



Engenharia de Atributos

Departamento de Engenharia Informática
Faculdade de Ciências e Tecnologia
Universidade de Coimbra

05, Janeiro 2023
Normal
Duração: 2:00h

Observe que:

Exame com consulta

São permitidos meios eletrónicos (computador, calculadora, etc.), porém com todas as comunicações desativadas.

Qualquer tentativa de fraude conduzirá à anulação da prova para todos os intervenientes.

1 – Considere o seguinte conjunto de variáveis categóricas adquiridas de um conjunto de 11 doentes:

Diabetes	S	S	S	S	N	N	S	S	N	N	S
Tabagismo	S	S	N	N	N	S	S	S	S	S	S
Consumo de Álcool	S	S	S	N	N	S	S	N	N	N	N
Antecedentes familiares	S	S	S	S	S	S	N	N	N	N	N
Risco Cardiovascular	A	A	A	A	B	B	B	A	A	A	A

Assumindo que S="Sim", N="Não" e que A="Alto", B="Baixo", considere que lhe é pedido que projete um preditor do risco cardiovascular usando os atributos "Diabetes", "Tabagismo", "Consumo de Álcool" e "Antecedentes Familiares". Nesse contexto, responda às seguintes questões:

- Quais são as vantagens decorrentes da redução da dimensionalidade do vetor de atributos? Que metodologias dispõe para o fazer?
- Apresente o ranking de atributos usando a metodologia Goodman Kruskal Lambda. Deverá apresentar todos os cálculos necessários.
- Apresente o ranking de atributos usando o teste Chi-quadrado. Deverá apresentar todos os cálculos necessários.

2 – Considere o conjunto de pontos (x,y) das classes C1 e C2, respetivamente:

Tabela 1: Pontos das classes C1 e C2.

Classe C1 (x,y)	Classe C2 (x,y)
P1=(4, 5)	P26=(6, 2)
P2=(3, 6)	P27=(13, 14)
P3=(4, 8)	P28=(17, 21)
P4=(8, 10)	P29=(28, 28)
P5=(6, 1)	P30=(24, 18)
P6=(18, 5)	P31=(30, 15)
P7=(8, 16)	P32=(29, 29)
P8=(3, 0)	P33=(33, 25)
P9=(18, 18)	P34=(32, 24)
P10=(19, 14)	
P11=(8, 9)	
P12=(2, 11)	
P13=(5, 4)	
P14=(8, 9)	
P15=(11, 19)	

P16=(5, 10)
P17=(12, 10)
P18=(14, 4)
P19=(4, 9)
P20=(2, 12)
P21=(28, 32)
P22=(22, 33)
P23=(22, 30)
P24=(34, 16)
P25=(15, 20)

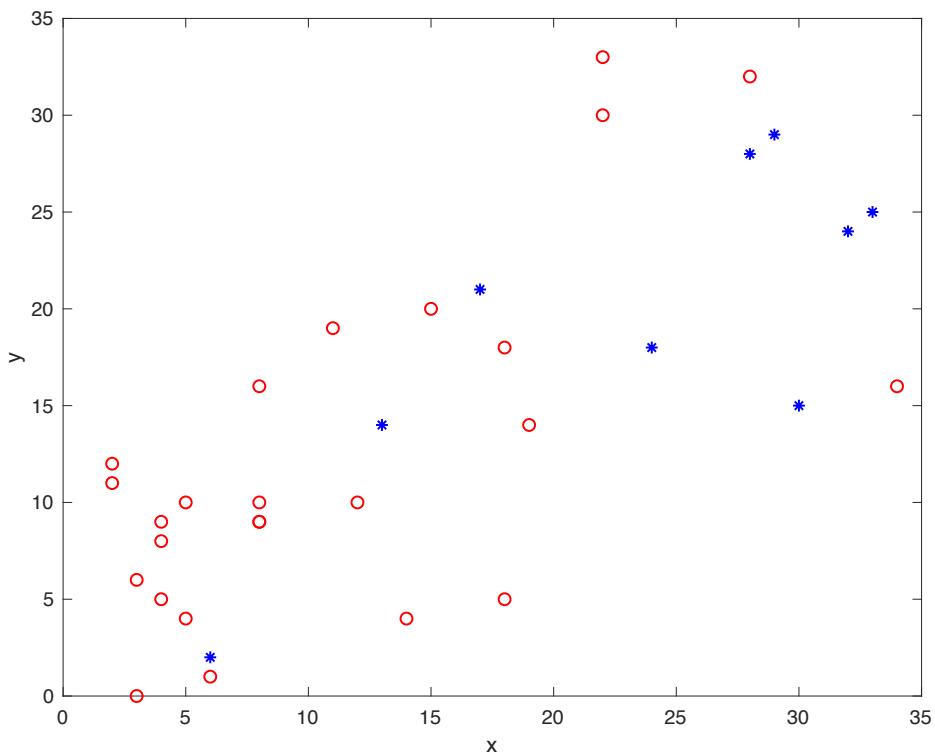


Figura 1: 'o' pontos da classe C1, '*' pontos da classe C2.

- Analise o desbalanceamento dos dados e a suas potenciais implicações.
- Usando a metodologia *Borderline SMOTE* com K=5, determine o conjunto de dados decorrentes do processamento dos pontos P24, P26 e P32. Considere que em cada iteração deverá gerar 2 pontos, sempre que aplicável. Apresente todos os cálculos necessários.
- Determine a projeção que maximiza a separabilidade das duas classes C1 e C2, usando os dados da tabela 1. Apresente todas as matrizes e indique todos os cálculos necessários.

3 - A Transformada de Fourier (TF) do sinal $x(t)$ é dada por

$$X(\omega) = \begin{cases} 0, & \omega < -20\pi \vee \omega > 20\pi \\ |20\pi - \omega|, & -20\pi < \omega < 20\pi \end{cases}$$

Considere agora o sinal $y[n]$, amostrado de $x(t)$ com uma frequência de amostragem 10Hz. Represente graficamente, em Hz, a TF de $x(t)$ e o resultado da aplicação da Discrete Time Fourier Transform (DTFT) ao sinal $y[n]$, para o domínio -20Hz a 20Hz.

4 - O sinal $x[n]$ foi amostrado de um sinal analógico $y(t)$ a cada 0.01s. Aplicou-se uma Transformada Discreta de Wavelet (DWT) de nível 3, obtendo-se os coeficientes de aproximação a_3 e os coeficientes de detalhe d_3 .

- Indique, justificando, as gamas de frequência presentes em a_3 e d_3 .
- Sabendo que queremos estudar o sinal na frequência de interesse 4Hz. Indique, justificando, qual o nível de decomposição da DWT adequado para obter essa frequência, e que tipo de coeficientes (aproximação ou detalhe) desse nível devem ser considerados?