 <small>ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO</small>	Tipo de Prova Exame Teórico – Época Especial	Ano letivo 2016/2017	Data 13-07-2017
	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. Num exame de SBD de escolha múltipla, o professor pretende determinar se há grupos questões que os alunos tendem a errar ou a acertar mais. Por exemplo, os alunos que erram uma pergunta sobre álgebra relacional também erram uma pergunta sobre junções.

4. Pretende-se determinar se há grupos de variáveis que ocorrem em conjunto (acertar ou não cada pergunta), numa situação muito similar ao market basket analysis

- B. É-lhe fornecido um dataset que descreve uma série de consultas num serviço de clínica geral de um hospital privado. O dataset contém informação sobre o paciente, os seus sintomas, o diagnóstico do médico e os medicamentos prescritos. Pretende-se que desenvolva um sistema para administração automática de medicamentos com base nas consultas conhecidas.

7. Uma vez que o que se pretende é decidir sobre o diagnóstico de novos pacientes com base em casos passados


- C. Existe um dataset que faz corresponder informação socioeconómica de alunos da ESTG à sua nota na UC de Inteligência Artificial (0 a 20). No futuro pretende-se vir a desenvolver um classificador para determinar se um aluno vai passar ou não à UC com base nessa informação. No entanto, é-lhe pedido previamente que prepare o dataset para essa tarefa.

5. É necessário discretizar a nota dos alunos, que é um valor entre 0 e 20, para uma variável discreta com duas classes possíveis: passou ou reprovou.

- D. Numa loja de produtos eletrónicos é mantido um perfil dos clientes que inclui, entre outros dados, a morada, a lista de marcas favorita ou o nível económico (estimado através do valor de compras passadas). Pretende-se descobrir se há conjuntos de clientes com perfis similares para o envio de publicidade direcionada.

2. Uma vez que se pretende encontrar grupos de instâncias com características similares.

- E. Um restaurante pretende determinar se há pratos que são pedidos com maior frequência em conjunto pelos seus clientes (e.g. bacalhau na brasa e pudim abade de priscos), para elaborar

 ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO	Tipo de Prova Exame Teórico – Época Especial	Ano letivo 2016/2017	Data 13-07-2017
	Curso Licenciatura em Engenharia Informática	Hora 10:00	
	Unidade Curricular Inteligência Artificial	Duração 2:00 horas	

menus com um preço pré-definido e que possam potenciar as vendas.

4. Pretende-se determinar se há grupos de variáveis que ocorrem em conjunto (os pratos pedidos por cada mesa).

- F. É-lhe fornecido acesso a um dataset construído na cantina da ESTG que contém informação sobre a quantidade de carne, hidratos e peixe vendidas por dia (em gramas), a quantidade de água gasta na preparação da comida (em mililitros), e o nº de refeições vendidas tanto ao almoço como ao jantar (em dezenas). O objetivo, no futuro, é otimizar a gestão do stock. No entanto pretende-se que pré-processe este dataset de forma a que os algoritmos a utilizar no futuro cheguem a melhores resultados

5. Uma vez que as variáveis podem ter valores muito diferentes (na ordem das dezenas de milhar para a comida e bebida, e na ordem das dezenas para as refeições vendidas), é necessário normalizar as variáveis para que todas tenham a mesma importância no momento da utilização de algoritmos de IA

- G. O direito do consumidor é o ramo do direito que lida com as relações jurídicas entre fornecedores de bens e serviços e seus consumidores (e.g. uma loja de eletrónica tem que proceder à reparação de um determinado artigo ainda dentro da garantia num prazo limite ou proceder à sua devolução). Pretende-se que modele este ramo do direito para que possa utilizar-se como um juiz eletrónico.

8. Uma vez que o sistema se deve basear na legislação vigente e a legislação pode ser interpretada como normas ou regras que estabelecem as consequências de cada evento ou ação e podem, assim, ser traduzidas para regras que podem ser interpretadas por um sistema baseado em regras

- H. É-lhe fornecido acesso a um dataset que faz corresponder as características de veículos (e.g. peso, velocidade) e de diferentes pisos e suas condições (e.g. molhado/seco, estado de conservação) à distância de travagem (em metros). Pretende-se que desenvolva um sistema de travagem automática para veículos que seja capaz de prever a distância de travagem e assim acionar automaticamente o travão para evitar acidentes.


3. Uma vez que se pretende prever um valor numérico (a distância de travagem) em função de um conjunto de variáveis independentes.

- I. Na prática da escalada, um aspeto importante é determinar a distância da queda, de forma a minimizar os potenciais acidentes e as suas consequências. É-lhe fornecido um dataset que faz corresponder os pesos de escaladores, a elasticidade e diâmetro das suas cordas, a distância de corda utilizada e a altura do escalador à distância de queda. Pretende-se que desenvolva um sistema capaz de determinar a distância da queda em função destas variáveis.

3. Uma vez que se pretende prever um valor numérico (a distância de queda) em função de um conjunto de variáveis independentes.

