**I. Pen-and-paper**

1. Começamos por calcular a matrix de desenho (de tamanho 8x4 pois são 8 observações e M=3)

Exemplificação dos cálculos para as primeiras duas linhas da matrix:

Para obter a função de aproximação , iremos de seguida obter o vetor w que minimiza o erro da função em relação aos dados de treino. Seguidamente obtemos o vetor **w**.

.

Assim acabamos por obter a expressão para a função de aproximação



|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| *1* | 1 | 2.837 | 3.375 |
|  | 3 | 3.071 | 0.005 |
|  | 2 | 2.603 | 0.364 |
|  | 0 | 2.3 | 5.291 |
|  | 6 | 5.076 | 0.853 |
|  | 4 | 2.606 | 1.944 |
|  | 5 | 2.211 | 7.766 |
|  | 7 | 7.323 | 0.104 |

Exemplificação do cálculo de

= 2.837

1. Para a binarização da variável começamos por ordenar os seus valores no conjunto de treino e obter uma mediana de 3.5. Assim, todos os valores menores que a mediana pertencerão a um dos bins, enquanto que os restantes valores farão parte do outro bin.

De seguida, proseguimos com a escolha das variáveis para a construção da árvore de decisão. Começamos por obter o ganho de informação para cada uma das variáveis, escolhendo para raiz a variável , uma vez que possui o maior ganho de informação.

O processo de escolha de variáveis é repetido para cada nó, até ser obtida a árvore na sua totalidade.

Para , nenhuma das variáveis apresenta ganho de informação pelo que o nó fica indeterminado, com 50% de probabilidade de classificação como P e 50% de probabilidade de classificação como N.

Para , ambas as variáveis possuem o mesmo valor de ganho de informação pelo que podemos optar por escolher qualquer uma das duas. Neste caso optamos por escolher .

Para , verificamos que a entropia da variável de output tem valor 0, pelo que este nó classifica a observação como P.

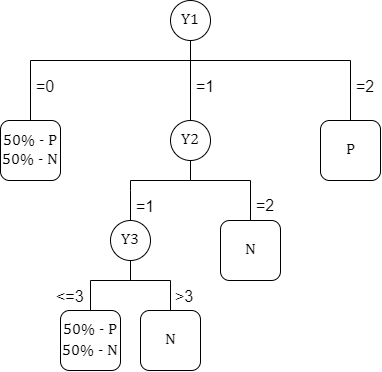
0

Para , , dado que a entropia da variável de output é também 0, obtemos um nó que classifica a observação como N.

Para , escolhemos a variável que resta uma vez que apresenta ganho de informação

Para , obtemos uma entropia de 1 para a variável de output, ficando o nó indeterminado e classificando as observações como P ou N com igual probabilidade.

Para , obtemos uma entropia de 0 para a variável de output, pelo que o nó é determinado e classifica as observações como N .



1. De acordo com a árvore de decisão obtida, podemos perceber que as observações de teste são classificadas como N e P respetivamente. Assim, podemos concluir que a eficácia dos modelos nestas observações é de 0% uma vez que as classificações corretas para as observações seriam P para e N .

**II. Programming and critical analysis**

1. Answer 6
2. Answer 7
3. Answer 8

**III. APPENDIX**

Paste your programming code here using Consolas 9pt or 10pt.

Use **highlighting** or colored text to facilitate the analysis by your faculty hosts.

**END**