 <small>ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO</small>	Tipo de Prova Exame Teórico – Época de Recurso	Ano letivo 2022/2023	Data 05-07-2023
	Curso Licenciatura em Engenharia Informática	Hora 10:00	
	Unidade Curricular Inteligência Artificial	Duração 2:30 horas	

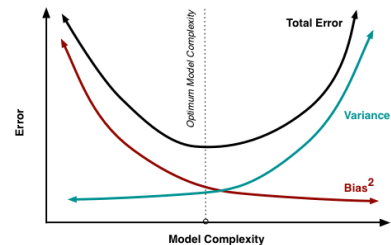
Observações:

- 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. Cada resposta errada desconta 0.25 valores da Parte 1.
- O enunciado deve ser entregue juntamente com a folha de resposta.

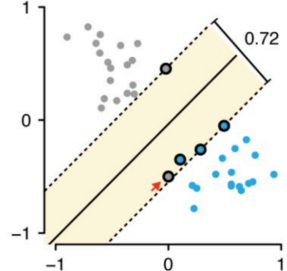
Número: _____ Nome: _____

PARTE I – Escolha Múltipla (10V)

- (1V) Considere a imagem que se apresenta à direita, que representa, de forma abstrata o compromisso (tradeoff) entre o erro por bias e por variance. Considere que se observa erro por variance num modelo do tipo Random Forest. Para tentar reduzir este tipo de erro pode-se:
 - Aumentar o número de épocas de treino
 - Diminuir o número de épocas de treino
 - Aumentar o número de árvores
 - Diminuir o número de árvores**
- (1V) Dos seguintes problemas, assinale aquele que poderia ser resolvido por um Algoritmo Genético.
 - Assinalar um e-mail como “spam” ou “não spam”, dadas as suas características
 - Prever o preço de uma casa, dada a sua área, localização e número de quartos
 - Elaborar o mapa de exames da ESTG, dadas as características das salas e de cada UC (e.g. ano curricular, semestre, número de inscritos)**
 - Nenhum dos restantes
- (1V) Considere que treinou dois modelos de regressão, para dois problemas de Machine Learning arbitrários, e que registou o RMSE de cada modelo. Assinale a afirmação correta.
 - O modelo com o RMSE mais baixo é garantidamente melhor
 - O modelo com o RMSE mais elevado é garantidamente melhor
 - O RMSE não é uma métrica válida para avaliação modelos de regressão
 - Sem informação adicional, não é possível concluir qual o melhor modelo**
- (1V) Quando comparados, genericamente, um modelo de Machine Learning tradicional A e um modelo B que faz parte de um Ensemble:
 - O modelo A é geralmente mais complexo que o modelo B**
 - O modelo A é geralmente menos complexo que o modelo B
 - O modelo A pode ter sido treinado com um sub-conjunto aleatório das instâncias existentes no dataset
 - O modelo A pode ter sido treinado com um sub-conjunto aleatório das colunas existentes no dataset



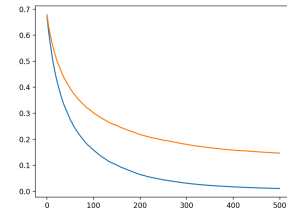
<p>P.PORTO</p> <p>ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO</p>	Tipo de Prova Exame Teórico – Época de Recurso	Ano letivo 2022/2023	Data 05-07-2023
	Curso Licenciatura em Engenharia Informática	Hora 10:00	
	Unidade Curricular Inteligência Artificial	Duração 2:30 horas	

5. (1V) Considere a seguinte afirmação: “As redes neuronais têm, no seu processo de treino, elementos que fazem com que duas redes com exatamente a mesma configuração, quando treinadas nos mesmos dados, resultem virtualmente sempre em modelos diferentes.”. Assinale a opção correta:
- A. A afirmação é incorreta
 - B. A afirmação é correta e isto deve-se à forma como os pesos são inicializados**
 - C. A afirmação é correta e isto deve-se ao algoritmo de backpropagation
 - D. A afirmação é correta e isto deve-se ao algoritmo de gradient descent
6. (1V) Considerando a imagem que se apresenta à direita, assinale a afirmação correta:
- A. A imagem representa um tipo de modelo que pode ser utilizado para regressão
 - B. A imagem representa um tipo de modelo que pode ser utilizado para classificação**
 - C. A imagem representa um tipo de modelo que pode ser utilizado para segmentação
 - D. A imagem não representa qualquer modelo de Machine Learning
- 
7. (1V) Considere que tem um conjunto de dados relativamente pequeno e que treinou duas Árvore de Decisão utilizando a mesma configuração (hyper-parâmetros). Após observar o RMSE de ambos os modelos, repara que estes têm performances bastante diferentes. Assinale a possível causa para esta diferença.
- A. A utilização de uma metodologia de train-test split para o treino e avaliação de cada um dos modelos**
 - B. O facto de o algoritmo de uma Árvore de Decisão inicializar a árvore aleatoriamente
 - C. A situação descrita é impossível pois não é possível calcular o RMSE numa árvore de decisão
 - D. Nenhuma das restantes
8. (1V) No decorrer das aulas foram abordadas várias metodologias para avaliar a performance de um modelo, entre eles a de validação cruzada (n-fold cross validation). Relativamente a esta metodologia, assinale a afirmação correta.
- A. Uma elevada variação de performance entre os diferentes modelos treinados indica uma forte capacidade de generalização
 - B. Uma elevada variação de performance entre os diferentes modelos treinados indica uma fraca capacidade de generalização**
 - C. Não é possível avaliar a capacidade de generalização de um modelo através de validação cruzada
 - D. Não é possível avaliar a performance de um modelo através de validação cruzada

<div>P.PORTO</div> <div>ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO</div>	Tipo de Prova Exame Teórico – Época de Recurso	Ano letivo 2022/2023	Data 05-07-2023
	Curso Licenciatura em Engenharia Informática	Hora 10:00	
	Unidade Curricular Inteligência Artificial	Duração 2:30 horas	

9. (1V) Considere a imagem que se apresenta à direita, que ilustra as chamadas curvas de aprendizagem, que resultam do treino de um modelo de ML. Assinale a afirmação correta.