- J. Um banco monitoriza todas as transações dos cartões de crédito dos seus clientes (descritas, entre outras, pelo seu montante, localização e hora do dia), assinalando ainda aquelas em que a transação constitui uma fraude (e.g. transação feita por um terceiro que foi capaz de clonar o cartão de crédito do cliente). É-lhe pedido que desenvolva um sistema de deteção de fraude em tempo real, para evitar este tipo de situações.

1. Uma vez que o que se pretende é atribuir um valor de entre dois possíveis (fraude ou não fraude) a instâncias

 <small>ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO</small>	Tipo de Prova Exame Teórico – Época Especial	Ano letivo 2016/2017	Data 13-07-2017
	Curso Licenciatura em Engenharia Informática	Hora 10:00	
	Unidade Curricular Inteligência Artificial	Duração 2:00 horas	

2. Considere o seguinte problema:

Pretende-se prever a afluência de pacientes nos serviços de emergência dos Hospitais públicos de Portugal ao longo de cada dia do ano de forma a fazer uma gestão mais inteligente dos recursos humanos e materiais.

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

(0.5V)

Para este tipo de problema poderia-se utilizar uma regressão, que permitiria prever o nº de clientes a partir de uma série de variáveis independentes.

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 comum haver um maior nº de acidentes e outros problemas aos fins de semana.
2. Cidade (obtido da localização do Hospital). Há cidades que, pelo seu nº de habitantes ou outras características, têm uma afluência maior de utentes nos hospitais.
3. Hospital (obtido da informação do Hospital). É importante porque, mesmo em cidades em que existam múltiplos hospitais, a afluência pode variar de hospital para hospital consoante diversos fatores (e.g. qualidade/rapidez do serviço, localização).
4. Temperatura (obtida do serviço web do IPMA). É importante pois em picos de calor ou frio existe uma maior afluência aos serviços de emergência.
5. Afluência em cada dia (obtida a partir dos registos do hospital). É a variável principal e permitirá, no futuro, treinar um modelo de regressão para prever a afluência, tal como pretendido.

2.3 Relativamente ao problema proposto, indique como procederia para realmente implementar um processo de recolha de dados. Isto é, como procederia para popular o dataset com dados relevantes e realistas para estudar o problema. Indique ainda se seria necessário, após a recolha dos dados e antes de aplicar o algoritmo escolhido na questão 2.1, a utilização de um ou mais algoritmos de pré-processamento, especificando qual ou quais.


(1V)

Para recolher dados para este dataset tentaria obter os dados relevantes dos registos de todos os Hospitais públicos de Portugal, nomeadamente do nº de utentes que cada um teve ao longo do tempo. Esta informação seria cruzada com a informação meteorológica e com a informação da localização do hospital. Se possível, e para melhorar o modelo, tentaria obter dados de vários anos.


Optaria ainda por normalizar os valores da afluência (que se situaria nos milhares), da temperatura (que oscilará, aproximadamente, entre 20 e 40 graus) e do dia da semana (que varia entre 1 e 7), de forma a que todas estas variáveis tenham a mesma relevância no treino do modelo.

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

- `posição(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 (valor entre 0 e 1), contra um determinado inimigo;

 <small>ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO</small>	Tipo de Prova Exame Teórico – Época Especial	Ano letivo 2016/2017	Data 13-07-2017
	Curso Licenciatura em Engenharia Informática	Hora 10:00	
	Unidade Curricular Inteligência Artificial	Duração 2:00 horas	

