Pattern Recognition
Reconhecimento de Padrões

2017/2018

Exame Recurso 28 June 2018 Duration: 2h30

Name: **Number:**

WARNING/AVISO

The Exam has a duration of 2h30m. The test is composed by five questions. The last question is a 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)	20		
2)	20		
3)	20		
4)	20		
5)	20		

Graded by:

Question 1 - Dimensionality Reduction & Fisher LDA

□ **20 pts**

Given the data available in Figure 1:

(a) Explain graphically which approach (PCA or LDA) you select if we aim to reduce data to one dimension, and ensuring the maximum possible separability between different classes. Justify.

(b) Compute the LDA projection vector $\mathbf{w}$. Are the direction pointed by $\mathbf{w}$ what you expected?

(c) Based on the LDA projection vector describe how you develop the Fisher LDA classifier for the data. To which class the classifier will assign the pattern $\mathbf{x} = [2 \;\; 3]^T$?

(d) Compute the decision function $d_{12}(\mathbf{x})$ and the separation hyperplane for the Fisher LDA classifier.

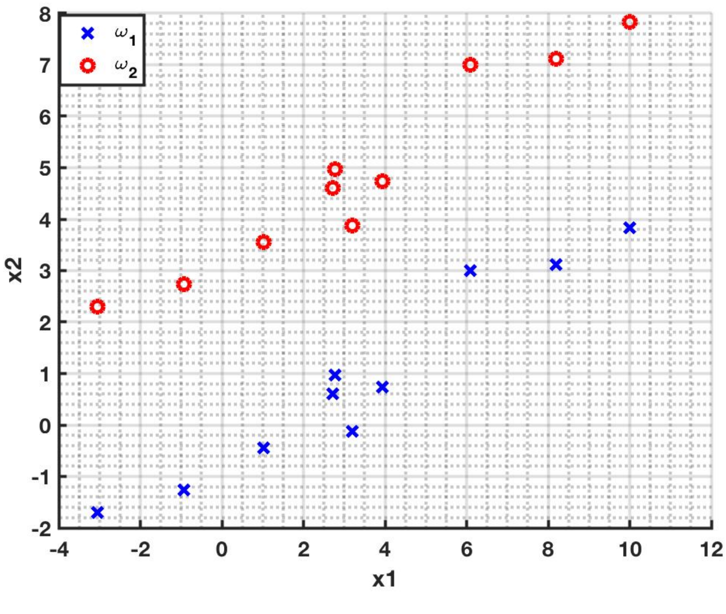


Figure 1: Binary classification problem described by features x_1 and x_2

Calculus support:

Mean vectors: $\mathbf{m}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ $\mathbf{m}_2 = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$

Within Scatter matrix: $\mathbf{S_w} = \begin{bmatrix} 281.92 & 130.32 \\ 130.32 & 64.50 \end{bmatrix}$

Inverse Within Scatter matrix: $\mathbf{S_w}^{-1} = \begin{bmatrix} 0.05 & -0.11 \\ -0.11 & 0.23 \end{bmatrix}$

Your answer 1):

a) LDA, copiem da consulta

b)

$$w = S_W^{-1}(m_1 - m_2) =$$

$$\begin{array}{cc} 2\times2 & 2\times1 \\ = \begin{bmatrix} 0.05 & -0.11 \\ -0.11 & 0.23 \end{bmatrix} & \begin{bmatrix} 0 \\ -4 \end{bmatrix} \end{array}$$

$$= \begin{bmatrix} 0.44 \\ -0.92 \end{bmatrix}$$, which doesn't surprise because is an almost vertical vector.

$w_0 =$

$$-\frac{1}{2}[4\ 4]\begin{bmatrix} 0.44 \\ -0.92 \end{bmatrix} =$$

$$= [-2\ -2]\begin{bmatrix} 0.44 \\ -0.92 \end{bmatrix} = 0.96$$

c) d)

$$g_i(x) = w^T x + w_0 \ , \ w_0 = -\frac{1}{2}(m_1+m_2)^T w$$

$$= [0.44\ -0.92]\, x + 0.96$$

$$\Downarrow x = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$\underbrace{\boxed{[0.44\ -0.92]\begin{bmatrix} 2 \\ 3 \end{bmatrix} + 0.96}}_{\to\ d_{1,2}(x)} = -0.92$$

