

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

Conhecimento e Raciocínio 2016/2017

Trabalho Prático

Para a realização do Trabalho Prático propõem-se 4 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 26 de Junho de 2017
- Época de Recurso: 23.59 do dia 10 de Julho de 2017

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



0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0
1	1	0	0	0	1	1	1	1	1	0	0	0	0	0	0
1	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0
0	1	0	0	0	0	0	1	1	1	1	0	0	0	0	0
0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0
0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0
0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0
0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1

Excerto do ficheiro binaryalphadigs.jpg e o primeiro algarismo “2” digitalizado em 20x16

O objetivo deste trabalho é a implementação de uma rede neuronal capaz de reconhecer os algarismos 0 a 9 e os caracteres A a Z, maiúsculos, quando escritos à mão e digitalizados em apenas 2 níveis, branco e preto (0 e 1), numa matriz de 20x16 pixéis. Os ficheiros binaryalphadigs.jpg e binaryalphadigs são, respetivamente, as imagens exemplos destes caracteres escritos à mão, e as matrizes 20x16 correspondentes a cada um deles.

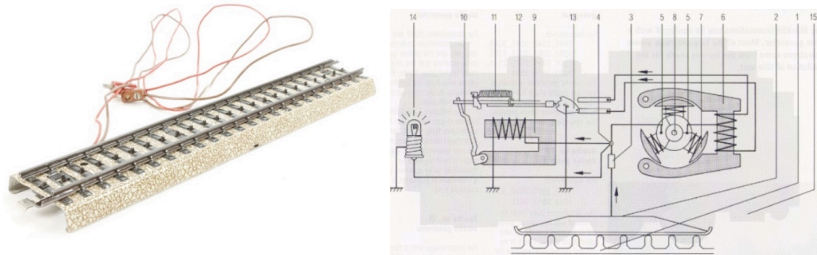
Para este trabalho a toolbox NNTOOL do Matlab pode ser usada como ferramenta auxiliar, mas as redes neurais devem ser implementadas de raiz.

Para este trabalho sugere-se a seguinte abordagem:

- 15% Comece por desenhar em matrizes de 5x7 pixéis os algarismos 0 a 9 (um exemplo de cada). Treine um rede neuronal capaz de reconhecer estes algarismos, testando-a com o dataset de treino.
- 15% Obtenha os coeficientes sinápticos de entrada, entre camadas e de polarização. Implemente um programa que simule a rede neuronal. Teste-a com as matrizes que criou em a)
- 20% A partir do ficheiro **binaryalphadigs.mat**, tome as matrizes correspondentes aos algarismos 0 a 9. Implemente e treine uma rede neuronal capaz de reconhecer estes algarismos. Utilize uma segmentação do dataset de 70%, 15%, 15% para treino, validação e teste. Observe a matriz de confusão, erros de treino e teste. Use várias configurações da rede. Obtenha a melhor.
- 30% Utilize agora o ficheiro **binaryalphadigs.mat** completo e repita c). Teste novas configurações da rede e optimize-a. Qual a melhor performance conseguida e com que modelo?
- 10% Obtenha os coeficientes sinápticos de entrada, entre camadas e de polarização para esse modelo. Implemente um programa que simule a rede neuronal. Teste-a com todo o ficheiro binaryalphadigs.mat
- 10% Desenhe manualmente alguns caracteres 0 a 9 e A a Z. Transcreva-os para uma matriz de 20x16. Desenvolva um pequeno programa para ler um ficheiro correspondente a um destes caracteres e aplicá-lo à rede obtida em e). Quais os resultados?
- 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 trabalho é criar um ES baseado em regras destinado ao diagnóstico de avarias em locomotivas analógicas escala H0 da Märklin. Trata-se de uma empresa alemã de brinquedos fundada em 1859, especialmente famosa pelos comboios elétricos (alguns modelos atingem hoje valores da ordem dos milhares de euros). O funcionamento é em corrente alternada (AC). Neste sistema os carris proporcionam a *massa* e um condutor central adicional, a *fase*. Assim, a tensão elétrica é aplicada às locomotivas pelas rodas e um “patim” que contacta com o condutor central dos carris.