- A. Num modelo de boa qualidade, a linha laranja (superior) indica tipicamente o erro de validação
- B. Num modelo de boa qualidade, a linha laranja (superior) indica tipicamente o erro de treino
- C. O eixo dos Y pode indicar a accuracy do modelo
- D. O eixo dos Y pode indicar a precision do modelo



10. (1V) Relativamente aos Algoritmos Genéticos, assinale a afirmação correta.

- A. Podem ser utilizados quer para problemas de classificação quer para problemas de regressão
- B. Podem ser utilizados tanto com conjuntos de dados grandes como pequenos
- C. Encontram sempre a melhor solução possível para o problema
- D. Encontram sempre mais que uma solução para o problema

PARTE II – Prolog (5V)


11. O atual conflito que decorre na Ucrânia após invasão da Rússia, tem e terá como uma das suas piores consequências a presença de milhares de minas terrestres por explodir, que representarão um risco para a população ucraniana durante décadas, até que todas as minas sejam detetadas e desativadas. Pretende-se implementar em Prolog um sistema de informação minimalista mas confiável para a gestão deste processo. Nesse sentido, considere a existência de uma Base de Conhecimento com a seguinte estrutura:

- mina(região, x, y, estado) – representa a existência de uma mina numa dada região (e.g. Kherson, Mariupol), em determinadas coordenadas (x, y) e num determinado estado. O estado pode conter os seguintes valores: x (explodida), a (ativa) ou d (desativada). Cada região tem uma grelha de coordenadas independente, pelo que as coordenadas se podem repetir entre regiões. No entanto, na mesma região, não pode existir mais que uma mina no mesmo par de coordenadas;
- desminagem(região, x, y, id_veiculo) – indica que uma mina na região e coordenadas especificadas foi desminada por um veículo de desminagem com um determinado identificador. O identificador é um número inteiro positivo.

11.1 (1V) Implemente em prolog o predicado multi_reg/1. Dado o identificador numérico de um veículo de desminagem, o predicado tem sucesso caso o veículo tenha efetuado desminagens em mais que uma região, falhando em qualquer outra situação.

11.2 (2V) Implemente em Prolog o predicado valida_desminagem/3, que será utilizado pelo exército ucraniano para validar a informação na Base de Conhecimento e minimizar a ocorrência de erros no reporte da informação ou a proliferação de notícias falsas. Dada uma região e um par de coordenadas, o predicado tem sucesso se (condições cumulativas):

- Existe uma mina nessa região/coordenadas no estado desativado;

 ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO	Tipo de Prova Exame Teórico – Época de Recurso	Ano letivo 2022/2023	Data 05-07-2023
	Curso Licenciatura em Engenharia Informática	Hora 10:00	
	Unidade Curricular Inteligência Artificial	Duração 2:30 horas	

- 11.3
(2V)
- Existe o registo de uma desminagem para essa região/coordenadas; e
 - Existe apenas uma mina registada nessa região/coordenadas.
- Implemente em Prolog o predicado `conta_activas/3`. Dada uma região, este predicado calcula o número total e a percentagem de minas ainda ativas nessa região.

PARTE III – Desenvolvimento (5V)

12.
(2.5V)
- Comente, de forma justificada, a seguinte afirmação, indicando ainda se concorda ou não com ela:
- “O hiper-parâmetro `max_depth`, que controla a profundidade máxima de uma árvore de decisão, deve ser definido de forma diferente quando se trata de uma árvore de decisão tradicional ou de uma Random Forest. Nomeadamente, o seu valor deve geralmente ser maior numa Random Forest do que numa árvore de decisão tradicional”.
13.
(2.5V)
- A tabela seguinte apresenta um conjunto de algoritmos e de conceitos abordados durante as aulas de Inteligência Artificial. Faça a correspondência entre cada conceito e o algoritmo que lhe está associado. Note que pode acontecer que um conceito esteja associado a mais que um algoritmo. Nesse caso é suficiente que faça a correspondência com uma das opções corretas. Pode ainda acontecer que não exista correspondência para alguns conceitos/algoritmos. Nesse caso, não é necessário fazer qualquer correspondência.

Conceito		Algoritmo	
A	Backpropagation	1	Neural Networks
B	Kernel Trick	2	Genetic Algorithms
C	Gradient Descent	3	Deep Learning
D	Fitness Function	4	Support Vector Machines
E	Feature Extraction	5	Decision Trees
F	Activation Function	6	Random Forest
G	Learning rate	7	K-Means
I	Epochs	8	K-NN
J	One-hot Encoding		
K	Pooling		
L	Kernel Function		
M	Distance/Similarity Metrics		