

Universidade Federal de Ouro Preto
BCC 740 - Inteligência Artificial
Prova 2

1. Considere a solução de uma regressão linear obtida pelo método dos mínimos quadrados dada por $\mathbf{w} = (X^t X)^{-1} X^t \mathbf{y}$, e responda:
 - (a) O que representa o vetor \mathbf{w} ?
 - (b) O que representa a matriz \mathbf{X} ?
 - (c) O que representa o vetor \mathbf{y} ?
 - (d) Como essa equação é obtida?
 - (e) É possível utilizar este método para resolver problemas em que a relação entre variável dependente (o alvo), y , e as variáveis independentes (atributos de entrada), \mathbf{x} , não é linear? Como?
2. Considere a base de dados abaixo:

Atributo1	Atributo2	Classe
1	2	Classe1
2	3	Classe1
3	4	Classe2
4	5	Classe2
5	20	Classe1
6	30	Classe1
7	40	Classe2
8	50	Classe2

Table 1: Exemplo de base de dados

- (a) Calcule o gini para a condição $Atributo1 \leq 4.5$. $I_G(p) = 1 - \sum_{i=1}^J p_i^2$
 - (b) Quais seriam condições ótimas, em relação ao gini, após selecionarmos como raiz da árvore de decisão o critério $Atributo1 \leq 4.5$. Desenhe a árvore completa.
3. Quando dizemos que um algoritmo de aprendizado de máquina “está aprendendo”, que processo algorítmico está acontecendo?
4. Qual tipo de problema resolvemos com regressão logística?
5. Como o vetor de pesos \mathbf{w} é obtido quando temos um modelo de regressão logística?
6. (O que é overfitting? Quais são os indícios de que um modelo está sofrendo de overfitting? De forma geral, o que deve ser feito para diminuir o overfitting?)
7. Quais as vantagens do algoritmo de busca em largura sobre o algoritmo de busca em profundidade? E quais as desvantagens?
8. Quando devemos utilizar o algoritmo de busca *branch and bound*?