



NOME: Aberto Abner Sento

AValiação II

1. (2,0) Considere o seguinte conjunto de dados e determine a classificação obtida para o exemplo $v_2v_4v_9$, utilizando o algoritmo naïve Bayes.

x1	x2	x3	Classe
v1	v3	v7	P
v1	v3	v8	N
v1	v4	v9	P
v2	v5	v9	N
v2	v6	v8	P

2. (1,0) Considere o seguinte conjunto de dados e determine a probabilidade $P(z_i | P)$ e $P(z_i | N)$, para $i = 1$ e 3.

	z1	z2	z3	z4	X
X1	0	0	1	0	N
X2	1	0	0	0	N
X3	0	1	1	1	P
X4	1	1	1	0	P
X5	0	0	1	0	N
X6	1	1	1	1	P
X7	0	1	0	1	N
X8	0	0	0	0	N

3. (2,0) Os dados a seguir provêm de um experimento para testar o desempenho de uma máquina industrial. O valor da capacidade da máquina (HP) foi coletado a diversas velocidades (rpmX100). Alguns parâmetros foram definidos na tentativa de encontrar um modelo de regressão linear adequado (definição da linha) para prever a capacidade da máquina. Analise os parâmetros e aponte o melhor conjunto (w_0, w_1); que apresenta a definição de uma linha melhor: $f(x) = w_0 + w_1x$

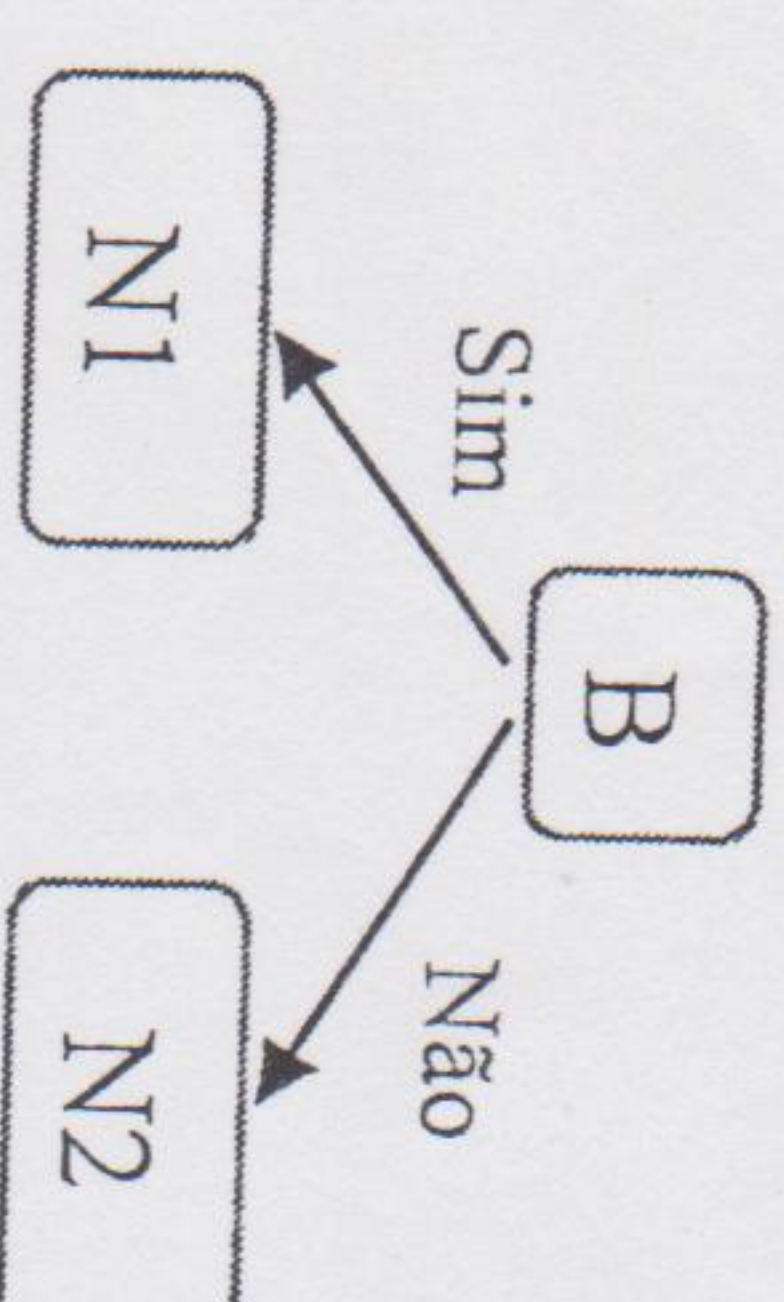
Velocidade (rpmX100)	Capacidade (HP)
22	64.03
20	62.47
18	54.94
16	48.84
14	43.73

