



Universidade do Minho

Licenciatura em Engenharia Informática

NÚMERO _____

NOME _____

Aprendizagem e Decisão Inteligentes

3º ano, 2º semestre

Ano letivo 2023/2024

Prova escrita

14 de maio de 2024



GRUPO 1

(6 valores)

QUESTÃO 1

RESPONDA ÀS QUESTÕES DESTE GRUPO EM FOLHA DE TESTE SEPARADA.

A resolução de problemas por aplicação de técnicas de *Machine Learning* envolve a identificação do tipo de problema em questão.

Para problemas de regressão, dê 3 exemplos de métricas de qualidade, explicando o seu significado.

QUESTÃO 2

O trabalho em grupo envolveu a realização de duas tarefas. Uma delas desenvolvia-se a partir de um *dataset* atribuído. A outra tarefa envolvia a seleção de um problema cujos dados eram escolhidos pelo grupo de trabalho.

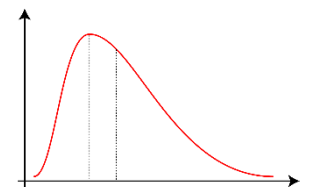
No que respeita à tarefa cujo *dataset* foi escolhido pelo grupo de trabalho:

- Indique o tipo de problema que o *dataset* reportava originalmente (classificação, regressão ou outro) descrevendo o(s) atributo(s) *target*.
- Identifique duas das técnicas de *Machine Learning* que foram aplicadas, relacionando com os resultados pretendidos para a abordagem ao problema.

QUESTÃO 3

Considere o gráfico à direita que representa a distribuição dos dados de um atributo contínuo com assimetria positiva, em que no eixo xx' estarão representados os valores do atributo e no eixo yy' a sua frequência.

Havendo a intenção de discretizar este atributo, explique as consequências de serem utilizadas estratégias de discretização de igual altura (*equal frequency binning*) e de igual largura (*equal width binning*).



GRUPO 2

(4 valores)

QUESTÃO 1

RESPONDA ÀS QUESTÕES DESTE GRUPO NO ESPAÇO RESERVADO PREENCHENDO OS ESPAÇOS VAZIOS COM EXPRESSÕES ADEQUADAS DE MODO QUE A DECLARAÇÃO SEJA CORRETA.

Numa Rede Neuronal Artificial, um neurónio é a _____ de composição da rede, o axónio é _____ e a sinapse é _____ entre _____ e _____.

QUESTÃO 2

Em *Machine Learning* a regressão linear é uma técnica para previsão de _____ e a regressão logística é uma técnica para previsão de _____.

QUESTÃO 3

Uma Árvore de Decisão é um grafo hierarquizado em que um _____ testa um atributo do *dataset*, cada _____ identifica uma alternativa do teste e cada _____ representa uma decisão.

QUESTÃO 4

Num paradigma da aprendizagem com supervisão é possível construir uma relação entre os valores _____ e os valores _____, uma vez que _____ utilizados para a aprendizagem _____ informações sobre os resultados pretendidos.

GRUPO 3

(6 valores)

PARA CADA AFIRMAÇÃO APRESENTADA, RESPONDA ASSINALANDO A SUA VERACIDADE (V) OU FALSIDADE (F). JUSTIFIQUE A RESPOSTA EXCLUSIVAMENTE NO ESPAÇO DISPONIBILIZADO.

NÃO SÃO CONSIDERADAS RESPOSTAS PARA AS QUAIS NÃO EXISTA JUSTIFICAÇÃO.

QUESTÃO 1

☐ *Machine Learning* é um paradigma de computação em que a característica essencial do sistema se revela pela sua capacidade de aprender com a experiência de modo autónomo e independente.

QUESTÃO 2

☐ Uma metodologia para análise de dados define que técnicas de *Machine Learning* deverão ser utilizadas no desenvolvimento de um projeto de aprendizagem automática para a resolução de problemas.

QUESTÃO 3

☐ Na fase de preparação de dados, a discretização de igual altura transforma um atributo numérico em categórico, e a discretização de igual largura transforma um atributo categórico em numérico.

QUESTÃO 4

☐ Técnicas de *Machine Learning* baseadas em Árvores de Decisão são utilizadas para a resolução de problemas de regressão.

QUESTÃO 5

☐ O treino de uma Rede Neuronal Artificial corresponde à aplicação de regras de aprendizagem que visam fazer variar os pesos das ligações entre neurónios.

QUESTÃO 6

☐ Numa Rede Neuronal Artificial *feedforward* podem existir ligações entre os neurónios de uma mesma camada intermédia.

GRUPO 4

(4 valores)

AS QUESTÕES DESTE GRUPO LISTAM AFIRMAÇÕES VERDADEIRAS (V) OU FALSAS (F).

