



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

## Conhecimento e Raciocínio 2021/2022

### Trabalho Prático

---

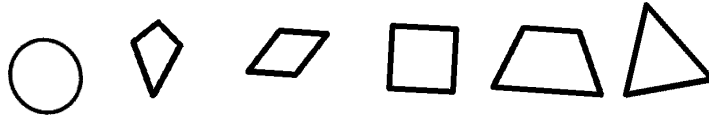
Para a realização do Trabalho Prático propõem-se 3 temas. Mais abaixo encontra-se a descrição detalhada de cada um deles e no Moodle será disponibilizado o material complementar necessário.

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 são de 2 alunos;
- Apenas um dos alunos do grupo deve fazer a escolha do tema pretendido no referendo disponibilizado no Moodle.
- A **data única de entrega** do trabalho é até às 23.59 do dia **18 de junho de 2022**;
- Devem ser entregues no Moodle o código e todos os ficheiros necessários para a execução e teste do trabalho, bem como o **PDF do relatório**;
- As defesas serão nos dias **20, 21, 22, 23 de junho**. Para a defesa, cada grupo terá de fazer a inscrição de 1 (**só 1**) dos seus elementos no Moodle, nos *slots* que para isso serão oportunamente disponibilizados;
- As defesas/dúvidas do tema Neural Networks serão com a Prof. Anabela Simões, do tema de Sistemas Periciais com a Prof. Inês Domingues e do tema CBR e Inferência Difusa com a docente da turma prática que frequentam (Anabela Simões/Inês Domingues);
- A defesa do trabalho é obrigatória e com a presença de todos os membros do grupo;
- O trabalho prático tem a cotação 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 e teste de diferentes arquiteturas de redes neuronais *feedforward* para classificar corretamente 6 formas geométricas (círculo, papagaio, paralelograma, quadrado, trapézio e triângulo):



No Moodle são fornecidos os ficheiros de imagens a preto e branco separadas por três pastas diferentes que devem ser usadas nas tarefas descritas de seguida.

**NOTA:** As imagens encontram-se no tamanho 224 x 224 pixéis, que em alguns computadores poderá levar a tempos e velocidades de treino muito elevadas. Caso seja necessário redimensioná-las, use as funções da toolbox de *image processing* do Matlab. Explique no relatório todo o pré-processamento feito às imagens. A imagem nunca deverá ter um tamanho inferior a 25x25.

Para este trabalho sugere-se a seguinte abordagem:

a) [10%]. 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 o reconhecimento dos caracteres da pasta “start”. Nesta pasta encontram-se **5 imagens** de cada forma geométrica.
- Use todos os exemplos que estão na pasta “start” no treino.
- Teste outras arquiteturas (topologias), funções de ativação e de treino. Registe os valores de desempenho (*accuracy*) das diferentes parametrizações e compare os resultados obtidos.

b) [30%]. Usando o modelo base implementado na alínea a) faça as alterações necessárias para implementar e testar várias topologias e parametrizações de RN de forma a obter um bom desempenho para a classificação dos caracteres fornecidos na pasta “train”. Nesta pasta encontram-se **50 imagens** de cada forma geométrica.

- Comece por usar uma segmentação do dataset de 70%, 15%, 15% para treino, validação e teste.
- Observe a matriz de confusão.
- Registe as precisões (*accuracy*) total e de teste.
- Explore e compare várias arquiteturas da rede (número de camadas/nº de neurónios).
- Teste diferentes funções de treino/ativação, diferentes segmentações na divisão dos exemplos. Registe os resultados para as várias redes neuronais que testar. Sugere-se a adaptação do ficheiro Excel dado nas aulas práticas, para registar resultados e obter as conclusões.
- Grave a(s) rede(s) neuronal(ais) com melhor(es) desempenho(s).

c) [25%]. Utilize agora as imagens da pasta **“test”**. Nesta pasta encontram-se 10 imagens de cada forma, imagens que não foram usadas no treino anterior.

Para esta tarefa use a melhor rede obtida em b).

- Sem re-treinar a rede verifique se a classificação dada pela RN é correta. Apresente os resultados obtidos.
- Agora volte a treinar a rede só com os exemplos da pasta **“test”**.
  - Agora teste a rede treinada anteriormente, separadamente para as imagens da **start, train e test**. Compare e registre os resultados obtidos em cada pasta.
- Volte a treinar a rede com todas as imagens fornecidas (**start + train + test**).
  - Agora teste a rede treinada anteriormente para as imagens de cada pasta em separado. Compare e registre os resultados obtidos.

d) [15%]. Desenhe manualmente algumas formas geométricas 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 melhor rede neuronal obtida em c). Quais os resultados?

e) [20%]. Desenvolva uma aplicação gráfica em Matlab que permita ao utilizador fazer as tarefas desenvolvidas anteriormente de forma fácil e intuitiva:

- Configurar a topologia da rede neuronal
- Escolher funções de treino / ativação
- Treinar a rede neuronal
- Gravar uma rede neuronal previamente treinada
- Carregar uma rede neuronal previamente treinada e aplicá-la a um *dataset*
- Desenhar uma nova forma, ou carregar um ficheiro de imagem onde esta já se encontre desenhada. Aplicar uma rede neuronal para classificar a forma geométrica desenhada.
- Visualizar os resultados da classificação
- Geração/gravação de ficheiros de resultados se achar relevante e necessário

f) Elabore um relatório do trabalho realizado. 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 para acompanhamento de pacientes suspeitos de cancro de próstata. Considere as seguintes indicações:

- Caso as suspeitas clínicas sejam baixas, o paciente deve ser observado anualmente (a data da próxima consulta deve ser indicada no ecrã ao médico de clínica geral). Caso as suspeitas clínicas sejam altas, o paciente deve ser referido à urologia.
- Caso a urologia decida fazer Ressonância magnética, e esta retorne um PI-RADS 1 a 3 ou não haja lesão focal, e o paciente seja considerado de baixo risco, este deverá ser observado de 6 em 6 meses (a data da próxima consulta deve ser indicada no ecrã ao médico de clínica geral).
- Caso a urologia decida fazer Ressonância magnética, e esta retorne um PI-RADS 1 a 3 ou não haja lesão focal, mas haja suspeitas clínicas ou bioquímicas, é feita uma biopsia guiada por ultrassom.
- Caso a urologia decida fazer Ressonância magnética, e esta retorne um PI-RADS 4 ou 5, é feita uma biopsia.
- Caso a urologia decida, em vez de uma Ressonância magnética, pode ser logo feita uma biopsia guiada por ultrassom.
- Caso a biopsia retorne negativa e não tenha sido ainda feita uma ressonância magnética deve-se considerar fazer uma ressonância magnética.
- Caso a biopsia retorne negativa e já tenha sido ainda feita uma ressonância magnética, deve ser feita uma biopsia guiada por ressonância magnética
- Caso a biopsia retorne positiva de baixo risco, deve ser feita uma vigilância ativa mensal (a data da próxima consulta deve ser indicada no ecrã ao médico de clínica geral)
- Caso a biopsia retorne positiva de médio ou alto risco, deve ser feita uma tomografia computadorizada.

Para este trabalho sugere-se a seguinte abordagem:

- a) [20%]. Implemente as regras que permitam ao médico de clínica geral identificar pacientes a ser encaminhados para a urologia.
- b) [30%]. Implemente as regras que permitam urologista decidir de que forma proceder.
- c) [10%] Considere a existência de lista de espera para a realização de ressonâncias magnéticas. O sistema deve indicar a data de realização da ressonância e atualizar possíveis datas posteriores dependentes desta.
- d) [20%]. Elabore um conjunto de testes que permita testar exaustivamente o funcionamento do sistema pericial desenvolvido. Garanta que os testes cobrem todos os cenários possíveis. Elabore um relatório detalhado com a análise feita aos testes executados e inclua essa informação no relatório final.
- e) [20%]. Complemente o sistema desenvolvido com uma interface gráfica que permita a inserção dos dados necessários e a apresentação dos resultados gerados.

f) 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. Uma má qualidade do relatório pode descontar até 50% na classificação total obtida nos pontos anteriores.

Nota: Recorde-se que deve utilizar um sistema pericial onde as regras não deverão possuir instruções do tipo “if”.

## TEMA 3 – CBR e INFERÊNCIA DIFUSA

É fornecido o dataset Melbourne\_HousePricing\_Samples (MHS) composto por 2059 registos de imóveis australianos na zona de Melbourne. Os preços dizem respeito a imóveis compostos apenas por habitação, apenas por terreno ou ambas, e distribuem-se pelas seguintes *Council Areas* conforme mostrado na tabela, que também apresenta os preços médios por m<sup>2</sup> da habitação e do terreno:

Número de Habitações	(Centro da) CouncilArea	Preço/m2 habitação	Preço/m2 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. Mas nesta biblioteca os atributos longitude e latitude serão substituídos 29 graus de pertença resultantes da fuzificação dos seus valores, feito da seguinte forma: 1) calcular a distância normalizada de cada imóvel ao centro de cada Council Area 2) O grau de pertença de cada imóvel a cada Council Area é dado por  $1-(distância\ normalizada)$  ao centro da Council Area. Pode obter as coordenadas dos centros das Council Areas no Google Maps.

**b) 25% Retrieve:** implemente a fase Retrieve. A definição das funções de distância local entre atributos fica ao seu critério, podendo para o efeito utilizar cálculos, tabelas ou outro método qualquer devidamente justificado. A ponderação de cada atributo fica ao seu critério. No entanto, um breve estudo preliminar sugeriu que os atributos Type, Localização (agora representada pelos 29 graus de pertença às diferentes Council Areas) e YearBuilt, têm especial importância.

**c) 10% Reuse (se situação 1):** 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.

**d) 40% Reuse (se situação 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 novo imóvel deverá ser inferido da seguinte forma:

1. Para a habitação:

i) Atendendo aos preços médios do  $m^2$  de habitação mostrados na tabela acima, definir  $n$  termos linguísticos  $TL_i$  (a seu critério. Exemplo: *baixo, médio...alto...*)

ii) Criar  $x$  regras de inferência difusa, a seu critério, do tipo “se o imóvel pertence a [CouncilArea] => o preço/ $m^2$  é  $TL_i$ ” (exemplo: *se o imóvel pertence a Port Phillip => o preço é muito alto*)

iii) Com base nos graus de pertença do novo caso a cada CouncilArea, disparar as regras e calcular o preço/ $m^2$  para o novo caso. O valor final obtém-se atendendo também à sua *BuildingArea*;

2. Repetir os passos i...iii) para calcular o preço/ $m^2$  para terrenos (neste caso usando o valor *LandSize*)

3. Calcular uma primeira estimativa do preço do imóvel do novo caso com base nos valores obtidos em 1 e 2. Poderá comparar o resultado com os imóveis mais semelhantes e, caso o resultado seja muito diferente, readaptar atendendo ao ano de construção, por exemplo. O que optar por aqui implementar com vista a uma adaptação mais perfeita, fica ao seu critério e imaginação.

**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 os novos casos devem ficar todos registados num “histórico” e que a passagem de novos casos deste histórico para a biblioteca de casos CBR deve ser feita numa opção de administração, sob supervisão de um administrador (backoffice).

**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.