



Pattern Recognition Reconhecimento de Padrões

2017/2018

Exame Normal 18 June 2018 Duration: 2h30

Name:

Number:

Practical Class:

WARNING/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 (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)	25		
5)	15		

Graded by:

Question 1 - Minimum Distance Classifiers (MDC)

□ 20 pts

(a) Describe how a MDC works, in which consists its training, and given new patterns how they are labeled.

(b) Given the data available in Figure 1:

- Explain why a Euclidean MDC is unable to classify data with zero error? Demonstrate graphically.
- Develop a Mahalanobis MDC for the data. To which class the classifier will assign the pattern $\mathbf{x} = [-2 \ 0]^T$?
- Compute the decision boundary $d_{12}(\mathbf{x})$ and the separation hyperplane for the Mahalanobis MDC. Are the hyperplane what you expect? Justify.

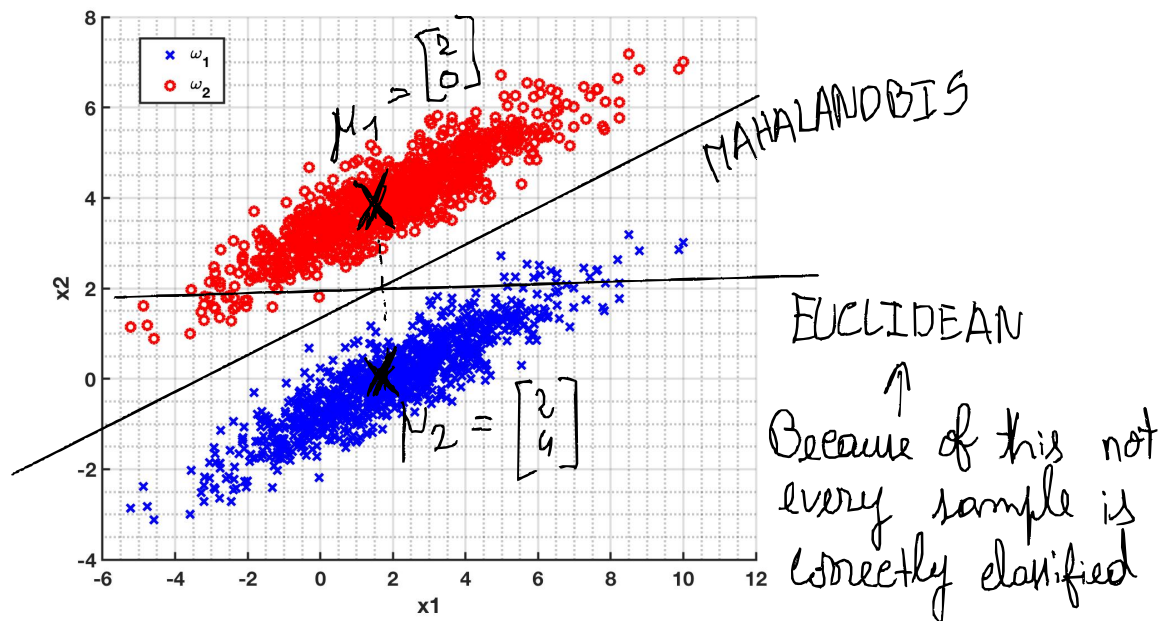


Figure 1: Binary classification problem described by features x_1 and x_2

$$\mu_1 - \mu_2 = \begin{bmatrix} 0 \\ -4 \end{bmatrix} \quad \mu_1 + \mu_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$$

Calculus support:

$$\text{Mean vectors: } \mu_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \quad \mu_2 = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

$$\text{Covariance matrices: } \mathbf{C} = \mathbf{C}_1 = \mathbf{C}_2 = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix}$$

$$\text{Inverse covariance matrix: } \mathbf{C}^{-1} = \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix}$$

$$g_1\left(\begin{bmatrix} -2 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 2 & -4 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} - 2 = -4 - 2 = -6$$

$$g_2\left(\begin{bmatrix} -2 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} -6 & 16 \end{bmatrix} \begin{bmatrix} -2 \\ 0 \end{bmatrix} - 26 = 12 - 26 = -14$$

