



DEIS - Departamento de Engenharia Informática e Sistemas  
ISEC - Instituto Superior de Engenharia de Coimbra

Conhecimento e Raciocínio 2017/2018

### Trabalho Prático

---

Para a realização do Trabalho Prático propõem-se 3 temas. Mais abaixo encontra-se a descrição detalhada de cada um deles e no Moodle será disponibilizado o material complementar de cada um.

No Moodle encontra-se um referendo para que possam escolher o tema que pretendem desenvolver. Apenas um dos alunos de cada grupo de trabalho deve selecionar o tema pretendido.

Os grupos de trabalho devem ser formados, no máximo, por **2 alunos**.

Existem duas datas possíveis para entrega, uma para a época normal e outra para a época de recurso. **Cada grupo apenas pode efetuar a entrega numa delas. Não serão aceites submissões nas duas datas.**

- Época Normal: 23.59 do dia 24 de Junho de 2018
- Época de Recurso: 23.59 do dia 08 de Julho de 2018

Devem ser entregues os seguintes elementos:

- Moodle: código e todos os ficheiros necessários para a execução e teste do trabalho; cópia pdf do relatório;
- No início da defesa deve ser entregue uma versão impressa do relatório.

Cada grupo realizará uma apresentação, com defesa, do seu trabalho. A defesa do trabalho é obrigatória.

**Datas para apresentação/defesa:** serão efetuadas defesas imediatamente após cada época de entregas. O mapa será divulgado pelos professores dos Laboratórios

O trabalho prático terá a cotação máxima de 10 valores (numa escala de 0 a 20).

## TEMA 1 – REDES NEURONAIS

Neste tema pretende-se que os estudantes aprofundem os seus conhecimentos sobre redes neuronais. O objetivo consiste na implementação de uma rede neuronal capaz de classificar corretamente um conjunto de folhas pertencentes a cerca de 35 espécies diferentes, que depois se subdividem em diferentes subespécies. Por exemplo, da **espécie Quercus** existem várias subespécies (*Kewensis*, *Nigra*, *Crassifolia*, etc)

No Moodle são fornecidos os ficheiros de imagens a preto e branco de 1564 imagens de diferentes folhas e no ficheiro excel encontra-se a classificação de cada uma delas no formato **espécie\_subespécie**. O **id** da folha corresponde ao nome do ficheiro *jpg* e existem alguns *ids* no excel que não se encontram na pasta das imagens e devem, portanto, ser ignorados.

Para este trabalho sugere-se a seguinte abordagem:

- a) [25%] Usando as funções de manipulação de imagem do Matlab converta as imagens fornecidas em matrizes binárias. Se achar necessário faça um tratamento prévio às imagens, como redimensionamento, ou outro que achar relevante.
  - Comece por uma rede neuronal de uma camada com 10 neurónios. Use a rede para treinar as folhas que se encontram na pasta **Folhas\_1**. Nesta pasta encontram-se várias folhas de diferentes espécies/subespécies. Use todos os exemplos no treino.
    - Faça uma primeira RN que faça a classificação por espécie.
    - Complemente a sua implementação de forma a fazer também a classificação por subespécie.
  - Teste várias topologias, funções de ativação e de treino, registre e compare os resultados obtidos. Registe conclusões.
- b) [20%] Repita os passos da alínea a) para o conjunto de imagens na pasta **Folhas\_2**. Utilize uma segmentação do dataset de 70%, 15%, 15% para treino, validação e teste. Observe a matriz de confusão, erros de treino e teste. Explore e compare várias configurações da rede. Obtenha a melhor, registre os resultados. Grave a(s) rede(s) neuronal(ais) com melhor desempenho.
- c) [15%] Utilize agora as imagens da pasta **Folhas\_3** que não foram usadas no treino anterior. Usando a(s) rede(s) treinada(s) e gravada(s) previamente verifique se a classificação dada pela(s) RN(s) é correta. Apresente os resultados obtidos. Posteriormente, repita o treino com estes novos exemplos, compare e registre os resultados obtidos.
- d) [10%] Desenhe manualmente algumas folhas que apresentem semelhanças com os exemplos usados no treino da rede. Transcreva os desenhos para matrizes binárias. Desenvolva um pequeno programa para ler um ficheiro correspondente a uma destas imagens e aplicá-lo à(s) rede(s) obtida(s) em c). Quais os resultados?
- e) [10%] Faça a extração de características tais como margens das folhas, texturas ou outras que considerar relevantes e use-as para otimizar o desempenho da RN. A extração de características poderá ser feita usando as funções de processamento de imagem do Matlab.
- f) [20%] Desenvolva uma aplicação gráfica em Matlab que permita ao utilizador fazer as tarefas desenvolvidas anteriormente de forma fácil e intuitiva:
  - Configurar a topologia das redes neuronais
  - Escolher funções de treino / ativação
  - Treinar as redes neuronais
  - Gravar uma rede neuronal previamente treinada
  - Carregar uma rede neuronal previamente treinada e aplica-la a um *dataset*
  - Desenhar uma nova folha, ou carregar um ficheiro de imagem onde esta já se encontre desenhada. Aplicar uma rede neuronal para classificar a folha
  - Visualizar os resultados da classificação.
  - Geração/gravação de ficheiros de resultados se achar relevante e necessário.
- g) Elabore um relatório do trabalho realizado e **imprima-o**. Uma má qualidade do relatório pode **descontar até 50%** na classificação total obtida nos pontos anteriores.

## TEMA 2 – SISTEMAS PERICIAIS

O objetivo deste tema é desenvolver um sistema pericial baseado em regras capaz de diagnosticar um Carcinoma Hepatocelular (CHC) num paciente, e indicar o melhor tratamento a aplicar em função do estágio em que a doença se encontra. Um Carcinoma Hepatocelular é um tipo de tumor maligno do fígado, sendo o mais frequente em adultos.

Para que exista coerência na comunidade médica no que toca às estratégias a seguir para esta doença, a *European Association for the Study of the Liver* em conjunto com a *European Organisation for Research and Treatment of Cancer* publicaram em 2012, no *Journal of Hepatology*, um conjunto de *guidelines* clínicas que definem estas estratégias. Este artigo encontra-se em anexo ao enunciado (EASL-EORTC-HCC-guidelines.pdf), e é nele que está a informação necessária para o desenvolvimento do sistema pericial. Desta forma, é recomendada uma leitura atenta à totalidade do artigo.

O sistema deverá permitir a inserção de factos pelos utilizadores finais, que serão os médicos, e, com base nos sintomas e nos resultados de exames realizados a um determinado paciente, irá indicar se ele tem ou não a doença CHC e, caso tenha, irá indicar também qual o estágio da doença e o tratamento mais indicado a aplicar. Desta forma, considere as seguintes alíneas:

- a) [25%] Implemente as regras que permitam ao sistema indicar se um determinado paciente tem a doença CHC ou se o diagnóstico é inconclusivo. Para esta componente tenha em particular atenção a informação apresentada no capítulo “Diagnosis” e a Figura 2 do artigo.
- b) [35%] Implemente as regras que permitam ao sistema indicar, para um paciente previamente diagnóstico com CHC, qual o estágio atual da doença e qual o tratamento mais correto a aplicar. Para esta componente tenha em atenção a informação apresentada no capítulo “Staging systems” do artigo, em particular o subcapítulo “BCLC classification: outcome prediction and treatment allocation” e a Figura 3.
- c) [25%] Elabore um conjunto de testes que permita testar exaustivamente o funcionamento do sistema pericial desenvolvido. Garanta que os testes cobrem todos os cenários possíveis de diagnóstico, estágios e tratamentos. Elabore um relatório detalhado com a análise feita aos testes executados e inclua essa informação no relatório final.
- d) [15%] Complemente o sistema desenvolvido com uma interface gráfica que permita a inserção dos dados necessários e a apresentação dos resultados gerados. Pode utilizar uma qualquer *framework* de Java para este fim, como a *Swing* ou a *JavaFX*.
- e) Elabore um relatório do trabalho realizado que, para além dos detalhes técnicos da implementação e da análise feita aos testes realizados com o sistema, deve incluir uma árvore de inferência ou um diagrama que permita uma simples leitura e compreensão das regras que o sistema implementa. Pode ilustrar o diagrama com imagens que considere convenientes. Um relatório com fraca qualidade pode levar a **descontos até 50%** na classificação total obtida nos pontos anteriores.

**Nota:** O desenvolvimento de um sistema deste tipo envolve que se tenha algum conhecimento do domínio em análise. Desta forma, é normal que, pontualmente, existam termos no artigo que o aluno não conhece e que não são explicados. Uma rápida pesquisa pela Internet irá esclarecer o aluno em relação a essas questões.

### TEMA 3 – CASE-BASED REASONING + LÓGICA DIFUSA

Neste tema pretende explorar-se o paradigma CBR e a lógica difusa, combinando-os numa mesma aplicação do ramo imobiliário, destinada a estimar o preço de venda de imóveis.

É fornecido o dataset *Melbourne\_HousePricing\_Samples* (MHS) contendo 2059 registos extraídos do dataset original *Melbourne\_HousePricing\_FULL*, também fornecido, com 34857 registos de casas australianas. Os preços dizem respeito a imóveis compostos apenas por habitação, apenas por terreno ou ambas.

Todos os records do dataset MHS contêm valores conhecidos de todos os seus atributos. Os imóveis distribuem-se pelas seguintes *Council Areas* da seguinte forma, e os preços médios por m<sup>2</sup> de habitação e de terreno foram calculados para cada *Council Area*:

Número de Habitações	CouncilArea	Preço por m2	Preço por m2 do terreno
93	Port Phillip City Council	10944.2	3599.8
60	Stonnington City Council	10906.6	3630.2
172	Boroondara City Council	10094.5	3346.1
67	Yarra City Council	9871.2	2004.2
111	Melbourne City Council	9779.0	5050.3
62	Monash City Council	8911.2	2433.7
66	Bayside City Council	8904.0	2815.9
119	Glen Eira City Council	7882.5	2906.8
162	Darebin City Council	7582.3	2667.4
152	Moreland City Council	7334.3	2484.6
30	Whitehorse City Council	7161.6	1916.6
130	Moonee Valley City Council	7057.8	2547.2
129	Maribyrnong City Council	6914.4	2117.4
68	Hobsons Bay City Council	6588.9	2602.4
63	Manningham City Council	6561.9	1982.4
96	Banyule City Council	6344.5	1713.8
54	Kingston City Council	6048.3	2023.2
27	Maroondah City Council	5845.8	1297.6
13	Greater Dandenong City Council	5246.0	1194.7
21	Knox City Council	5141.5	1248.3
105	Brimbank City Council	4578.8	1264.2
9	Casey City Council	4249.8	1082.1
26	Frankston City Council	4139.5	1035.8
10	Nillumbik Shire Council	4133.2	1168.1
89	Hume City Council	4018.7	1070
5	Macedon Ranges Shire Council	3993.3	58.4
51	Whittlesea City Council	3725.4	1263.9
43	Wyndham City Council	3635.7	1038.2
26	Melton City Council	2648.7	785.7

- a) [10%] Biblioteca de Casos: Utilize o dataset MHS como uma biblioteca de casos de um sistema CBR. Nesta biblioteca os atributos longitude e latitude serão fuzzificados da seguinte forma:
1. Calcular a distância de cada imóvel ao centro de cada Council Area;
  2. O grau de pertença de cada imóvel a cada Council Area é dado por  $1 - (\text{distância normalizada})$  ao centro da Council Area. Pode obter as coordenadas destes centros através do Google Maps, por exemplo.
- b) [15%] **Retrieve**: Desenvolva uma form para introdução de dados relativos a um novo imóvel. Implemente a fase Retrieve do sistema CBR com base em distância global linear. A definição das funções de distância local entre atributos fica ao seu critério, podendo para o efeito utilizar cálculos, tabelas ou outro qualquer método devidamente justificado. Não esquecer que:

