 <small>ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO</small>	Tipo de Prova Exame Teórico – Época Normal	Ano letivo 2018/2019	Data 19-06-2019
	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.
- A Parte 1 deste exame é constituída por questões de escolha múltipla. As mesmas devem ser respondidas na folha de resposta.
- O enunciado deve ser entregue juntamente com a folha de resposta.

Número: _____ Nome: _____

PARTE I

1. (1V) No treino de um modelo de aprendizagem supervisionada, o algoritmo deve:
 - A. Favorecer overfitting e generalização
 - B. Evitar overfitting e generalização
 - C. Evitar overfitting e favorecer generalização
 - D. Favorecer overfitting e evitar generalização
2. (1V) Uma Rede Neuronal:
 - A. Tem no mínimo duas camadas
 - B. Tem no mínimo três camadas
 - C. Tem sempre mais inputs que outputs
 - D. Tem sempre mais outputs que inputs
3. (1V) Durante o treino de uma Rede Neuronal a aprendizagem ocorre:
 - A. Acrescentando novas camadas à Rede Neuronal
 - B. Acrescentando novos neurónios a camadas existentes na Rede Neuronal
 - C. Acrescentando ligações entre neurónios
 - D. Fazendo alterações nos pesos das ligações entre neurónios
4. (1V) Considere a existência da variável numérica *idade*, num determinado dataset com informação demográfica. Pretende transformar-se esta variável numa enumeração com os valores {criança, jovem, adulto, idoso}. Identifique a técnica adequada:
 - A. Normalização
 - B. Nenhuma das restantes
 - C. Classificação
 - D. Discretização
5. (1V) Uma Árvore de Decisão de classificação:
 - A. Tem uma folha por cada valor diferente da variável dependente
 - B. Tem um valor específico de uma variável independente (ou uma gama de valores) em cada ramo
 - C. Tem um nível por cada variável dependente
 - D. Tem um ramo por cada variável dependente
6. (1V) No treino de um modelo com N-Fold Cross Validation:
 - A. São treinados N modelos
 - B. O erro é dado pelo erro médio de N modelos
 - C. O erro é dado pelo erro médio de N+1 modelos
 - D. Nenhuma das restantes

 <small>ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO</small>	Tipo de Prova Exame Teórico – Época Normal	Ano letivo 2018/2019	Data 19-06-2019
	Curso Licenciatura em Engenharia Informática	Hora 10:00	
	Unidade Curricular Inteligência Artificial	Duração 2:00 horas	

7. (1V) Assinale a opção verdadeira:
- A. A Confusion Matrix pode ser criada tanto para um modelo de regressão como de classificação
 - B. A ROC curve pode ser criada em qualquer problema de classificação
 - C. A ROC curve pode ser criada em qualquer problema de regressão
 - D. Nenhuma das restantes

8. (1V) O Case Based Reasoning é um algoritmo incluído na categoria de:
- A. Instance-based Learning
 - B. Supervised Learning
 - C. Unsupervised Learning
 - D. Reinforcement Learning

PARTE II

9. (2V) Nas aulas de IA foram abordadas duas grandes formas de aprendizagem: supervisionada e não supervisionada. Indique em que consiste cada uma delas, indicando ainda as suas diferenças fundamentais e um exemplo de aplicação para cada uma delas.

A aprendizagem supervisionada permite o treino de modelos que fazem corresponder um conjunto de variáveis independentes a uma variável dependente. Para isto, é necessária a utilização de um dataset devidamente etiquetado, contendo tanto as variáveis independentes como a(s) dependente(s). Ou seja, os modelos que resultam de aprendizagem supervisionada permitem prever um output em função de um ou mais inputs. Na aprendizagem não supervisionada não existem variáveis dependentes. Assim, este tipo de aprendizagem apenas permite identificar padrões ou grupos nos dados. Isto é, encontrar conjuntos de dados que tenham características semelhantes. Um exemplo de aplicação de aprendizagem supervisionada é a utilização de um algoritmo de Random Forest para treinar um modelo que prevê se um aluno passa ou não numa determinada UC, dadas variáveis como as horas de estudo, a assiduidade ou o seu QI. Um exemplo de aplicação de aprendizagem não supervisionada é a utilização de um dataset de dados de clientes de uma superfície comercial (e.g. idade, estado civil, morada) para encontrar grupos de clientes semelhantes, para eventualmente criar publicidades direcionadas.

10. (2V) Considere o excerto do dataset que se apresenta de seguida, que descreve algumas características dos alunos da Unidade Curricular de Inteligência Artificial. O dataset contém as seguintes variáveis:

- _id – O identificador numérico único de cada aluno
- QI – O nível de QI do aluno
- N_Irmaos – O número de irmãos do aluno
- Sal_AgregadoF – O salário, em euros, do agregado familiar
- Nota – Uma enumeração com três valores possíveis {fraco, médio, bom} descrevendo a nota qualitativa do aluno na UC

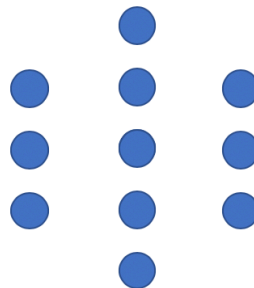
_id	QI	N_Irmaos	Sal_AgregadoF	Nota
1	80	2	1500	Fraco
2	120	1	2340	Bom

P.PORTO ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO	Tipo de Prova Exame Teórico – Época Normal	Ano letivo 2018/2019	Data 19-06-2019
	Curso Licenciatura em Engenharia Informática	Hora 10:00	
	Unidade Curricular Inteligência Artificial	Duração 2:00 horas	

3	94	3	1400	Médio
...

