 <small>ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO</small>	Tipo de Prova Exame Teórico – Avaliação Final	Ano letivo 2016/2017	Data 22-06-2017
	Curso Licenciatura em Engenharia Informática	Hora 10:00	
	Unidade Curricular Inteligência Artificial	Duração 2:00 horas	

Observações:

- Pode trocar a ordem das questões, desde que as identifique convenientemente
- Qualquer tentativa de fraude implica a anulação do exame

1. (5V)


Considere os seguintes algoritmos/abordagens abordados na UC de Inteligência Artificial:

1. Classificação
2. Segmentação/Clustering
3. Regressão Linear
4. Associação
5. Normalização
6. Discretização
7. Raciocínio Baseado em Casos
8. Sistema Baseado em Regras

Faça corresponder, justificando, cada problema descrito de seguida com o algoritmo/abordagem que achar mais apropriado para a sua resolução.

Cada resposta deve seguir o formato Letra → Número : Justificação

- A. É-lhe fornecida uma base de dados de imagens médicas pré-processadas: cada pixel que poderá, eventualmente, representar uma mutação foi já marcado com uma cor diferente. Pretende-se que facilite o trabalho do médico que vai analisar as imagens, apontando as áreas de maior concentração de células com mutação.
- B. É-lhe fornecido um dataset descrevendo o preço de uma série de casas em função da área (em mm^2), do nº de quartos, do nº de telhas e do nº de pisos. Pretende-se que prepare este dataset de forma a melhorar a qualidade do modelo de previsão de preços que venha a ser desenvolvido no futuro.
- C. Pretende-se prever a média de primeiro ano de um caloiro da ESTG com base na sua média de acesso ao ensino superior e na sua assiduidade.
- D. É-lhe fornecido um dataset gerado num hospital, que faz corresponder os sintomas de cada um dos pacientes que deram entrada no Serviço de Urgências a um dos 5 níveis de prioridade da escala de Manchester (Azul, Verde, Amarelo, Laranja e Vermelho). Pretende-se que desenvolva um modelo que represente o processo de tomada de decisão dos profissionais do Serviço de Urgências durante a triagem.
- E. A nave colonizadora Covenant mantém, em papel, um conjunto de protocolos que determinam como agir em cada circunstância. No entanto, e como se veio a constatar, nem sempre os Humanos estão inclinados a seguir os protocolos. Pretende-se que implemente estes protocolos no computador de bordo das próximas colonizadoras, para impedir novos desastres no futuro.
- F. Numa prova automobilística com 20 participantes, pretende-se prever a posição em que cada piloto ficará com base em variáveis como os anos de experiência do condutor, o tipo de piso, a cilindrada do veículo ou os resultados passados.
- G. Um hospital pretende que desenvolva um sistema automatizado de diagnóstico que seja capaz de aprender através da observação do trabalho dos médicos que o vão utilizar. O sistema deverá ser capaz de registar, para cada paciente, quais os sintomas e a sua gravidade, bem como qual o

 ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO	Tipo de Prova Exame Teórico – Avaliação Final	Ano letivo 2016/2017	Data 22-06-2017
	Curso Licenciatura em Engenharia Informática	Hora 10:00	
	Unidade Curricular Inteligência Artificial	Duração 2:00 horas	

diagnóstico do médico. Espera-se que, com o tempo, a performance do sistema no diagnóstico médico vá melhorando.

- H. O Louvre tem um total de aproximadamente 38.000 objetos em exposição o que faz com que, com frequência, um visitante não seja capaz de os observar todos. Pretende-se determinar se há objetos que, independentemente da sua localização no museu, são frequentemente observados em conjunto pelos visitantes. O objetivo é colocar esses objetos mais próximos, de forma a melhorar a eficiência dos percursos dos visitantes e assim permitir-lhes observar mais obras.
- I. Entre 1980 e o presente, o preço do Gigabyte nos discos rígidos tem vindo a descer desde os \$1.000.000 até a um valor inferior a \$0.01. É-lhe fornecido acesso a um dataset contendo o preço do Gigabyte em cada ano entre 1980 e 2017. Pretende-se prever o preço do Gigabyte ao longo dos próximos anos.
- J. É-lhe fornecido um dataset que faz corresponder o aspeto de cavalos de corrida (caracterizado pela condição dos dentes, cascos, e postura) à sua idade em anos (arredondada à 2ª casa decimal). Pretende-se, no futuro, vir a classificar a idade do cavalo como sendo “jovem” ou “velho” no entanto é primeiro necessário, para isso, transformar o dataset. É-lhe pedido que faça esta tarefa.