Carris Marklin tipo “M” e esquema de uma locomotiva analógica (séries 3000 e 3100)

O documento anexo **Marklin_Service_Manual.pdf** é um manual de manutenção de locomotivas Märklin. No capítulo 9 – *Correcting Problems with Locomotives* - as secções 9.1 a 9.18 apresentam uma série de testes e procedimentos com vista ao diagnóstico e reparação de avarias. Pretende implementar-se um ES que substitua parte destas secções, proporcionando assim todo o suporte necessário aos modelistas e colecionadores. Para isso, sugere-se a seguinte abordagem:

- 20% Traduza e compreenda todo o texto das secções 9.1 a 9.5. Esquematize-o numa árvore de inferência ou num diagrama que torne simples a sua leitura e compreensão. Ilustre o diagrama com as imagens consideradas convenientes. Elabore um relatório contendo todos os resultados deste trabalho.
- (10% por cada secção implementada) Elabore um sistema pericial que implemente as tarefas de diagnóstico e indique os procedimentos corretivos correspondentes às secções 9.1 a 9.5. Este sistema deve permitir a entrada dos factos pelo utilizador.
- 20% Elabore uma bateria de testes composta por um conjunto de situações que permitam testar (quase) exaustivamente o funcionamento do sistema pericial. Teste todos estes casos. Elabore um relatório detalhado dos testes desenvolvidos e dos resultados obtidos.
- 10% Complemente o trabalho com um interface gráfico destinado a permitir a entrada de dados e a afixação do diagnóstico e dos procedimentos de manutenção. Alguns destes procedimentos serão mais perceptíveis se acompanhados de uma figura ilustrativa...
- 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.

Terminologia:

- Pickup shoe:** Patim da locomotiva, que deve tocar no condutor central da via (a *fase*)
- Layout:** O desenho do percurso na maquete, composto por carris
- Metal coupler:** O acoplador é a peça, feita de metal, que liga a locomotiva à 1ª carruagem
- Contact plate:** Placa de contacto, é uma pequena peça junto do motor, por onde a corrente eléctrica chega ao motor e às lâmpadas de iluminação. Está ligada ao *pickup shoe* por um fio.

TEMA 3 – CASE-BASED REASONING

Este trabalho tem como objetivo a implementação de 3 modelos CBR de áreas distintas e usando ferramentas diferentes: 1) MyCbr e identificação de padrões 2) MatLab e predição 3) Linguagem à escolha e diagnóstico

- 1) 20% Reporte-se aos ficheiros descritos no TEMA1 – REDE NEURONAL. Faça um pré-processamento adequado dos dados de modo a poder importá-los como base de casos para o MyCBR. Uma vez importados, configure as funções de semelhança e realize alguns testes com vista à avaliação do reconhecimento de padrões por este sistema. Analise os resultados. Faça um relatório do trabalho realizado.
- 2) Reporte-se aos ficheiros descritos a respeito do TEMA4 – REDE BAYESIANA.
 - a. 10% Leia 70% do ficheiro por.xls para o Matlab (records extraídos aleatoriamente), de modo a constituir uma Case Library de um sistema CBR. NOTA: use apenas 2 ou 4 classes (intervalos) para o valor de G3, pelas razões apontadas no TEMA4
 - b. 10% Em Matlab, implemente a fase Retrieve do paradigma CBR com base em distância linear, funções simétricas e com ponderação 1 para todos os atributos. Aplique os restantes 30% de exemplos ao sistema e verifique a taxa de acerto
 - c. 10% Modifique a função de distância para incluir as seguintes ponderações e repita os testes b). Compare resultados.