Desenhe uma possível arquitetura de uma Rede Neuronal para prever a nota de alunos. Note que é suficiente desenhar cada uma das camadas que considerar necessárias e os seus neurónios, não sendo necessário desenhar as ligações entre os neurónios. Faça ainda as considerações que achar necessárias sobre as variáveis a utilizar.

Não é necessário utilizar a variável `_id` uma vez que não tem qualquer relevância para prever a nota do aluno. Uma possível arquitetura teria 3 neurónios na camada de input (um para cada variável independente relevante), um número arbitrário de neurónios na(s) hidden layer(s) e 3 neurónios de output (um para cada valor possível da nota).




11. (2V) No decorrer das aulas de IA foram estudadas diferentes formas diferentes de avaliar a performance de um modelo durante o seu treino. Uma delas consiste na divisão prévia do dataset em três datasets disjuntos (train, validation e test). Indique qual o objetivo desta divisão e, particularmente, qual a utilidade do dataset de test.

O principal objetivo desta divisão é adotar uma abordagem cientificamente válida para o treino de modelos. Especificamente, os datasets de treino e validação são usados para treinar o modelo e avaliar a sua performance durante o treino. Já o dataset de test é utilizado após o treino do modelo, para avaliar a sua performance com dados que este não conheceu durante o treino. Isto permite estimar quão bem o modelo se comportará no “mundo real” bem como avaliar a capacidade de generalização do modelo.

12. Considere a seguinte tabela que mostra os limites de velocidade nas estradas Portuguesas, em km/h, em função do tipo de estrada e da categoria do veículo:

Estrada Categoria	Urbana	Nacional	Autoestrada
A	50	60	120
B	50	90	120
C	50	90	100

Considere ainda que se um condutor tiver a carta há menos de 2 anos, os limites de velocidade aplicáveis

 <small>ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO</small>	Tipo de Prova Exame Teórico – Época Normal	Ano letivo 2018/2019	Data 19-06-2019
	Curso Licenciatura em Engenharia Informática	Hora 10:00	
	Unidade Curricular Inteligência Artificial	Duração 2:00 horas	

são diminuídos em 10 km/h.

12.1
(1.5V)

Modele, em Prolog, a informação que consta na tabela acima.

```
limite(_, urbana, 50):-!.
limite(a, nacional, 60):-!.
limite(_, nacional, 90):-!.
limite(c, autoestrada, 100):-!.
limite(_, autoestrada, 120).
```

12.2
(1.5V)

Implemente, em Prolog, o predicado excesso/4 que determina se um condutor circula ou não em excesso de velocidade, dada a idade da sua carta de condução (em anos), a categoria do veículo, o tipo de estrada em que circula e a sua velocidade. Indique ainda um exemplo de utilização do predicado.

```
excesso(I,C,T,V):-I<2, limite(C,T,L), L1 is L - 10, V > L1,!.
excesso(_,C,T,V):-limite(C,T,L), V > L.
```

13
(3V)

Considere a funcionalidade do GoogleMaps que ilustra, através de três cores diferentes (verde, amarelo e vermelho) a intensidade do trânsito nas estradas. Esta funcionalidade apenas permite visualizar o estado do trânsito em tempo real. Admita agora que se pretende desenvolver um modelo com a capacidade de prever a intensidade do trânsito em datas futuras, para que os utilizadores do GoogleMaps possam melhor planear as suas viagens de forma a evitar trânsito. Para este desafio, indique:

- Que fontes de informação poderia utilizar
- Que variáveis seriam extraídas dessas fontes de informação
- Qual a estrutura do dataset
- (se aplicável) que tarefas de preparação de dados aplicaria
- Que algoritmo poderia utilizar para treinar um modelo adequado

a) Sensores GPS colocados nos pontos das ruas onde se pretende contar o tráfego, Calendário, Relógio, Google Maps, Sensores de passagem laser colocados nos pontos onde se pretende contar o tráfego

b) Dos sensores GPS extrair as coordenadas (latitude e longitude);

Do calendário extrair a data, dia da semana, se é feriado ou não


Do relógio extraía a hora

Da API do Google Maps extraía a rua que corresponde a cada par de coordenadas

Dos sensores de passagem laser extraía o momento em que cada carro passa

c) O dataset teria uma linha por cada carro que passa em frente a cada sensor de passagem e teria a seguinte estrutura

Latitude, longitude, rua, data, dia da semana, feriado, hora

 <small>ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO</small>	Tipo de Prova Exame Teórico – Época Normal	Ano letivo 2018/2019	Data 19-06-2019
	Curso Licenciatura em Engenharia Informática	Hora 10:00	
	Unidade Curricular Inteligência Artificial	Duração 2:00 horas	

- d) Seria necessário transformar este dataset. Primeiro, agrupar os dados por estrada, coordenadas, data e hora, contando o número de linhas. De seguida discretizar esta nova variável com a contagem para ter 3 valores possíveis (intensidade de trânsito). Esta é a variável dependente. Por último, normalizar todas as variáveis numéricas.
- e) Para este problema poderia utilizar-se o algoritmo de Random Forest