R: Will be assigned to w_2

Question 2 - k-Nearest Neighbors (kNN)

☐ **20pts** Consider the data in Fig. 2.

b)

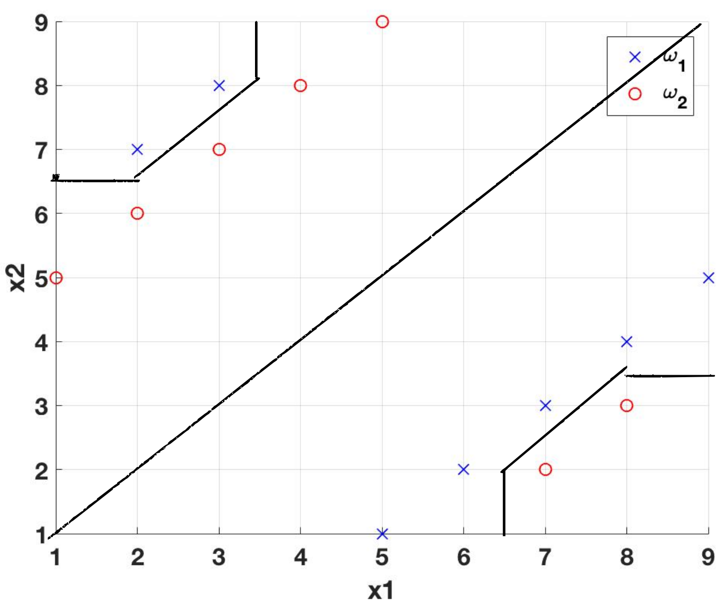


Figure 2: Binary classification problem described by features x_1 and x_2

Consider a k-nearest neighbor classifier using Euclidean distance metric.

(a) Describe how kNN works.

(b) Draw the decision boundary when k=1.

(c) What value of k minimize the training error for this dataset? What is the resulting training error?

(d) Why might too small values of k be bad?

Your answer to 2):

a) KNN works by assigning a point to the class the majority of the k neighbours have. k points are the closest points given some distance metric.

c) $k=1$, $e=0$

d) Because they lead to low training error but may cause overfitting

Question 3 - Receiver Operator Characteristics (ROC)

☐ **20 pts** Consider the values of the decision functions of two linear classifiers applied to a binary dataset. Consider the "1" class as the positive class.

$d_1(\mathbf{x})$	0.96	0.85	0.71	0.27	0.38	-0.31	-0.42	-0.56	-1.00	-0.89
$d_2(\mathbf{x})$	0.48	0.13	0.88	-0.08	0.11	-0.43	-0.79	-0.03	-1.00	-0.8061
Class	1	1	1	1	1	2	2	2	2	2

(a) What are ROC curves? How we can use them for feature ranking and for classifiers comparison?

(b) By using ROC curves, which classifier related to $d_1(\mathbf{x})$ and $d_2(\mathbf{x})$ is the best? Justify. (**Suggestion**: Consider three thresholds, e.g. -1, 0 and 1.)

Your answer to 3):

a) ROC curves are the plot the proportion of TPR vs FPR as a threshold changes. A perfect classifier would have AUC = 1 so we can leverage ROC to compare classifiers, closest to one wins. The feature ranking comes from using only that feature to plot the ROC and see its discriminative power.

$T = -1$:

$d_1 \rightarrow [1,1,1,1,1,1,1,1,1,1]$

$d_2 \rightarrow [1,1,1,1,1,1,1,1,1,1]$

	$\hat{P}$	$\hat{N}$
P	5 TP	0 FN
N	5 FP	0 TN

$TPR_{1,2} = \frac{TP}{TP+FN} = 1 \qquad FPR_{1,2} = 0 = \frac{TN}{TN+FP}$

$T = 0$:

$d_1 \rightarrow [1,1,1,1,1,2,2,2,2,2] \qquad d_2 \rightarrow [1,1,1,2,1,2,2,2,2,2]$

$TPR_1 = 1 \qquad TPR_2 = 4/5$

$FPR_1 = 1 \qquad FPR_2 = 1$

$T = 1$:

$d_1 \rightarrow [2,2,2,2,2,2,2,2,2,2] \qquad d_2 \rightarrow [2,2,2,2,2,2,2,2,2,2]$

$TPR_{1,2} = 0 \qquad FPR_{1,2} = \frac{5}{5} = 1$

R: d1 é o melhor classificador.

