Pattern Recognition
Reconhecimento de Padrões

2015/2016

Exame de Recurso 30 June 2016 Duration: 2h00

Name: **Number:** **Practical Class:**

AVISO

The Exam has a duration of 2h00m. The test is composed by five questions. The last question is a Matlab practical question. Each question must be answered in the framed box below (and following) it. Questions may be answered in Portuguese or English. This is a closed book test. You may use only 1 A4 manuscript with your 'own' notes. You are allowed to use a calculator machine. Violation of the rules ends up with exam cancellation, course failure and eventually you may be subject to disciplinary procedure. If you have any questions, you may ask. Good Luck!

Question	pts	Results	Graded by:
1)	15		
2)	15		
3)	20		
4)	20		
5)	30		

Graded by:

Name: **Number:** **Practical Class:**

Question 1 - Fisher Linear Discriminant Analysis (F-LDA)

□ **15 pts**

(a) Describe F-LDA. How is it used to discriminate patterns from two classes?

(b) Given the data available in Figure 1:

- Compute the projection vector and the bias.
- Compute the values of the discriminant function.
- Is the classification perfect for the training data?

$\mu_1 - \mu_2 = \begin{bmatrix} 0.3 \\ 0.35 \end{bmatrix}$

$\mu_1 + \mu_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

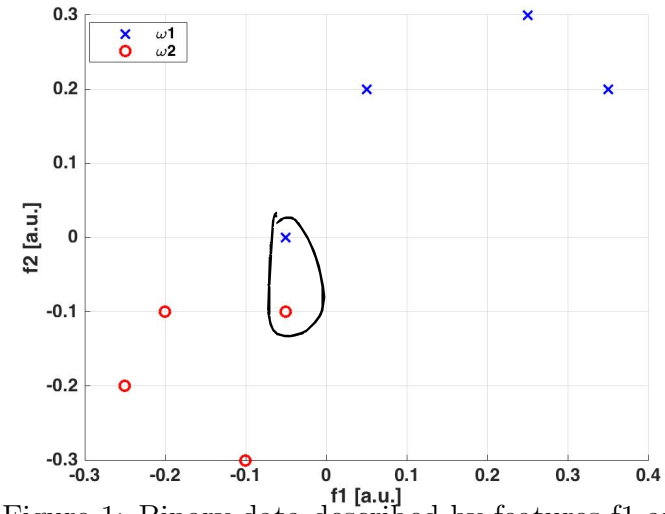


Figure 1: Binary data described by features f1 and f2

Calculus support:

Mean vectors: $\mu_1 = \begin{bmatrix} 0.150 \\ 0.175 \end{bmatrix}$ $\mu_2 = \begin{bmatrix} -0.150 \\ -0.175 \end{bmatrix}$

Between-class scatter matrix: $\mathbf{S}_b = \begin{bmatrix} 0.0900 & 0.1050 \\ 0.1050 & 0.1225 \end{bmatrix}$

Within-class scatter matrix: $\mathbf{S}_w = \begin{bmatrix} 0.1250 & 0.0500 \\ 0.0500 & 0.0750 \end{bmatrix}$ $\mathbf{S}_w^{-1} = \begin{bmatrix} 10.9091 & -7.2727 \\ -7.2727 & 18.1818 \end{bmatrix}$

Your answer 1):

a) Fisher LDA for a 2 class problem, consists in projecting data into a 1D space maximizing between class separability, while simultaneous minimizing within class variability.

b) Projection vector and bias:

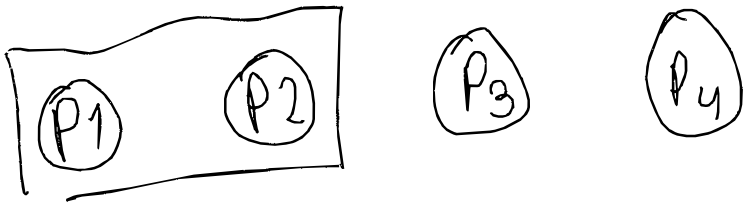
$$w = S_w^{-1}(m_1 - m_2) \quad 2\times2 \quad 2\times1$$

$$= \begin{bmatrix} 10.9091 & -7.2727 \\ -7.2727 & 18.1818 \end{bmatrix} \begin{bmatrix} +0.3 \\ +0.35 \end{bmatrix} = \begin{bmatrix} +0.7273 \\ 4.1818 \end{bmatrix}$$

$$g_1(x) = w^T x + w_0 = \begin{bmatrix} -0.7273 \\ 4.1818 \end{bmatrix}^T x + 0 \gtrless 0 \Rightarrow \text{Class 1 else Class 2}$$

$$-\frac{1}{2}(\mu_1 + \mu_2)^T w = 0$$

R: No, two samples overlap.



Question 2 - Hierarchical Clustering

□ **15pts**

a) Highlight the main advantages of hierarchical clustering in relation to k-means.

b) Consider the patterns P1, P2, P3, P4 and P5, and the inter-pattern distance represented in the following matrix:

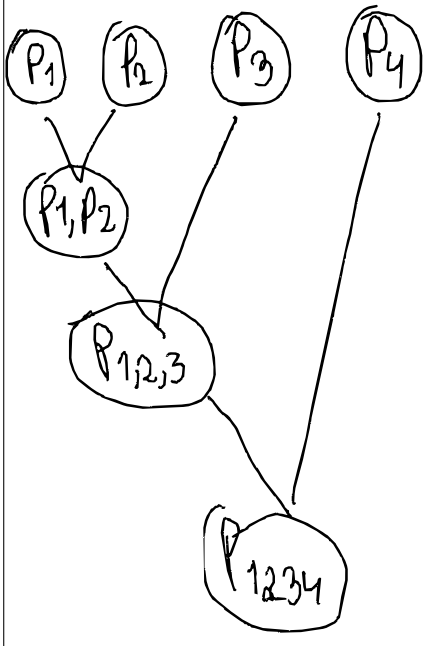
	P1	P2	P3	P4
P1	0	1	4	6
P2	-	0	2	7
P3	-	-	0	3
P4	-	-	-	0

Develop the dendrograms considering the **Single** and **Complete** approaches.

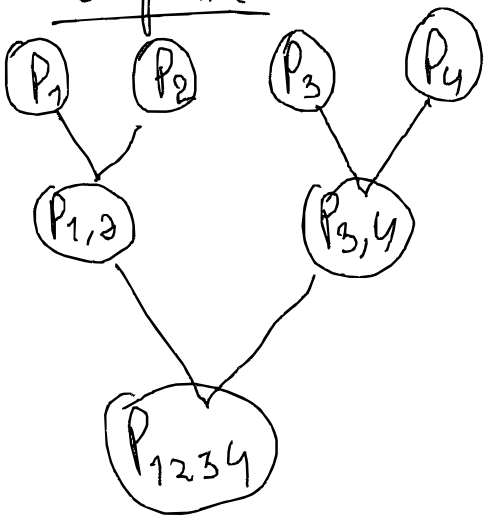
Your answer to 2):

a) Deterministic, no need to predifine the number of clusters, produces a dendogram beeing highly interpretable.

b) Single



Complete



α_1 : Classify image as Apple

α_2 : Classify image as Peach

Question 3 - Bayes Decision Theory

□ **20 pts**

Suppose that you are hired to develop a decision system that aims to discriminate images containing apples (A) from images containing peaches (P). You receive the information from the customer that classifying wrongly an A as P is two times more costly than wrongly classifying a P as A.

a) Define the loss function, the Bayes rule and the Bayes risk.

b) Explain how you incorporate the statement "*classifying wrongly an A as P is two times more costly than wrongly classifying a P as A*" in the classification process.

Your answer to 3):

a) Loss Function: $\lambda(\alpha_i \mid w_k)$

$2\,\lambda(\alpha_1 \mid w_2) = \lambda(\alpha_2 \mid w_1) \Rightarrow$

$\Rightarrow 2\,\lambda_{12} = \lambda_{21}$, For the sake of notation

$l_i = \sum_{k=1}^{k} \lambda(\alpha_i \mid w_k)\, P(w_k \mid x)$

Let $g_k(x) = f(P(w_k) \mid x)$ we can define the bayes classification rule as: $g_1(x) = g_2(x)$ so:

$\lambda_{21} P(w_1) P(x \mid w_1) = \lambda_{12} P(w_2) P(x \mid w_2)$

$\Leftrightarrow 2\lambda P(w_1) P(x \mid w_1) = \lambda P(w_2) P(x \mid w_2)$

$\Leftrightarrow 2\, P(w_1) P(x \mid w_1) = P(w_2) P(x \mid w_2)$

$\Rightarrow \frac{2P(w_1)}{P(w_2)} \underset{w_2}{\overset{w_1}{\lessgtr}} \frac{p(x \mid w_2)}{p(x \mid w_1)}$

Question 4 - Support Vector Machines

□ **20 pts**

Consider a linear SVM classifier given by $\mathbf{w} = [-1.08 \quad -7.50]^T$ and $b = 3.66$. (a) Compute the margin.

(b) Based on the following input patterns, represented by the features f1 and f2, plot the ROC curve and compute the related classifier performance.

	1	2	3	4	5	6	7	8	9	10
f1	0.30	0.37	-0.62	-0.30	0.09	0.57	-0.37	0.28	-0.82	-0.34
f2	0.40	0.76	0.07	0.85	0.40	0.91	-0.04	0.31	0.27	0.39
Class	2	2	1	2	1	2	1	1	1	2

In the previous table **"1" represents the positive class** and **"2" represents the negative class**.

Your answer 4):

a) $R = \frac{2}{\|w\|} = \frac{2}{7,58} = 0.264$

b) $d_{1,2}(x) = w^T x + b = [-1.08 \quad -7.50] \begin{bmatrix} f1 \\ f2 \end{bmatrix} + 3.66$

$= -1.08 f_1 - 7.50 f_2 + 3.66$ if ≥ 0 classify + else

calculate the score for each one of the samples:

[0.336; -2.44; 3,8; -2,391; 0.5628; -3,78; 4.3596; 1,033; 2.5206; 1.1022]

Sorted samples: [7, 3, 9, 10, 8, 5, 1, 4, 2, 6]

FPR : [0, 0, 0, 0.2, 0.2 ...]

TPR : [0.2, 0.4, 0.6, 0.6, 0.8 ...]

38

5

Cont. your answer to 4):

$$AUC = \frac{R_+ - \frac{n_+(n_+ + 1)}{2}}{n_+ n_-}$$

$$= \frac{38 - 15}{25} \overset{5}{\approx} 0.92$$