PARA CADA QUESTÃO DESTE GRUPO, UMA AFIRMAÇÃO INCORRETAMENTE ASSINALADA ANULA OUTRA ASSINALADA CORRETAMENTE. UMA AFIRMAÇÃO NÃO ASSINALADA NÃO É CONSIDERADA.**QUESTÃO 1**A Figura 1 apresenta um excerto de um *dataset* utilizado no decorrer das aulas com dados sobre os passageiros do desastre do Titanic.

Figura 1
Dataset "titanic.xlsx"

	A	B	C	D	E	F	G	H	I	J	K	L
1	RowId	Survived	PClass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
2	1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7,25		S
3	2	1	1	Cumings, Mrs. John Bradley	female	38	1	0	PC 17599	71,2833	C85	C
4	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 31	7,925		S
5	4	1	1	Futrelle, Mrs. Jacques Heath	female	35	1	0	113803	53,1	C123	S
6	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8,05		S
7	6	0	3	Moran, Mr. James	male		0	0	330877	8,4583		Q
8	7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	51,8625	E46	S
9	8	0	3	Palsson, Master. Gosta Leon	male	2	3	1	349909	21,075		S
10	9	1	3	Johnson, Mrs. Oscar W (Elise	female	27	0	2	347742	11,1333		S
11	10	1	2	Nasser, Mrs. Nicholas (Adel	female	14	1	0	237736	30,0708		C
12	11	1	3	Sandstrom, Miss. Marguerite	female	4	1	1	PP 9549	16,7	G6	S
13	12	1	1	Bonnell, Miss. Elizabeth	female	58	0	0	113783	26,55	C103	S
14	13	0	3	Saunders, Mr. William Henry	male	20	0	0	A/5. 2151	8,05		S
15	14	0	3	Andersson, Mr. Anders Johan	male	39	1	5	347082	31,275		S
16	15	0	3	Vestrom, Miss. Hulda Amanda	female	14	0	0	350406	7,8542		S
17	16	1	2	Hewlett, Mrs. (Mary D King	female	55	0	0	248706	16		S
18	17	0	3	Rice, Master. Eugene	male	2	4	1	382652	29,125		Q
19	18	1	2	Williams, Mr. Charles Eugene	male		0	0	244373	13		S

- ☐ As células vazias, como por exemplo «F7» e «K2», representam *missing values*.
- ☐ Os dados da coluna «A» representam conhecimento que não tem relevância analítica.
- ☐ Não é possível preencher as células vazias da coluna «K» por se tratar de um atributo alfanumérico.
- ☐ Não é possível discretizar os dados da coluna «F» por conter células vazias.
- ☐ Apesar de a coluna «B» estar preenchida com números, representa conhecimento categórico.
- ☐ Os dados da coluna «L» representam conhecimento binário.

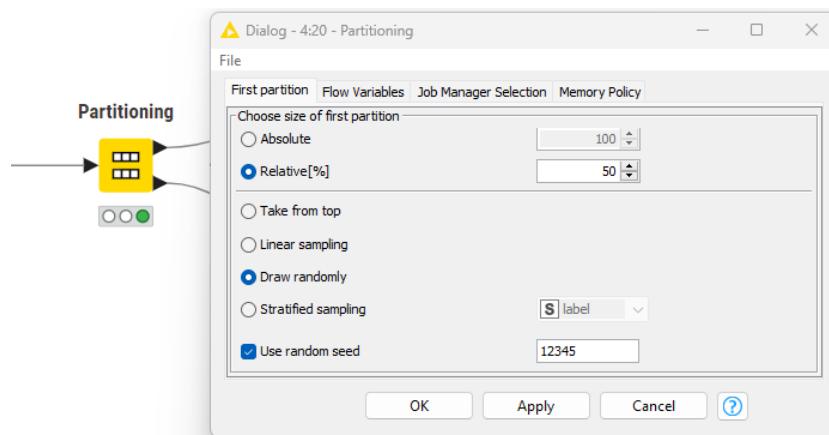
QUESTÃO 2Em *Machine Learning*, técnicas de segmentação (*clustering*):

- ☐ São técnicas usadas para construir modelos para previsão de resultados contínuos.
- ☐ Seguem um paradigma de aprendizagem por reforço.
- ☐ São técnicas de aprendizagem automática.
- ☐ Seguem um paradigma de aprendizagem supervisionada.
- ☐ São técnicas usadas para construir modelos de classificação.
- ☐ Seguem um paradigma de aprendizagem não supervisionada.

QUESTÃO 3

Considere o nodo «Partitioning» ilustrado na Figura 2.

Figura 2
Configuração de
nodo Knime



- ☐ O nodo implementa uma estratégia de validação de modelos denominada *cross validation*.
- ☐ A configuração do nodo é adequada para permitir a replicabilidade das experiências.
- ☐ A opção «Draw randomly» é incompatível com a opção «Use random seed».
- ☐ A opção «Relative [%]» tem de ser sempre superior a 50.
- ☐ O nodo «Partitioning» implementa uma estratégia de treino designada *hold-out validation*.
- ☐ Quando assinalada, a opção «Stratified sampling» cria uma nova amostra do *dataset* designada «label».