Ranked attributes:

0.52703	32	G2	0.04713	4	address
0.35642	31	G1	0.04557	24	famrel
0.30073	15	failures	0.04454	25	freetime
0.2338	21	higher	0.04159	29	health
0.11172	1	school	0.03857	12	guardian
0.08271	27	Dalc	0.03645	22	internet
0.07854	16	schoolsup	0.02707	2	sex
0.07198	7	Medu	0.02511	19	activities
0.06346	8	Fedu	0.02417	20	nursery
0.06117	14	studytime	0.02335	17	famsup
0.05823	18	paid	0.02129	23	romantic
0.05503	10	Fjob	0.01638	5	famsize
0.05412	9	Mjob	0.00996	6	Pstatus
0.05349	11	reason	0	30	absences
0.05029	13	traveltime	0	3	age
0.04979	26	goout			
0.04812	28	Walc			

- d. 20% Complete o ciclo CBR incluindo as restantes fases Revise e Retain. Um caso só deve ser retido se contiver algo de novo, uma lição...
- 3) 30% Reporte-se à documentação descrita a respeito do TEMA2-SISTEMA PERICIAL. Analise o texto das seções 9.1. e 9.2. A partir dele projete um sistema CBR destinado ao diagnóstico destas falhas. Pode programá-lo numa linguagem qualquer, à sua escolha.
- 4) 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 4 – REDE BAYESIANA

O objetivo deste trabalho é implementar uma rede bayesiana que permita prever o resultado final de um aluno nas disciplinas de português e de matemática em momentos distintos do ano letivo.

Os ficheiros **por.xls** e **mat.xls** representam alunos de português e de matemática cujos atributos se encontram descritos no ficheiro student.txt. De entre estes atributos devem destacar-se G1, G2 e G3 que representam, respetivamente, as classificações numa escala de 0 a 20 obtidas no final do 1º período, 2º período e final do ano. Existem 382 alunos comuns entre português e matemática, descritos por atributos iguais (exceto G1 a G3) nos ficheiros por.xls e mat.xls.

Para o ficheiro por.xls:

- a) 20% Feito um estudo preliminar através de uma ferramenta adequada, concluiu-se que os atributos mais relevantes para a previsão da classe G1 são failures, higher, school, Dalc e studytime. Implemente uma rede Bayesiana para a predição de G1. NOTA: alguns destes atributos, em especial G1, têm muitos valores diferentes, pelo que se torna penoso calcular todas as probabilidades condicionais envolvidas. Deve por isso proceder à discretização ou generalização prévia dos atributos, através do Excel, Matlab ou outra ferramenta à sua escolha. Para G1 poderá começar por usar, por exemplo, apenas 2 classes: abaixo de 10 valores e acima de 10. Tente também com mais algumas classes...
- b) 10% Teste a performance da rede. Crie um ficheiro com os casos selecionados para o teste; importe os casos para o Genie (File-> OpenDataFile); selecione-os e faça Data->Copy; preencha a correspondência entre nós da rede e atributos dos casos. De retorno ao interface da rede bayesiana, poderá agora aplicá-los um a um à rede e verificar os resultados. Anote-os. Elabore um relatório do trabalho realizado em a) e b)
- c) 10% Feito um estudo preliminar através de uma ferramenta adequada, concluiu-se que os atributos mais relevantes para a previsão da classe G2 são: G1, failures, Famrel, Fedu e Medu. Tomando em atenção a NOTA feita em a), modifique a rede desenhada em a) para incluir agora G2.
- d) 10% Teste a performance da rede seguindo um procedimento análogo ao descrito em b). Elabore um relatório do trabalho realizado em c) e d)
- e) 10% Feito um estudo preliminar através de uma ferramenta adequada, concluiu-se que os atributos mais relevantes para a previsão da classe G3 são G2, G1, failures, Dalc, Fedu e Medu. Tomando em atenção a NOTA feita em a), modifique a rede desenhada em c) para incluir agora G3.
- f) 10% Teste a performance da rede seguindo um procedimento análogo ao descrito em b). Elabore um relatório do trabalho realizado em e) e f).
- g) 10% Faça um estudo que permita concluir acerca da previsibilidade da evolução de alunos ao longo do ano, isto é, compare os resultados previstos para G1, G2 e G3 quando o mesmo aluno progride ao longo do ano: previsões na matrícula, depois do 1º período (G1 conhecido) e depois do 2º período (G2 conhecido).
- h) 10% Há 382 alunos comuns entre por.xls e mat.xls (atributos iguais). Identifique-os usando o Matlab ou outra ferramenta ao seu critério. Será a rede desenhada em e) aplicável para a previsão dos resultados de matemática? Teste utilizando pelo menos 20 exemplos.
- i) 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.