Question 4 - Non-Linear Support Vector Machines

□ **20 pts**

Consider the data presented in Fig. 3. The dotted circumferences indicate class limits and define R_1 and R_2.
Consider that a non-linear SVM is applied to the data. The SVM apply a transformation to the data characterized by $f(\mathbf{x}) = \begin{bmatrix} x_1^2 \\ x_2^2 \end{bmatrix}$.

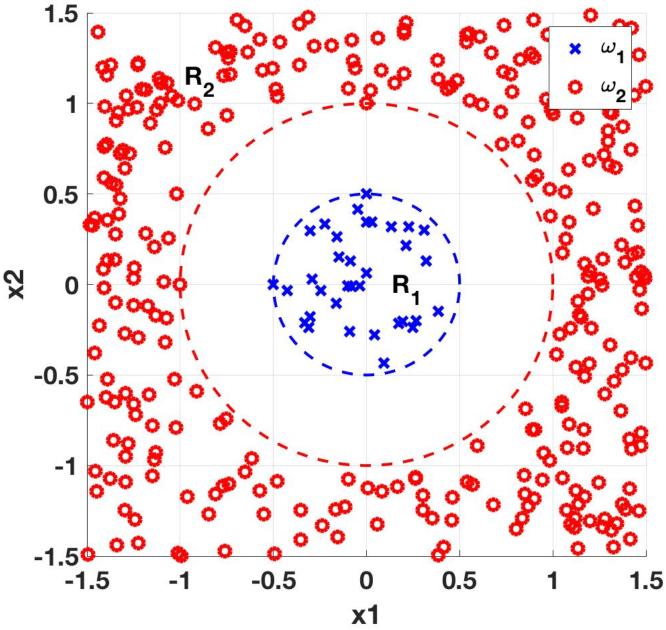


Figure 3: Binary classification problem described by features x_1 and x_2

(a) Describe how a non-linear SVM operates over the data.

(b) In this case what is the SVM kernel?

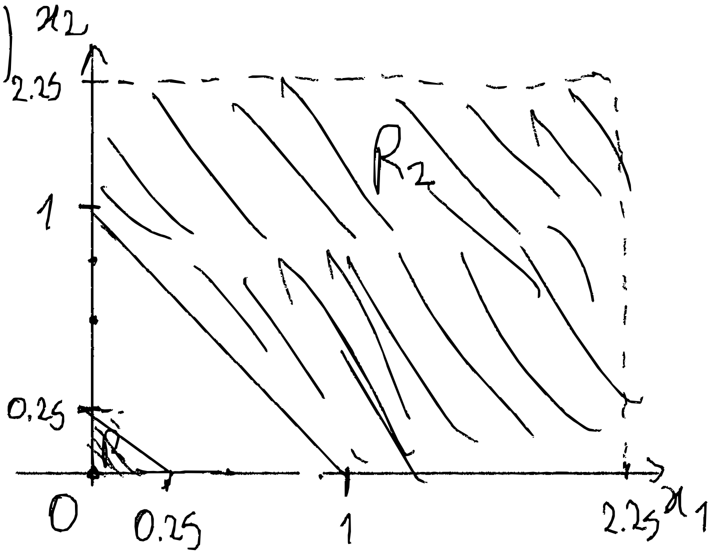
(c) Make an approximate graphical representation of R_1 and R_2 in the transformed space.

(d) Identify possible support vectors and develop the decision function $d_{12}(\mathbf{x}) = \mathbf{w}^T f(\mathbf{x}) + b$. To which class belongs the pattern $\mathbf{x} = [0 \quad -0.9]^T$

Your answer to 4):

a) A non-linear SVM maps the data to a high dimensional space where it's easier to classify with linear decision surfaces.

b) Polynomial

c) 

$x_2 = -x_1$

d) Support vectors: $\begin{bmatrix} -1 \\ 0 \end{bmatrix} \rightarrow w_2 \begin{bmatrix} -0.5 \\ 0 \end{bmatrix}$ $\quad b = \frac{1+0.25}{2} = 0.65$

$d_{1,2}(x) = w^T f(x) + b =$

$= x_1^2 + x_2^2 - 0.65$