A → 2 : As técnicas de clustering organizam elementos próximos, que partilham das mesmas características, em grupos. Revelar grupos contendo grandes concentrações de células mutadas facilitaria o trabalho do médico

B → 5 : A saber, o principal problema do dataset é a grande diferença entre as gamas de valores de algumas das suas variáveis: área e nº telhas terão valores eventualmente muito altos enquanto que nº de quartos e nº de pisos valores muito baixos. A normalização permite que as gamas de valores das variáveis sejam uniformes, impedindo que algumas variáveis “escondam” outras no momento da criação do modelo

C → 3 : Uma vez que se trata da previsão de um valor numérico contínuo (entre 0 e 20) a abordagem mais adequada seria a Regressão Linear, que permite construir um modelo linear de uma variável em função de outra(s).

D → 1 : Visto que o que se pretende é atribuir uma classe (um dos cinco níveis de prioridade da triagem de Manchester) a cada observação

E → 8 : Uma vez que os protocolos são já um conjunto de regras que dizem como atuar em cada circunstância, estes poderiam ser implementados num sistema baseado em regras (por exemplo prolog), que poderia ser facilmente consultado para determinar como decidir em cada situação.

F → 1 : Visto que o que se pretende é atribuir uma classe (um valor discreto entre 0 e 20), a abordagem mais adequada é a classificação

G → 7 : Uma vez que o que se vai guardar são casos concretos de diagnóstico e visto que se espera que o sistema vá aprendendo com a observação da realidade e melhorando, a abordagem mais adequada é o Raciocínio Baseado em Casos

H → 4 : Pretende-se determinar se há objetos que são observados em conjunto numa mesma visita, i.e., se quem observa A e B também observa C. A melhor abordagem para este problema, com semelhanças óbvias ao Market Basket Analysis, é a pesquisa de regras de associação

I → 3 : Uma vez que se trata da previsão de um valor numérico contínuo a abordagem mais adequada seria a Regressão Linear, que permite construir um modelo linear de uma variável em função de outra(s).

J → 6 : A variável idade é numérica e tem que ser transformada numa variável discreta (com dois valores

<p>P.PORTO</p> <p>ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO</p>	Tipo de Prova Exame Teórico – Avaliação Final	Ano letivo 2016/2017	Data 22-06-2017
	Curso Licenciatura em Engenharia Informática	Hora 10:00	
	Unidade Curricular Inteligência Artificial	Duração 2:00 horas	

possíveis), sendo para isso necessário utilizar um algoritmo de discretização

2. Considere o seguinte problema:

Pretende-se estudar a afluência de trânsito numa artéria de uma determinada cidade, bem como os fatores que o influenciam. O Objetivo é desenvolver um sistema capaz de prever, com base nesses fatores, a afluência de carros na artéria em cada momento específico de cada dia do ano.

2.1 Indique, justificando, qual o algoritmo ou técnica, de entre os abordados na aula, que utilizaria para implementar este sistema.

(1V)

O algoritmo mais adequado seria um algoritmo de classificação que poderia, por exemplo, classificar o trânsito como sendo "ligeiro", "médio" ou "intenso"

2.2 Admita agora que vai efetivamente construir um dataset para modelar o problema descrito. Indique, justificando, 5 atributos que o dataset deveria conter. Para cada um indique, ainda, porque processo poderiam os seus valores ser obtidos.

(2.5V)

1. Dia da semana, obtido de um calendário. É importante pois há dias da semana em que o trânsito é mais intenso (dias de trabalho vs. fim de semana)
2. Hora do dia, obtido de um relógio. É importante pois o trânsito varia ao longo do dia (e.g. mais intensidade em horas de ponta, em que as pessoas se dirigem para ou voltam do trabalho)
3. Data (dia/mês/ano), obtido de um calendário. É importante pois a mesma hora e o mesmo dia da semana podem ter intensidades de trânsito diferentes em diferentes alturas do ano (e.g. comparar Maio com Agosto)
4. Feriado, obtido de um calendário. Os feriados, em que há menos pessoas a dirigirem-se para o trabalho, influenciam de forma significativa o trânsito. É ainda importante devido à existência de feriados móveis, que podem calhar em datas diferentes de ano para ano.
5. Intensidade de trânsito, obtido a partir de câmaras de vigilância de trânsito. Conta, de forma aproximada, o nº de carros em cada momento na artéria a estudar e guardando um valor discretizado, para permitir a utilização de um algoritmo de classificação. Constitui a variável de classe do dataset

3. No contexto da framework Robocode, considere a existência dos seguintes factos:

- `posicao(X,Y, Inimigo)` – representa a última posição conhecida de um dado inimigo;
- `distancia(Inimigo, Dist)` – calcula a distância a que se encontra um dado inimigo;
- `eu(X,Y)` – representa a posição atual do meu robot;
- `accuracy(Inimigo, P)` – representa a eficácia do robot (em percentagem), contra um determinado inimigo;
- `disparo(Inimigo, Acertou)` – representa um disparo feito a um determinado inimigo e o resultado (hit/no);


3.1 Defina o predicado `naArea/2` que determina se um dado inimigo se encontra dentro do quadrado cujo centro é o meu robot e cujo lado é também dado como argumento. Mostre ainda como utilizaria o predicado.

