minho em projetos de investigação, e podendo contribuir para as estender e melhorar, contando com apoio de um tutor especialista na área. A lista de propostas para esta componente é dada em anexo.

#### Opção 2:

Neste caso, o grupo deverá começar por escolher um conjunto de dados que irá ser usado na análise de dados. Os dados poderão ser retirados de algum tipo de publicação (e.g. artigo científico) devendo ter uma dimensão razoável, preferivelmente acima dos 5000 exemplos. Não se recomenda a escolha de conjuntos de dados de competições onde existam já muitas soluções disponíveis por parte da comunidade, mas participar em competições a decorrer com soluções inovadoras é uma opção.

A análise dos dados deverá ser adequada às características dos dados escolhidos e ao seu domínio de aplicação, sendo indicadas abaixo um conjunto de tarefas recomendadas (algumas poderão não fazer sentido dependendo dos dados escolhidos e outras não mencionadas poderão ser igualmente consideradas relevantes):

- carregamento e preparação dos dados (incluindo se necessário a geração dos atributos);
- exploração inicial, incluindo sumarização, visualização e pré-processamento;
- aprendizagem não supervisionada (redução de dimensionalidade, clustering);
- seleção de atributos;
- aprendizagem supervisionada (usando métodos “tradicionais”) incluindo a otimização dos seus hiper-parâmetros. Poderão ser usados e avaliados (de forma robusta) diversos métodos de Aprendizagem Máquina mais tradicional, de acordo com as características dos dados;
- metodologias de *deep learning*, comparando os resultados com os métodos anteriores, devendo ser exploradas alternativas em termos das configurações destes métodos, podendo ser testadas também classes distintas de modelos (e.g. *feedforward*, recorrentes, convolucionais).

Deverá ser criado um (ou vários) *Jupyter Notebooks* contendo os scripts usados e os resultados obtidos, bem como incluindo a explicação das razões para cada um dos métodos e discutindo-se os resultados obtidos. Estes podem usar ficheiros de código adicional desenvolvido pelo grupo (módulos incluindo funções ou classes). Todo o código desenvolvido por terceiros que seja usado deve ser devidamente identificado.

### **Datas e condições de entrega:**

Em qualquer das opções, o grupo deve colocar o código desenvolvido num repositório público (e.g. github), onde esteja também a documentação e os recursos necessários, devendo na página de entrada identificar claramente o conteúdo das várias pastas. Este repositório deverá ser continuamente atualizado com os resultados do trabalho e será avaliado numa fase intermédia e no final.

No final, o grupo deve escrever um relatório sucinto do trabalho desenvolvido (máximo de 10 páginas), que será complementado pelos ficheiros relevantes no repositório (código, notebooks, tabelas de resultados, etc).

### **Prazos e condições de entrega:**

- Escolha do tema para o trabalho – submissão no e-learning até dia **26 de março**.
  - No caso da opção 1, deverão apenas indicar entre duas e quatro opções dentro da lista apresentada (por ordem de preferência);
  - No caso da opção 2, deverão indicar claramente os dados a analisar e incluir um plano de trabalhos preliminar com os objetivos a atingir (máximo 1 página)
- Indicação do URL do repositório: **2 de maio**, sendo a avaliação inicial realizada entre 3 e 5 de maio;
- Apresentação à turma: **21 maio** (10 minutos por grupo) – objetivos detalhados; metodologias a usar/ plano de trabalhos; resultados preliminares já obtidos;
- Prazo final para atualização final dos repositórios: **6 junho**, sendo a avaliação final realizada entre 7 e 9 de junho; os grupos poderão ser convocados para uma defesa do trabalho nesta fase.

### **Avaliação:**

Peso na avaliação:

- Avaliação inicial – 30%
- Apresentação – 30%
- Avaliação final (incluindo defesa se necessário) – 40%

Notem que a avaliação é individual podendo, em casos justificados, haver diferenças de notas entre os vários elementos de um mesmo grupo