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?

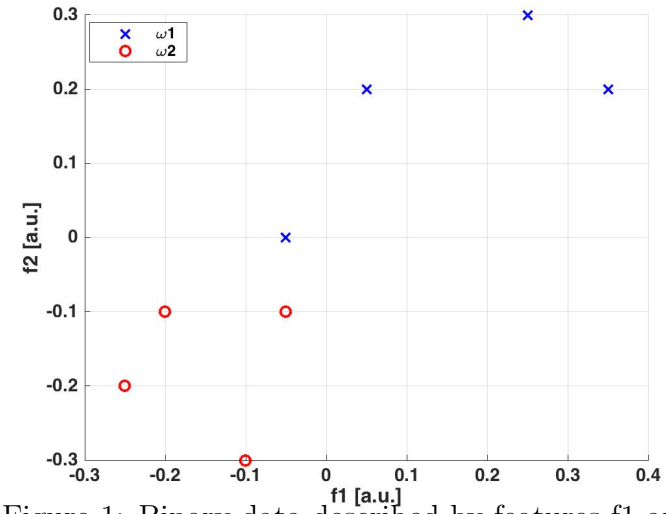

Figure 1: Binary data described by features f1 and f2

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):

Cont. your answer to 1):

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):

Cont. your answer to 2):

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):

Cont. your answer to 3):

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.

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):

Cont. your answer to 4):

Pattern Recognition
Reconhecimento de Padrões

2015/2016

PR Practical Question

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

Question 5 - kNN classifiers

□ **30 pts**

Write two Matlab functions to training and testing kNN classifiers. The functions should have the following prototypes:

- **function model=kNN_training(Xtr,Ttr,k)**

- **function [ss,sp]=kNN_testing(Xte,Tte,model)**

Where:

- **model** is a structure with fields that contain the kNN parameters needed for testing;
- **Xtr** is a matrix with dimensions D x Ptr, being D the problem dimensionality and Ptr the number of patterns in the training data;
- **Ttr** is the target vector in the training with dimensiona 1 x Ptr, and with"1" labeling positive patterns and "2" labeling negative patterns;
- **k** is the number of neighbors ;
- **Xte** is a matrix with dimensions D x Pte, being D the problem dimensionality and Pte the number of patterns in the testing data;
- **Tte** is the target vector in the testing data with dimensiona 1 x Pte, and with"1" labeling positive patterns and "2" labeling negative patterns;
- **ss** is the sensitivity on the testing data;
- **sp** is the specificity on the testing data;

Important note: You are advised to not use any built-in function from STPRtool and Matlab (e.g. knnrule.m, knnclass.m, fitcknn.m and predict.m).

Matlab Help:

IDX = knnsearch(X,Y,'K',k): Finds the k nearest neighbor in X for each point in Y by considering the Euclidean distance. X is an MX-by-N matrix and Y is an MY-by-N matrix. Rows of X and Y correspond to patterns and columns correspond to features. IDX is matrix with dimensions MY-by-k. Each row in IDX contains the indexes of the k nearest neighbor in X for the corresponding row in Y. If Y is a single pattern, IDX will be a vector with dimensions 1-by-k.

S = struct('field1',VALUES1,'field2',VALUES2,...): Creates a structure array with the specified fields and values.
For example St=struct('Vector',[3;4],'Matrix',[1 2;2 1]) creates a structure with two fields one containing a vector and other a matrix. For instance, the matrix can be obtained by doing St.Matrix.

I = find(X): Returns the linear indices corresponding to the nonzero entries of the array X. X may be a logical expression.

Your answer to 5):

Cont. your answer to 5):