TÉCNICO LISBOA

Aprendizagem 2023/24

Lição de casa eu

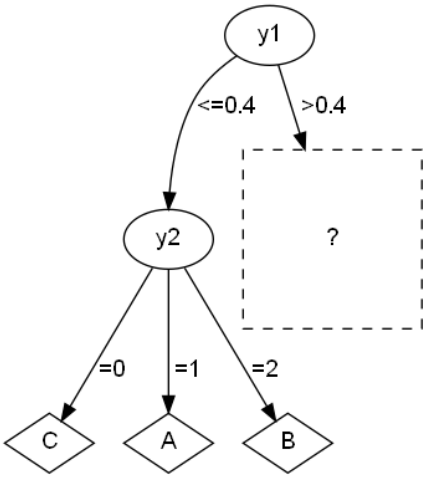
Prazo: 29/9/2023 (sexta-feira) 23h59 via Fenix como PDF

- Envie Gxxx.ZIP no Fenix, onde xxx é o número do seu grupo. O ZIP deve conter dois arquivos: Gxxx.pdf com seu relatório e Gxxx.ipynb com seu notebook python
- É possível enviar várias vezes no Fenix para evitar problemas de última hora. No entanto, apenas a última submissão é mantida
- A troca de ideias é incentivada. No entanto, se a cópia for detetada após a autorização automática ou manual, o trabalho de casa é anulado e as diretrizes do IST aplicam-se aos partilhadores e consumidores de conteúdos, independentemente da intenção subjacente.
- Consulte o FAQ antes de postar perguntas aos anfitriões do corpo docente

I. Caneta e papel[11v]

Considere a árvore de decisão parcialmente aprendida do conjunto de dados . é descrito por quatro variáveis de entrada – uma numérica com valores em [0,1] e 3 categóricas – e uma variável alvo com três classes.

	1	2	3	4	fora
1	0,24	1	1	0	A
2	0,06	2	0	0	B
3	0,04	0	0	0	B
4	0,36	0	2	1	C
5	0,32	0	0	2	C
6	0,68	2	2	1	A
7	0,9	0	1	2	A
8	0,76	2	2	0	A
9	0,46	1	1	1	B
10	0,62	0	0	1	B
11	0,44	1	2	2	C
12	0,52	0	2	0	C



1) [5v] Complete a árvore de decisão fornecida usando ganho de informação com entropia de Shannon (registro2). Considere que: i) são necessárias no mínimo 4 observações para dividir um nó interno, e ii) as decisões serão colocadas em ordem alfabética crescente em caso de empate.

2) [2.5v] Desenhe a matriz de confusão de treinamento para a árvore de decisão aprendida.

3) [1.5v] Identifique qual classe tem a pontuação F1 de treinamento mais baixa.

4) [1v] Considerandosim2ser ordinal, avalie sesim1esim2são correlacionados usando o coeficiente de Spearman.

5) [1v] Desenhe os histogramas relativos condicionais de classe desim1usando 5 caixas igualmente espaçadas em [0,1]. Desafio: encontrar a divisão da raiz usando as regras discriminantes destas distribuições empíricas.

TÉCNICO LISBOA

Aprendizagem 2023/24

Lição de casa eu

Prazo: 29/9/2023 (sexta-feira) 23h59 via Fenix como PDF

II. Programação[9v]

Para responder às seguintes perguntas, considere usar oaprenderDocumentação da API e os notebooks na página do curso como orientação. Mostre em seu relatório PDF o código e os resultados correspondentes.

Considere ocoluna_diagnóstico.arffdados disponíveis na aba lição de casa, compreendendo 6 características biomecânicas para classificar 310 pacientes ortopédicos em 3 classes (normal, hérnia de disco, espondilolistese).

1) [1,5v] Aplicarf_classifdeaprenderavaliar o poder discriminativo das variáveis de entrada. Identifique a variável de entrada com maior e menor poder discriminativo.
   Trace as funções de densidade de probabilidade condicional de classe dessas duas variáveis de entrada.

2) [4v] Usando uma divisão estratificada de treinamento-teste 70-30 com uma semente fixa (estado_aleatório=0), avaliam em um único gráfico as precisões de treinamento e teste de uma árvore de decisão com limites de profundidade em {1,2,3,4,5,6,8,10}e os parâmetros restantes como padrão.

   *[opcional]*Observe que o limiar dividido de variáveis numéricas em árvores de decisão não é determinístico emaprender, portanto, você pode optar por calcular a média dos resultados usando 10 execuções por parametrização.

3) [1.5v] Comente os resultados, incluindo a capacidade de generalização entre ambientes.

4) [2v] Para implantar o preditor, uma equipe de saúde optou por aprender uma única árvore de decisão (estado_aleatório=0) usando*todos*dados disponíveis como dados de treinamento e garantindo ainda que cada folha tenha um mínimo de 20 indivíduos para evitar riscos de overfitting.

   eu. Trace a árvore de decisão.

   ii. Caracterize uma condição de hérnia identificando as associações hérnia-condicionais.

FIM