



- Observações:
- A prova deverá ser realizada individualmente, sem consulta;
 - A interpretação das questões é parte integrante da avaliação;

Declaro que realizei esta prova sem consultar outras pessoas e materiais e que estou ciente do código de ética da USP.

- 1) Construa uma árvore de regressão, sem poda, para a base de dados abaixo e estime o valor de Y para a tupla [S N]. **(2 pontos)**

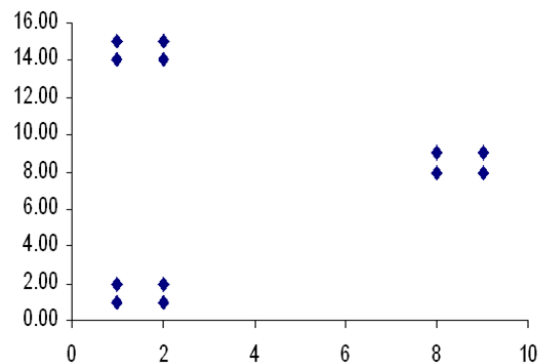
A ₁	A ₂	Y
S	S	1
S	N	5
S	S	1
S	N	3
N	S	7
N	N	6
N	S	5

- 2) Considere a base de treinamento abaixo. Estime o valor de Y para a tupla [1 1] de acordo com o método dos vizinhos mais próximos (k -NN, com $k=3$) utilizando a distância Euclidiana. **(2 pontos)**

A ₁	A ₂	Y
0	1	3
1	1	2
1	0	1
5	5	0
0	0	-1
2	2	2
4	4	3

- 3) Obter os centroides do algoritmo k -means, com $k=3$, nos dados da tabela/gráfico abaixo a partir dos protótipos [2 12], [1 4] e [8 6] após 2 iterações. **(2 pontos)**

Objeto x_i	x_{i1}	x_{i2}
1	1	2
2	2	1
3	1	1
4	2	2
5	8	9
6	9	8
7	9	9
8	8	8
9	1	15
10	2	15
11	1	14
12	2	14



- 4) Considere uma partição de referência formada por duas categorias $\mathbf{P}=\{\mathbf{P}_1, \mathbf{P}_2\}$, sendo $\mathbf{P}_1=\{\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_6\}$ e $\mathbf{P}_2=\{\mathbf{x}_2, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_7\}$, e um conjunto de grupos $\mathbf{C}=\{\mathbf{C}_1, \mathbf{C}_2\}$ obtidos por um algoritmo particional, sendo $\mathbf{C}_1=\{\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$ e $\mathbf{C}_2=\{\mathbf{x}_2, \mathbf{x}_6, \mathbf{x}_7\}$. Calcular o índice externo de *Rand* (2 pontos).
- 5) Considere que os centróides obtidos pelo algoritmo *k-means*, com $k=3$, são aqueles obtidos por inspeção visual. Considerando a silhueta simplificada, calcule o valor de $s(i)$ para $\mathbf{x}_3=[1,1]$. Por simplicidade, considere que a medida de distância é a soma das diferenças absolutas dos valores para cada atributo, isto é, distância = $|\Delta_{x1}| + |\Delta_{x2}|$. (2 pontos)

Objeto \mathbf{x}_i	x_{i1}	x_{i2}
1	1	2
2	2	1
3	1	1
4	2	2
5	8	9
6	9	8
7	9	9
8	8	8
9	1	15
10	2	15
11	1	14
12	2	14