Parâmetros:

- a. ☒ (3.456, 1.712) Erro: 31978,3
b. ☐ (5.995, 2.711) Erro: 95833,49
c. ☐ (6.144, 2.925) Erro: 112162,15

4. (1,5) Considere o conjunto de dados da questão 2 e determine a classificação obtida para o exemplo 1010 utilizando o algoritmo k-Nearest Neighbor, com $k=3$

5. (2,0) Em modelos de árvore de decisão, a definição da melhor divisão dos nodos é de fundamental importância. Sendo assim, utilize o índice Gini para verificar o grau de impureza do nodo B.



	Pai	N1	N2
C1	5	4	1
C2	5	2	3
Gini = 0.5		Gini: <u>0,2</u>	<u>0,6</u>

6. (1,5) A avaliação de desempenho de um modelo de classificação pode ser feita utilizando contagens tabuladas em uma matriz de confusão. Analise a seguinte matriz de confusão e determine a taxas de erro do modelo, precisão da classe 1, e erro da classe 2.

	Valor verdadeiro	
	Classe 1	Classe 2
Valor previsto		
Classe 1	28	3
Classe 2	5	19

ANEXO

Teorema de Bayes:

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)}$$

Distância Euclidiana:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

Distância de Manhattan:

$$d(p, q) = \sum_{i=1}^n |p_i - q_i|$$

Índice Gini:

$$Gini(t) = 1 - \sum_{i=1}^c [p(i/t)]^2$$

Ganho Δ (qualidade da divisão):

$$\Delta = I(pai) - \sum_{j=1}^k \frac{N(v_j)}{N} I(v_j)$$

Folha de respostas
Roberto Abreu Dentu

1) $P_{\text{prob}}(P|V_2V_4V_9)$

multiplicar Juros de soma

$$P_{\text{prob}} = \frac{P_n(V_2|P) \times P_n(V_4|P) \times P_n(V_9|P) \cdot P(P)}{P_n(V_2) \times P_n(V_4) \times P_n(V_9)}$$

$$P_{\text{prob}} = \frac{\left(\frac{1}{3} \times \frac{1}{3} \times \frac{1}{3}\right) \cdot \frac{3}{5}}{\frac{2}{5} \times \frac{1}{5} \times \frac{2}{5}} = \frac{1 \cdot 0,6}{1} = 0,6$$

$P_{\text{prob}}(N|V_2V_4V_9)$

$$P_{\text{prob}} = \frac{P_n(V_2|N) + P_n(V_4|N) + P_n(V_9|N)}{P_n(V_2) \times P_n(V_4) \times P_n(V_9)} \cdot P(N)$$

$$P_{\text{prob}} = \frac{\left(\frac{1}{2} + 0 + \frac{1}{2}\right) \cdot \frac{2}{5}}{\frac{2}{5} + \frac{1}{5} + \frac{2}{5}} = \frac{1 \cdot 0,4}{1} = 0,4$$

Como $0,6 > 0,4$, então a classe é **P**

2) $i = 1$

$$P(Z_1|P) = \frac{P(P|Z_1) \cdot P(Z_1)}{P(P)} = \frac{\frac{2}{3} \cdot \frac{3}{8}}{\frac{3}{8}} \approx 0,66$$

$$P(Z_1|N) = \frac{P(N|Z_1) \cdot P(Z_1)}{P(N)} = \frac{\frac{1}{3} \cdot \frac{3}{8}}{\frac{5}{8}} = 0,2$$

