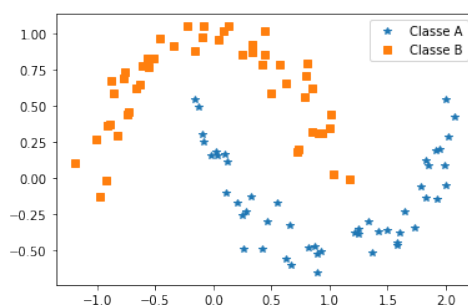


Universidade Federal de Ouro Preto  
BCC 325 - Inteligência Artificial  
Prova 2  
Prof. Rodrigo Silva

1. Para se obter os pesos,  $\mathbf{w}$ , de um modelo de regressão linear utilizamos a expressão  $\mathbf{w} = (X^t X)^{-1} X^t \mathbf{y}$ . Considerando a seguinte base de dados, responda:

Atributo1	Atributo2	Classe
1.2	2	3
2.3	3	5
3.5	4	6
4	5	8
5.7	2	2
6.7	2.5	9

- Defina  $X$  em termos da base de dados apresentada. Considere a necessidade de determinar a constante que representa a interceptação do modelo com o eixo vertical.
  - Defina o vetor  $\mathbf{y}$  em termos da base de dados apresentada.
  - Explique como a equação  $\mathbf{w} = (X^t X)^{-1} X^t \mathbf{y}$  é obtida.
  - É possível utilizar este método para resolver problemas em que a relação entre variável dependente (atributo alvo),  $y$ , e as variáveis independentes (atributos de entrada),  $\mathbf{x}$ , não é linear? Como?
  - Apresente  $X$  para o caso em que suspeitamos que a relação entre  $\mathbf{x}$  e  $y$  é quadrática.
2. Considere os dados abaixo:



- Qua tipo de problema resolvemos com regressão logística?
- É possível resolver este problema com um aplicação direta do algoritmo de regressão logística com as variáveis definidas nos eixos  $x$  e  $y$ ? Por quê?
- O que poderia ser feito para que este problema seja resolvível com regressão logística?
- Como o vetor de pesos  $\mathbf{w}$  é obtido quando temos um modelo de regressão logística?

3. Quando dizemos que um algoritmo de aprendizado de máquina “está aprendendo”, que processo algorítmico está acontecendo?
4. 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

- (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 essa árvore.
5. 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?
6. Quais as vantagens do algoritmo de busca em largura sobre o algoritmo de busca em profundidade? E quais as desvantagens?
7. Quando devemos utilizar o algoritmo de busca A\*?