Your answer 1):

a) In a MDC, each new sample is assigned to the class of the nearest prototype. In training, we compute the prototypes (often the mean vector of the training set).

$$x = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

b)

$$g_k(x) = m_k^T \underset{1 \times 2}{\mathbf{C}^{-1}} \underset{2 \times 1}{x} - 0.5 m_k^T \underset{2 \times 2}{\mathbf{C}^{-1}} \underset{2 \times 1}{m_k}$$

$$\begin{aligned} g_1(x) &= \begin{bmatrix} 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix} x - 0.5 \begin{bmatrix} 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \\ &= \begin{bmatrix} 2 & -4 \end{bmatrix} x - \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 & -4 \end{bmatrix} x - \begin{bmatrix} 1 & -2 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 2 & -4 \end{bmatrix} x - 2 \end{aligned}$$

$$\begin{aligned} g_2(x) &= \begin{bmatrix} 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix} x - 0.5 \begin{bmatrix} 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \\ &= \begin{bmatrix} -6 & 16 \end{bmatrix} x - \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \\ &= \begin{bmatrix} -6 & 16 \end{bmatrix} x - \begin{bmatrix} -3 & 8 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} -6 & 16 \end{bmatrix} x - 26 \end{aligned}$$

Cont. your answer to 1):

$$\begin{aligned}d_{12}(x) &= (m_1 - m_2)^T C^{-1} [x - 0.5(m_1 + m_2)] \\&= \begin{bmatrix} 0 \\ -4 \end{bmatrix}^T \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix} [x - \begin{bmatrix} 2 \\ 2 \end{bmatrix}] \\&\quad \begin{matrix} 1 \times 2 & 2 \times 2 \end{matrix} \\&= \begin{bmatrix} 0 & -4 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix} [x - \begin{bmatrix} 2 & 2 \end{bmatrix}^T] \\&= \begin{bmatrix} 8 & -20 \end{bmatrix} x - \begin{bmatrix} 8 & -20 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} \\&= \begin{bmatrix} 8 & -20 \end{bmatrix} x - (16 - 40) = \begin{bmatrix} 8 & -20 \end{bmatrix} x + 24\end{aligned}$$

the hyperplane is $d_{12}(x) = 0$

$$\begin{bmatrix} 8 & -20 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 24 = 0 \Rightarrow 8x_1 - 20x_2 + 24 = 0$$

↑
Linear Hyperplane
which makes sense
since $C_1 = C_2$

$$\textcircled{P_{1,2}} \rightarrow \textcircled{P_3}$$

$$16 + 4 = 20$$

$$\textcircled{P_3} \xrightarrow{9} \textcircled{P_4}$$

Question 2 - Hierarchical Clustering

□ 20pts

a) b) Hierarchical clustering algorithms outputs a dendrogram. Describe what is a dendrogram?

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	5
P2	-	0	2	6
P3	-	-	0	3
P4	-	-	-	0

Develop the dendrogram considering the **Ward's** approach (Note: Consider the Lance-Williams formula).

Your answer to 2):

a) A dendrogram is a visual representation of the clusters' hierarchy.

$$b) d_{(i \cup j)k} = \alpha_i d_{ik} + \alpha_j d_{jk} + \beta d_{ij} + \gamma |d_{ik} - d_{jk}|$$

	P1	P2	P3	P4
P1	0	1	4	5
P2		0	2	6
P3			0	3
P4				0

$$\alpha_i = \frac{n_i + n_k}{n_i + n_j + n_k}$$

$$\alpha_j = \frac{n_j + n_k}{n_i + n_j + n_k}$$

$$\beta = -\frac{n_k}{n_i + n_j + n_k}$$

$n_l \equiv$ number of elements in cluster

Cont. your answer to 2):

$$d_{(i \cup j)k} = \alpha_i d_{ik} + \alpha_j d_{jk} + \beta d_{ij} + \gamma |d_{ik} - d_{jk}|$$