$i = 3$

$$P(Z_3|P) = \frac{P(P|Z_3) \cdot P(Z_3)}{P(P)} = \frac{\frac{3}{5} \cdot \frac{5}{8}}{\frac{3}{8}} = 1$$

$$P(Z_3|N) = \frac{P(N|Z_3) \cdot P(Z_3)}{P(N)} = \frac{2/5 \cdot 5/8}{5/8} = 0,4$$

$$3-a) f(x) = 3,456 + 1,712 \cdot x$$

Valores Inventados
onde é velocidade é capacidade

$$f(\text{enno}) = (22 - (3,456 + 1,712 \cdot 64,03))^2 + (20 - (3,456 + 1,712 \cdot 62,47))^2 + (18 - (3,456 + 1,712 \cdot 59,94))^2 + (16 - (3,456 + 1,712 \cdot 48,84))^2 + (14 - (3,456 + 1,712 \cdot 43,73))^2$$

$$f(\text{enno}) = 8294,72 + 8172,99 + 6322,36 + 5050,95 + 4137,28$$

$$f(\text{enno}) = 31978,3$$

$$3-b) f(\text{enno}) = (22 - (5,995 + 2,711 \cdot 64,03))^2 + (20 - (5,995 + 2,711 \cdot 62,47))^2 + (18 - (5,995 + 2,711 \cdot 54,94))^2 + (16 - (5,995 + 2,711 \cdot 48,84))^2 + (14 - (5,995 + 2,711 \cdot 43,73))^2$$

$$f(\text{enno}) = 24831,56 + 24133,98 + 19665,50 + 14981,81 + 12220,64$$

$$f(\text{enno}) = 95833,49$$

$$3-c) f(\text{enno}) = (22 - (6,144 + (2,925 \cdot 64,03)))^2 + (20 - (6,144 + (2,925 \cdot 62,47)))^2 + (18 - (6,144 + (2,925 \cdot 54,94)))^2 + (16 - (6,144 + (2,925 \cdot 48,84)))^2 + (14 - (6,144 + (2,925 \cdot 43,73)))^2$$

$$f(\text{enno}) = 29388,84 + 28516,65 + 22154,38 + 17689,26 + 14413,02$$

$$f(\text{enno}) = 112162,15$$

$$4) X_1 = \sqrt{(1-0)^2 + 0^2 + 0^2 + 0^2} = 1 \quad N \quad X_7 = \sqrt{4} = 2$$

$$X_2 = \sqrt{0^2 + 0^2 + (-1)^2 + 0^2} = 1 \quad N \quad X_8 = \sqrt{(-1)^2 + 0^2 + (-1)^2 + 0^2} = \sqrt{2}$$

$$X_3 = \sqrt{(-1)^2 + (-1)^2 + 0^2 + 1^2} = \sqrt{3}$$

$$X_4 = \sqrt{0^2 + 1^2 + 0^2 + 0^2} = 1 \quad P \quad \text{Mencione: } X_1, X_2 \text{ e } X_4$$

$$X_5 = \sqrt{(-1)^2 + 0^2 + 0^2 + 0^2} = 1 \quad \text{Classe} = N$$

$$X_6 = \sqrt{0^2 + 1^2 + 0^2 + 1^2} = \sqrt{2}$$

Folha de resposta
Roberto Albuquerque

Letura em
horizontal, não
em vertical

5) C2 (P1, P2, P3)

$$G_{int}(N1) = 1 - ((4/5)^2 + (2/5)^2) = 0,2 \rightarrow 4/6 \quad 2/6$$

$$G_{int}(N2) = 1 - ((1/5)^2 + (3/5)^2) = 0,6 \rightarrow 1/4 \quad 3/4$$

$$N1 \rightarrow 0,2 \quad \frac{6}{10} \quad \frac{4}{10}$$

$$\Delta N1 = 0,5 - \left[\left(\frac{4 \cdot 0,2}{5} \right) + \left(\frac{2 \cdot 0,2}{5} \right) \right] = 0,5 - 0,24 = 0,26,$$

$$\Delta N2 = 0,5 - \left[\left(\frac{1 \cdot 0,6}{5} \right) + \left(\frac{3 \cdot 0,6}{5} \right) \right] = 0,5 - 0,48 = 0,02,$$

=====

$$6) \text{ tx enno modelo} = \frac{3+5}{28+3+5+19} \approx 0,145$$

$$1) \text{ Precisão classe 1} = \frac{28}{28+5} \approx 0,848$$

$$\text{enno classe 2} = \frac{3}{19+3} \approx 0,142$$

-----X-----X-----X-----