1. alguns atributos são parte da descrição do caso e outros atributos constituem a sua solução. Apenas os atributos descritores entrarão na fórmula de cálculo da semelhança entre casos;
  2. A ponderação de cada atributo fica ao seu critério, devendo para o efeito realizar testes com base em casos selecionados do dataset original *Melbourne\_HousePricing\_FULL*. No entanto, um breve estudo preliminar sugeriu que os atributos Type, CouncilArea, YearBuilt e Bedroom2 têm especial importância. O sistema deve mostrar os 10 imóveis mais semelhantes ao atual, ordenados por grau decrescente de semelhança.
- c) [10%] **Reuse 1:** terminado o Retrieve, se as localizações do imóvel do novo caso e do caso passado a ele mais semelhante forem “suficientemente parecidas”, a avaliação do novo imóvel será a mesma da do caso passado mais semelhante. Para implementar o conceito de “suficientemente parecidas” baseie-se nos graus de pertença de cada caso às diversas *CouncilAreas*, conforme implementado em a). Para fazer o *tunning* desta função de semelhança deverá também realizar testes com casos novos obtidos a partir do dataset original e avaliar os resultados.
- d) [50%] **Reuse 2:** caso contrário, se as localizações do imóvel do novo caso e do caso passado a ele mais semelhante forem “diferentes”, então o preço do caso mais semelhante deverá ser adaptado da seguinte forma:
1. Calcular o preço/m<sup>2</sup> de habitação a aplicar ao novo caso, com base na sua localização:
    - i. A partir da tabela dada, definir  $k$  termos linguísticos  $TL_i$  ( com  $i \in [1...k]$  ) correspondentes a  $M$  classes de preços/m<sup>2</sup>. O valor de  $k$  fica ao seu critério e em função dos testes de validação que realizar;
    - ii. Criar  $N$  regras de inferência difusa do tipo “se imóvel pertence a  $[CouncilArea]$  => o preço/m<sup>2</sup> é  $TL_i$  ;
    - iii. Com base nos graus de pertença do novo caso a cada *CouncilArea*, disparar as regras e calcular o preço/m<sup>2</sup> para habitação
  2. Repetir os passos i)...iii) para calcular o preço/m<sup>2</sup> para terrenos
  3. Calcular uma primeira estimativa do preço do imóvel do novo caso com base em  $P_1$ ,  $P_2$ , área do imóvel e área do terreno;
  4. Calcular o preço final do imóvel do novo caso como uma média ponderada entre o valor calculado em 3. e o valor obtido a partir do Retrieve, em b) (ou seja, o preço do imóvel mais semelhante ao novo caso). Os fatores de ponderação destes valores serão empíricos e devem ser testados utilizando para o efeito casos obtidos a partir do dataset original, *Melbourne\_HousePricing\_FULL*
- e) [15%] **Revise e Retain:** a implementar de forma “standard” de acordo com os princípios gerais do paradigma CBR. Deve tomar em linha de conta que:
1. Os atributos do novo caso podem não ter sido todos fornecidos no preenchimento inicial da form referida em b)
  2. O preço de avaliação proposto pelo sistema e resultante de c) pode ser alterado pelo utilizador
  3. Os novos casos devem ficar todos registados numa tabela “histórica”
  4. A passagem de novos casos desta tabela histórica para a biblioteca de casos CBR deve ser feita sob supervisão de um administrador, devendo o sistema assinalar quais os casos prováveis candidatos a retenção.
- f) Relatório: Elabore um relatório do trabalho realizado e imprima-o. Uma má qualidade do relatório pode **descontar até 50%** na classificação total obtida nos pontos anteriores.