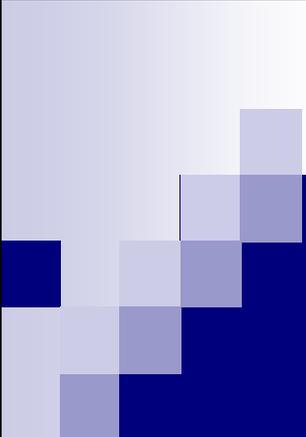


Datamining II – Exemplos de problemas para exame

V 1.0, V.Lobo, EN/ISEGI, 2013



Exercícios “tipo” para os exames

(com os protestos do) **Victor Lobo**

Questões de escolha múltipla

- **Se quisermos estimar a taxa de falhas que um dado classificador terá quando for usado em novas situações, devemos usar a taxa de erro que foi calculada para:**
 - O conjunto de treino.
 - O conjunto de validação.
 - O conjunto de teste.
 - Não é possível estimar esse valor.
- **O método das “Janelas de Parzen” serve para:**
 - Obter clusters de dados que minimizem a distância entre dados de um mesmo cluster (distância intra-cluster), e maximizem as distâncias inter-clusters.
 - Fazer previsões (classificações ou regressões) para dados novos, usando uma base de dados históricos para os quais a variável que se pretende prever é conhecida.
 - Obter clusters de dados que dividam a base de dados em “grupos naturais”.
 - Todas as respostas anteriores estão correctas.

Datamining II – Exemplos de problemas para exame

V 1.0, V.Lobo, EN/ISEGI, 2013

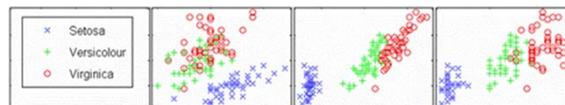
- A tabela ao lado apresenta as notas de um conjunto de alunos. Se tiver que substituir o valor em falta para a nota de Álgebra do João, usando o método do vizinho mais próximo, escolhe o valor:

	SAD	GIS.	Álgebra
Manel	18	17	16
João	14	13	?
Maria	14	15	16
Pedro	10	10	11
Roberto	15	16	16
José	14	13	15
Hildefonso	13	14	13

- a) 13 b) 14 c) 15 d) 16

- Tendo em conta a figura, que representa vários gráficos bi-variados de um mesmo conjunto de dados multidimensionais, poderemos dizer que as classes linearmente separáveis são:

- a) A classe Setosa das Classes Versicolour e Verginica;
- b) A classe Vergínica das Classes Versicolour e Setosa;
- c) A classe Versicolour das Classes Verginica e Setosa;
- d) Não existem classes linearmente separáveis;



- Um estimador naive de Bayes:

- Só faz sentido usar se as probabilidades condicionais de cada variável (ou característica) forem independentes das probabilidades condicionais das outras variáveis.
- Apesar de assumir que as probabilidades condicionais de cada variável (ou característica) são independentes das probabilidades condicionais das outras variáveis, o estimador naive de bayes pode dar boas estimativas, e deve ser usado quando a base de dados tem muitas variáveis (logo. consequência da praga da dimensionalidade, é esparsa .
- Faz previsões ingénuas pois parte do princípio que o teorema de Bayes se aplica a casos reais.
- Faz decisões “de faca” no sentido que divide o espaço de soluções em 2 partes iguais.

Datamining II – Exemplos de problemas para exame

V 1.0, V.Lobo, EN/ISEGI, 2013

Imagine que, à semelhança de um caso apresentado nas aulas, estava a trabalhar para uma agência imobiliária e queria escolher, baseado na idade e no rendimento anual dos clientes, o tipo e valor de casa que vai tentar vender. Como a imobiliária é muito recente, tem apenas dados sobre 7 clientes que compraram casas. O tipo de casa está classificado em 3 classes: casos do tipo A,B, e C. Para além disso, em cada classe há uma grande variedade de preços. Os dados existentes são os apresentados na tabela seguinte. Para além dos dados em bruto, onde as idades estão em anos, e o rendimento em Euros, são apresentados os mesmos dados normalizados por "Min/Max", isto é, normalizados para valores entre 0 e 1.

1) A certa altura aparece um cliente novo, o Sr.Sousa, que tem 58 anos e um rendimento de 30.000€ anuais. Usando a técnica do vizinho mais próximo, prevê que o Sr.Sousa vai comprar uma casa do tipo A, B, ou C ?

2) Construa uma rede neuronal, composta por um prescepção apenas, para prever o valor cada a adquirir. Inicialize os pesos sinápticos com os valores que quiser. Mostre o que acontece a esses pesos sinápticos quando a rede é treinada uma vez com o Sr.Almeida e Sr.Cadaval

Dados						
Cliente	Idade	Rendimento	Normalizados		Casa a comprar	
			I	R	Tipo	Valor
Almeida	23	20,000	0.00	0.00	C	130,000
Boavista	26	40,000	0.09	0.29	A	300,000
Cadaval	54	90,000	0.89	1.00	A	900,000
Damasio	42	70,000	0.54	0.71	A	750,000
Ernerso	33	50,000	0.29	0.43	B	200,000
Fonseca	26	30,000	0.09	0.14	B	250,000
Goncalves	58	32,000	1.00	0.17	C	200,000