Pattern Recognition/Pattern Recognition Techniques

2014/2015

Exame Recurso 26 June 2015 Duration: 2h30

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

AVISO

The Exam has a duration of 2h30m. 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 it. Questions may be answered in Portuguese or English. This is a closed book test. You are allowed to use a calculator machine. As consultation you may use only 1 Page A4 with your own notes. Violation of the last rule 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)	20		
2)	20		
3)	10		
4)	20		
5)	30		

Graded by:

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

Question 1 - Principal Component Analysis

□ **20 pts**

Consider the dataset represented in the table bellow described by feature 1 (f_1) and by feature 2 (f_2), and with class labels defined by Class (c), where "1" and "2" represents the positive and negative class, respectively.

Feature 1 (f_1)	2.5	0.5	2.2	1.9	3.1	2.3	2	1	1.5	1.1
Feature 2 (f_2)	2.4	0.7	2.9	2.2	3.0	2.7	1.6	1.1	1.6	0.9
Class (c)	1	2	1	2	1	1	2	1	2	2

1. In the context of PCA is there any unnecessary information in the table? Justify?

2. Compute the variance explained by the principal components, as well as the direction of projection. Represent them in the space spanned by f_1 and f_2.
   (Note: The eigenvalues of a given matrix $\mathbf{M}$ *can be found by solving* $|\lambda\mathbf{I} - \mathbf{M}| = 0$. *The eigenvectors are the vectors* $\mathbf{V}$ *that satisfy* $\mathbf{MV} = \lambda\mathbf{V}$. *)*

Your answer to 1):

Your answer to 2):

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

Question 2 - Minimum distance classifiers

□ **20 pts**

Find the generic discriminant function of a minimum distance classifier when the Standardized Euclidean distance is considered. Consider as class prototypes the mean and that the squared Standardized Euclidean distance is given by:

$$d^2 = (\mathbf{x} - \mathbf{m}_k)^T \mathbf{D}^{-1} (\mathbf{x} - \mathbf{m}_k), \tag{1}$$

where D is the diagonal matrix with diagonal elements given by v_j^2, which denotes the variance of the j-th feature.

Your answer:

Your answer:

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

Question 3 - kNN

□ **10pts**

1. If a data set has significant class noise, do we want to use a smaller or larger value of k? Why? Explain graphically your answer.

2. Does kNN tend to overfit with smaller values of k or with larger values of k? Why? Explain graphically your answer.

Your answer to 1):

Your answer to 2):

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

Question 4 - SVM

□ **20pts**

1. What equations are used for classification in a SVM?
2. In a linear SVM describe the influence of the free parameter? Explain graphically your answer.

Your answer to 1):

Your answer to 2):

Pattern Recognition/Pattern Recognition Techniques

2014/2015

PR/PRT Practical Question

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

Practical Question

☐ **30 pts**

Receiver Operating Characteristics (ROC)

The Excel file SignalNoise.xls contains 100 samples of signal + noise where some samples correspond to events that one wants to detect. Develop a Matlab script that by changing the value of a detection threshold develops a ROC curve and computes the threshold that originates the best performance values. Note: for this question you are not allowed to use the function *roc* from STPRTool or the function *perfcurve*).
Consider that the SignalNoise.xls has the following structure:

A	B	C	D	E	F	G	H	I	J
normal	noise (qui.square)	occurrence times (Poisson)	normal	signal (amp)	Gain	signal w/ gain	signal+noise	threshold	Detections
-0,3	0,09	4	-0,8221	0,675839	4		0,0901393	1	0
-1,28	1,632	10	0,0391	0,001532			1,6324743		1

The 100 signal + noise samples are in the column "sinal+noise" and the classification labels are in the column "Detections".
Useful Matlab Functions:

- S = load(FILENAME): Loads the variables from a MAT-file into a structure array S.
- I = find(X): Returns the linear indices I corresponding to the nonzero entries of the array X. X may be a logical expression.
- D = pdist(X): Computes the Euclidean distance between pairs of objects in m-by-n data matrix X.

Your answer: