 ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO	Tipo de Prova Exame Teórico – Época de Recurso	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. Um médico de família tem vindo a registar, ao longo dos últimos anos e para cada paciente que examina, os seus sintomas, o diagnóstico, a medicação prescrita e o resultado da intervenção. Pretende-se que desenvolva um sistema para automatizar a prescrição de medicação a um paciente, dados os seus sintomas.
7. Uma vez que o que existe é um conjunto de casos específicos que ocorreram no passado e que na medicina, com frequência, os mesmos medicamentos são atribuídos aos mesmos sintomas, seria possível desenvolver um sistema baseado na semelhança a casos passados para prescrever a medicação para cada paciente
- B. Pretende-se perceber as condições que poderão estar a causar a extinção de uma determinada espécie animal. Assim, é-lhe fornecido um dataset que contém dados georreferenciados sobre o nº de animais dessa espécie, a quantidade de alimento disponível, o nível de desflorestação, o nº de incêndios ou outros desastres naturais, e o nº de predadores. Pretende-se que descubra quais as causas mais prováveis para a diminuição do nº de exemplares da espécie estudada.
4. Pretende-se perceber como duas ou mais variáveis se relacionam entre si. Por exemplo, sempre que o nº de desastres naturais e o nível de desflorestação aumenta, o nº de exemplares diminui
- C. Em cenários de emergência em que os meios médicos são escassos, é frequente fazer uma triagem que separa os pacientes com possibilidades de sobreviver daqueles sem qualquer hipótese, para focar os recursos existentes nos primeiros. Um médico transmite-lhe o processo de decisão que geralmente é seguido nestes casos e pede-lhe que implemente um sistema que o automatize.
8. Uma vez que o médico nos diz explicitamente como estas decisões são tomadas, i.e. com base em que critérios o que deve ser feito, esta informação pode ser traduzida em regras que podem ser utilizadas por um SBR
- D. É-lhe fornecido um dataset que descreve os produtos de uma linha de produção de biscoitos

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

embalados que foram rejeitados ou aprovados e quais as suas características no momento da decisão (e.g. peso da embalagem, estado da embalagem). É-lhe pedido que desenvolva um sistema baseado nesse dataset para automatizar a rejeição/aprovação de produtos na linha de produção.

1. Uma vez que o que se pretende é atribuir um valor de entre um conjunto finito e discreto (aprovado/rejeitado)

- E. Pretende-se que desenvolva um sistema para gerir receitas de culinária, que seja capaz de sugerir receitas similares às receitas já utilizadas ou marcadas como favoritas pelo utilizador. O sistema deve ainda ser capaz de adaptar receitas (e.g. a receita é para 4 pessoas mas o utilizador pretende cozinhar para 6).

7. Uma vez que o que se pretende é raciocinar por analogia, i.e., encontrar e adaptar casos similares

- F. A quantidade de azeitona produzida depende, entre outros fatores, das características do solo, da área plantada, do número de oliveiras, da quantidade de chuva e da temperatura durante o ano. É-lhe dado um dataset com esta informação e é-lhe pedido que preveja a quantidade de azeitona que será produzida em determinadas plantações.

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

- G. É-lhe dado acesso a um dataset que faz corresponder, para cada estudante, a sua média de entrada na Universidade, a sua idade e o nível de escolaridade dos pais à sua média de fim de curso. No futuro pretende-se vir a treinar um classificador para determinar, à entrada da Universidade, se o aluno conclui com média superior ou inferior/igual a 15. Neste momento, pedem-lhe que prepare o dataset para essa tarefa.

6. Seria necessário discretizar a nota final do aluno em duas classes:  $>15$  e  $\leq 15$ , para posteriormente treinar o classificador

- H. Pretende-se que desenvolva um sistema para substituir os juizes num tribunal, que baseie a sua tomada de decisão na legislação vigente.

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


- I. Num festival de verão, o consumo de cerveja é influenciado por variáveis como o nº de pessoas, a temperatura ou a hora do dia. Pretende-se que desenvolva um sistema que preveja a quantidade de cerveja consumida em cada momento de um festival.

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

- J. É-lhe fornecido um dataset descrevendo os perfis (e.g. idade, localização, quantidade de dinheiro gasto por mês, produtos favoritos) de um grupo de clientes de uma loja online. Pretende-se encontrar clientes com perfis similares para enviar publicidade direcionada.