- `disparo(Inimigo, Acertou)` – representa um disparo feito a um determinado inimigo e o resultado (hit/no);
- 3.1 (2V) Defina o predicado `dispara/2` que determina se, avistando-se um determinado robô, se lhe deve disparar (e nesse caso com que potência) ou não. O robot dispara de acordo com os seguintes critérios:
- Dispara com potência 3 se a distância é inferior ou igual a 50 (independentemente do inimigo);
 - Se a distância está entre 50 e 100 (inclusive):
 - Dispara com potência 3 se o inimigo é o `SittingDuck`;
 - Dispara com potência 2 se o inimigo é qualquer outro que não o `SittingDuck` ou o `Crazy`
 - Se a distância está entre 100 e 200 (inclusive) dispara com potência entre 0 e 2, proporcionalmente à accuracy contra esse robot.
- ```
dispara(I,3):-distancia(I,D), D < 50.
dispara(sittingDuck,3):-distancia(sittingDuck,D), D > 50, D <= 100.
dispara(I,2):-I \= sittingDuck, I \= Crazy, distancia(I,D), D > 50, D <= 100.
dispara(I,P):-distancia(I,D), D > 100, D <= 200, accuracy(I,A), P is 2 * A.
```
- 3.2 (2V) Admita agora a inexistência do predicado `accuracy/2`, definido acima, e defina-o.
- ```
accuracy(I,P):-findall(I,disparo(I,hit),Lh),
findall(I,disparo(I,_),Lt), length(Lh,TamH), length(Lt,TamT), P is TamH/TamT.
```
- 3.3 (1V) Defina o predicado `dadosValidos/0`. O objetivo deste predicado é que verifique a validade dos factos existentes na base de conhecimento. Especificamente, o predicado deve falhar se houver informação sobre a accuracy de um determinado inimigo mas não houver registo de disparo a esse inimigo.
- ```
dadosValidos:- \+ (accuracy(I,_), \+ disparo(I,_)).
```
4. (1V) Indique um cenário de aplicação realista de um algoritmo de discretização. Indique ainda qual é o principal objetivo deste tipo de algoritmos.
- O principal objetivo dos algoritmos de discretização é o de transformar uma variável numérica numa variável discreta que, em vez de poder tomar um número potencialmente infinito de valores, passa a poder tomar apenas valores de um conjunto finito e discreto. Geralmente, estes algoritmos aplicam-se quando se pretende diminuir a complexidade dos dados e/ou aplicar algoritmos de classificação uma vez que o output destes algoritmos é uma classe e não um valor numérico. Por exemplo, as notas dos alunos de uma UC (entre 0 e 20) podem ser discretizadas para as seguintes classes: Reprovado, Suficiente, Bom, Excelente
5. (1.5V) Indique qual o papel e importância da função de similaridade num mecanismo de Raciocínio Baseado em Casos. Indique ainda de que forma é possível atribuir importâncias diferentes a diferentes variáveis.
- Num mecanismo de CBR, a função de similaridade tem como objetivo quantificar quão parecidos ou similares dois casos são. É de importância fundamental pois sem ela seria impossível esta quantificação e

|                                                                                                                                                        |                                                 |                         |                    |
|--------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------|-------------------------|--------------------|
| <br><small>ESCOLA<br/>SUPERIOR<br/>DE TECNOLOGIA<br/>E GESTÃO</small> | Tipo de Prova<br>Exame Teórico – Época Especial | Ano letivo<br>2016/2017 | Data<br>13-07-2017 |
|                                                                                                                                                        | Curso<br>Licenciatura em Engenharia Informática | Hora<br>10:00           |                    |
|                                                                                                                                                        | Unidade Curricular<br>Inteligência Artificial   | Duração<br>2:00 horas   |                    |

logo a escolha do caso mais similar, que é um elemento central num sistema de CBR. A função de similaridade é, essencialmente, uma soma pesada. Assim, é possível atribuir diferentes pesos a cada variável, dando uma maior importância a determinadas variáveis de acordo com a sua relevância para o problema

6. Comente, indicando se concorda com ela ou não, a seguinte afirmação:

(1.5V) Num sistema de IA, a aprendizagem pode ser supervisionada ou não supervisionada. Indique a principal diferença entre estes dois tipos de aprendizagem e dê exemplos de um algoritmo em cada um destes tipos de aprendizagem.

Na aprendizagem supervisionada são utilizados datasets que contêm dados de treino no sentido em que as instâncias já fazem corresponder um input a um output, que permitem ao modelo ajustar-se de forma a produzir um resultado similar. Na aprendizagem não supervisionada tal não acontece, tendo o algoritmo que encontrar um padrão ou uma estrutura nos dados que não é à partida conhecida. Um exemplo de um algoritmo de aprendizagem supervisionado é a classificação, enquanto que a segmentação é um exemplo de um algoritmo de aprendizagem não supervisionado.

7. Considere a seguinte afirmação:

(1V) *Vivemos num mundo com excesso de dados mas com falta de conhecimento.*

Comente a afirmação, indicando ainda o papel da Inteligência Artificial e do Data Mining neste contexto.

De facto, atualmente a quantidade de dados que são gerados por virtualmente todas as nossas atividades (virtuais ou não) é esmagador. Tudo o que nós fazemos, desde enviar um e-mail, consultar um site, conduzir, fazer uma chamada, jogar um jogo, etc. gera dados que são guardados algures e podem nunca vir a ser utilizados. Há assim um potencial enorme de descoberta de conhecimento que não é explorado porque a velocidade a que os dados são gerados e crescem é superior à nossa capacidade de os tratar. Nesse sentido, a IA e o Data Mining têm dado um contributo fundamental com técnicas e algoritmos que permitem, de forma quase automática, analisar estes dados e extrair daí conhecimento, ajudando-nos a perceber melhor como o mundo à nossa volta e nós próprios funcionamos.

8. Considere a seguinte afirmação:

(1V) *Para se aplicar corretamente um algoritmo de clustering/segmentação é fundamental saber-se de antemão qual o número de grupos que se pretende encontrar nos dados.*

Comente a afirmação, indicando ainda a sua veracidade.

A afirmação é falsa pois quando se utiliza um algoritmo de clustering não se sabe à partida quantos grupos se vão encontrar pois as instâncias não se encontram catalogadas. É o algoritmo que determina o nº de grupos que serão encontrados.