	P_1	P_2	P_3	P_4
P_1	0	1	4	5
P_2		0	2	6
P_3			0	3
P_4				0

$$\alpha_i = \frac{n_i + n_k}{n_i + n_j + n_k}$$

$$\alpha_j = \frac{n_j + n_k}{n_i + n_j + n_k}$$

$$\beta = -\frac{n_k}{n_i + n_j + n_k}$$

$n_i \equiv$ number of elements in cluster

Smallest distance is $d_{12}=1$ merge and update

$$n_{C_1} = 2 \quad n_{P_3} = 1 \quad n_{P_4} = 1$$

$$d_{C_1 P_3} = \frac{2}{3} \times 4 + \frac{2}{3} \times 2 - \frac{1}{3} = 3,67$$

$$d_{C_1 P_4} = \frac{2 \times 5}{3} + \frac{2}{3} \times 6 - \frac{1}{3} = 7$$

	C_1	P_3	P_4
C_1	0	3,67	7
P_3		0	3
P_4			0

(...)

Question 3 - Multivariate Normal Bayesian Classification

- **20 pts** Consider the conditional probability density functions in Figure 2 that are generated by a multivariate Gaussian process. The mean vectors for both classes are: $\mu_1 = [-1 \ -1]^T$ and $\mu_2 = [1 \ 1]^T$. The covariance matrix is the identity matrix for both classes. Consider also that patterns from ω_1 and ω_2 appears in equal prevalence.

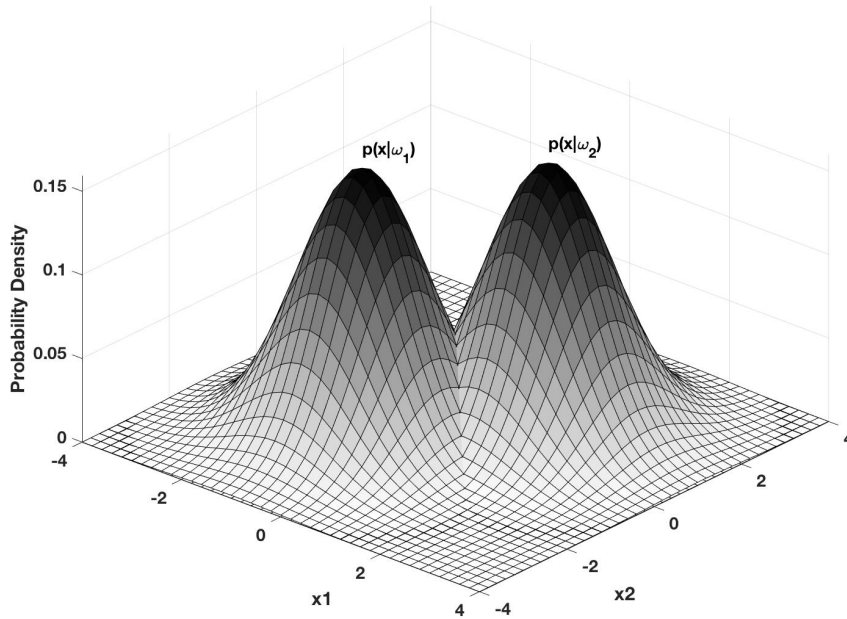


Figure 2: Conditional probability density functions defined in a space spanned by features x_1 and x_2 .

- What is the problem dimensionality? Justify.
- Starting with the Bayes rule develop the Likelihood Ratio Test (LRT).
- Develop the generic decision function of the form $d_{12}(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0$. Specify all the steps.
- Specify $d_{12}(\mathbf{x})$ for the concrete values of this exercise. What is the relation between this classifier and an Euclidean MDC? Justify.

$$\mu_1 = [-1 \ -1]^T \quad \mu_2 = [1 \ 1]^T \quad \begin{bmatrix} -1 \\ -1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$

Your answer to 3):

a) 2 because there are two features.