2. Uma vez que o que se pretende é encontrar segmentos de clientes que partilhem de características similares

2. Considere o seguinte problema:

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

*Pretende-se estudar modelar a distância de travagem dos veículos (curta, média ou longa) em 5 tipos de asfalto diferentes que uma empresa comercializa. O objetivo é decidir, em cada estrada que se asfalte e mediante as características dos veículos que por ela circulam, o tipo de asfalto mais indicado de forma a reduzir a distância de travagem e assim evitar acidentes.*

2.1  
(0.5V) Indique, justificando, qual o algoritmo ou técnica, de entre os abordados na aula, que utilizaria para implementar este sistema.

O algoritmo mais adequado seria um algoritmo de classificação uma vez que o objetivo é, para cada instância, determinar a distância e esta assume um de três valores discretos possíveis.

2.2  
(2.5V) 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.

1. Tipo de piso (de entre os 5 existentes). Obtido por observação do piso. É importante porque diferentes tipos de piso têm efeitos diferentes na distância de travagem (e.g. quanto menos rugoso maior a distância de travagem)
2. Peso do carro. Obtido de uma balança industrial. É importante porque quanto maior o peso do carro, menor a distância de travagem
3. Condição do piso (e.g. seco, molhado). Obtido por observação do piso. É importante porque a existência de água, por exemplo, aumenta a distância de travagem
4. Nº de eixos do carro. Obtido do livrete do carro ou por observação do carro. É importante porque um maior nº de eixos implica um maior nº de pneus, logo uma menor distância de travagem
5. Distância percorrida após travagem (em cms). Obtida com uma fita métrica. É a variável fundamental neste problema e é a variável que se pretende estudar em função das restantes.


2.3  
(1V) 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.

Para recolher dados para este dataset escolheria carros com diferentes características e testaria cada um em cima de cada um dos pisos a estudar. Para isso, aceleraria o carro até uma certa velocidade (igual para todos) e, num determinado ponto, travaria a fundo. Após a imobilização do carro, mediria a distância percorrida entre a sua posição final e o local do início da travagem.

Uma vez que a distância de travagem é medida em centímetros, seria necessário discretizar este valor para que passasse a ter apenas 3 valores possíveis, tal como especificado no enunciado: curta, média ou longa

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;
- `disparo(Inimigo, Acertou)` – representa um disparo feito a um determinado inimigo e

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

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 `velhosConhecidos/1` que determina se já alguma vez disparamos a um determinado inimigo.

```
velhosConhecidos(I):-disparo(I,_).
```

4.  
(1V)


Indique um cenário de aplicação realista de um algoritmo de normalização. Indique ainda qual é o principal objetivo deste tipo de algoritmos.

O principal objetivo dos algoritmos de normalização é o de atribuir a todas as variáveis gamas de valores iguais, para que todas as variáveis tenham a mesma importância no momento de treino do modelo. Um exemplo concreto poderia ser por exemplo em datasets robocode contendo as coordenadas X e Y (que variam entre 0 e 800 ou mais), a potência de disparo (que varia entre 0 e 3) e a velocidade dos robots (que varia entre -8 e 8).

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 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.

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

6. Comente, indicando se concorda com ela ou não, a seguinte afirmação:
- (1V) A utilização de uma abordagem de Raciocínio Baseado em Casos na modelação de um problema ou no suporte à tomada de decisão apenas faz sentido em cenários em que seja necessária aprendizagem.
- Não concordo com a afirmação. Na minha opinião, um sistema CBR pode ser interessante enquanto abordagem para a resolução de problemas mesmo que não exista aprendizagem. Nomeadamente, se já houver uma base de casos suficientemente grande que represente o problema de forma adequada, não é necessário aprender novos casos. No entanto, o sistema pode continuar a ser utilizado para encontrar soluções para cada problema encontrado. Para além disso, pode-se ainda tirar proveito do mecanismo de raciocínio por analogia e do mecanismo de adaptação do CBR
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. Tal como abordado nas aulas, existe uma grande diferença de performance entre um sistema baseado em regras e um sistema baseado em casos. Apoiando-se no que já sabe sobre as duas alternativas, explique a que se deve esta diferença.
- (1.5V) A diferença entre as duas abordagens deve-se, essencialmente, à forma como o conhecimento é estruturado em cada um deles. Num sistema baseado em regras o conhecimento é representado sob a forma de factos ou predicados com um valor lógico, o que faz com que a sua utilização seja extremamente simples e eficiente, mesmo para representar grandes quantidades de informação. Já num sistema baseado em casos, o conhecimento é representado caso a caso. Quer isto dizer que se houver uma grande quantidade de informação a representar isto implica a utilização de um elevado número de casos. Para além disso, sempre que é necessário utilizar este conhecimento é necessário percorrer todos os casos existentes e comparar cada um deles com o caso atual (através da função de similaridade) o que é um processo muito complexo computacionalmente e que aumenta de forma considerável o tempo de pesquisa.