(2V)

`naArea(I, L):-eu(X,Y), posicao(Xi,Yi,I), L1 is L/2, L1X is X-L1, L2X is X+L1, L1Y is Y-L1, L2Y is Y+L1, Xi > L1X, Xi < L2X, Yi > L1Y, Yi < L2Y.`

<p>P.PORTO</p> <p>ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO</p>	Tipo de Prova Exame Teórico – Avaliação Final	Ano letivo 2016/2017	Data 22-06-2017
	Curso Licenciatura em Engenharia Informática	Hora 10:00	
	Unidade Curricular Inteligência Artificial	Duração 2:00 horas	

- 3.2
(2V) Defina o predicado `inimigosPerto/2` que calcula a lista de inimigos que se encontram a uma distância inferior ao valor dado como argumento. Mostre ainda como utilizaria o predicado.
- ```
inimigosPerto(D, L):-findall(I, (posicao(, , I), distancia(I, D1), D1 < D), L).
```
- 3.3  
(1V) Defina o predicado `sobreposto/1` que determina se um determinado inimigo está na mesma posição que o meu robot.
- ```
sobreposto(I):-eu(X,Y), posicao(X,Y,I).
```
4.
(1.5V) Comente a seguinte afirmação, indicando ainda se concorda ou não com ela:
- Um dos problemas do Raciocínio Baseado em Casos é que se baseia em factos cuja relevância estatística é desconhecida.*
- Concordo com a afirmação. De facto, podem existir casos numa base de casos que não são, de todo, representativos do problema. Estes casos podem resultar, por exemplo, de uma combinação muito particular de acontecimentos e não voltarem a acontecer nunca mais. Seja como for, quando olhamos para dois casos numa base de casos, é impossível percebermos qual dos dois é mais representativo do problema, o que seria muito importante para suportar a tomada de decisão.
5.
(1V) Indique, fundamentando e descrevendo-a, uma característica que ache fundamental num sistema de Inteligência Artificial.
- Uma das características fundamentais num sistema de Inteligência Artificial é a capacidade de aprendizagem. De facto, um sistema considerado inteligente deve ser capaz de aprender, de adquirir e utilizar novos conhecimentos e, desta forma, adaptar-se às alterações no ambiente e resolver cada vez melhor os problemas com que se depara.
6.
(2.5V) A fase de preparação dos dados é uma das mais importantes no processo de Extração de Conhecimento. Identifique e explique brevemente 2 problemas que a preparação de dados permite resolver e 2 tarefas típicas na preparação de dados, não necessariamente relacionados.
- Dois problemas:
- 1) a existência de dados de fontes distintas que precisam de ser integrados. Isto acontece com frequência quando se tem mais que um dataset diferente a descrever o mesmo problema
 - 2) A falta de valores em alguns atributos. Isto acontece quando, por falha ou erro, por exemplo, os valores de determinados atributos não estão disponíveis
- Duas tarefas típicas na preparação de dados:
- 1) Discretização: permite converter uma variável contínua em uma variável discreta
 - 2) Normalização: permite que variáveis com intervalos de valores muito diferentes possam ser transformadas de forma a que tenham todas o mesmo intervalo de valores

 ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO	Tipo de Prova Exame Teórico – Avaliação Final	Ano letivo 2016/2017	Data 22-06-2017
	Curso Licenciatura em Engenharia Informática	Hora 10:00	
	Unidade Curricular Inteligência Artificial	Duração 2:00 horas	

7.
(1.5V)

Considere o dataset ilustrado ao lado, que representa a temperatura em 7 cidades ao longo do dia. Para este dataset indique qual o seu problema e que abordagem utilizaria para resolvê-lo. Justifique a sua resposta indicando qual o objetivo principal da abordagem seguida.

O problema deste dataset é a falta de alguns valores nos seus atributos

Para resolver o problema optaria por preencher cada valor em falta com o valor médio do seu atributo. Isto é indicado porque em cada atributo há muitos poucos valores em falta e seria feito com o objetivo de preencher os valores em falta sem alterar de forma significativa a distribuição dos valores dos atributos.

	C1	C2	C3	C4	C5	C6	C7
09:00	19	17	20	18	19	17	20
10:00	15	15	16	16	17	19	17
11:00	15	18	19	17	20	17	18
12:00	20	?	20	15	17	16	19
13:00	18	19	17	18	17	17	17
14:00	19	17	17	15	?	17	17
15:00	17	16	20	17	18	16	18
16:00	17	17	20	15	15	15	15
17:00	15	17	17	17	20	17	19
18:00	17	18	17	?	17	20	19
19:00	16	15	18	17	16	19	20
20:00	16	?	18	20	19	15	20
21:00	16	17	15	15	17	16	?
22:00	17	18	16	17	16	20	17
23:00	20	17	17	17	16	18	18
00:00	18	17	19	17	?	15	16