b) The decision surface for a Bayesian classifier can be written as:

$$P(w_1|x) > P(w_2|x) \text{ then } w_1 \text{ else } w_2$$

$$\Rightarrow \frac{P(w_1) P(x|w_1)}{\cancel{P(x)} > 0} > \frac{P(w_2) P(x|w_2)}{\cancel{P(x)} > 0}$$

$$\Rightarrow \frac{P(w_1)}{P(w_2)} > \frac{P(x|w_2)}{P(x|w_1)} = \Lambda(x)$$

$$c) w = C^{-1} (m_1 - m_2) = \begin{matrix} 2 \times 2 & 2 \times 1 \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \begin{bmatrix} -2 \\ -2 \end{bmatrix} \end{matrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$

$$w_0 = -0.5 (m_1^T C^{-1} m_1 - m_2^T C^{-1} m_2)$$

$$= -0.5 \left(\begin{matrix} 1 \times 2 & 2 \times 2 \\ \begin{bmatrix} -1 \\ -1 \end{bmatrix} & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \begin{bmatrix} -1 & -1 \end{bmatrix} \end{matrix} - \begin{matrix} 1 \times 2 & 2 \times 2 \\ \begin{bmatrix} 1 \\ 1 \end{bmatrix} & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 1 \end{bmatrix} \end{matrix} \right)$$

$$= -0.5 \left(\begin{matrix} 1 \times 2 & 2 \times 1 \\ \begin{bmatrix} -1 & -1 \end{bmatrix} & \begin{bmatrix} -1 \\ -1 \end{bmatrix} \end{matrix} - \begin{matrix} 1 \times 2 & 2 \times 1 \\ \begin{bmatrix} 1 & 1 \end{bmatrix} & \begin{bmatrix} 1 \\ 1 \end{bmatrix} \end{matrix} \right)$$

$$= -0.5 (2 - 2) = 0$$

$$d_{12}(x) = [-2 \ -2]^T x$$

↳ Igual a euclidean MDC

Question 4 - Multiclass Support Vector Machines

□ 25 pts

Consider a three class linear SVM classification strategy based on the One-vs-One strategy, implemented by considering the error-correcting output codes (ECOC) approach. The parameters \mathbf{w} and b for each one of the classifiers are:

Classifier	Problem	\mathbf{w}	b
SVM1	$\omega_1 (+)$ vs $\omega_2 (-)$	$[-1 \ 1]^T$	1
SVM2	$\omega_1 (+)$ vs $\omega_3 (-)$	$[-1 \ -1]^T$	7
SVM3	$\omega_2 (+)$ vs $\omega_3 (-)$	$[0 \ -1]^T$	2

- (a) Compute the separation margin for SVM1.
 (b) What is the coding design matrix? If training One-vs-All was selected what should be the coding design matrix?
 (c) Is $\mathbf{x} = [6 \ 1]^T$ a support vector? If yes, what is its class? Justify your answers.
 (d) What is the accuracy if the following testing patterns $\mathbf{x} = [x_1 \ x_2]^T$ are considered?
 Note: Consider the Hamming distance in the decoding phase.

x_1	2	3	5
x_2	0	0	5
True class(ω_k)	1	2	3

Your answer 4):

$$a) R = \frac{2}{\|\mathbf{w}\|} = \frac{2}{\sqrt{2}} = \frac{2\sqrt{2}}{2} = \sqrt{2}$$

b) $\omega_1 \ \omega_2 \ \omega_3$

SVM1	1	-1	-1
SVM2	-1	1	-1
SVM3	-1	-1	1

← OAA

OAO

$\omega_1 \ \omega_2 \ \omega_3$

SVM1	1	-1	0
SVM2	1	0	-1
SVM3	0	1	-1

Cont. your answer to 4):

c) $x = [6 \ 1]^T$, support vector iff $|w^T x_i + b| = 1$

SVM 1:

$$\left| \overset{1 \times 2}{[-1 \ 1]} \overset{2 \times 1}{\begin{bmatrix} 6 \\ 1 \end{bmatrix}} + 1 \right| = 1 \Rightarrow |-4| = 1, \text{ não é s.v.}$$

SVM 2:

$$\left| [-1 \ -1] \begin{bmatrix} 6 \\ 1 \end{bmatrix} + 7 \right| = 1 \Rightarrow |0| = 1, \text{ não}$$

SVM 3:

$$\left| [0 \ -1] \begin{bmatrix} 6 \\ 1 \end{bmatrix} + 2 \right| = 1 \Rightarrow |1| = 1, \text{ é}$$

R: É support vector da classe 2.

d) $\begin{bmatrix} 2 \\ 0 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \end{bmatrix}$

For $x = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ and equivalent for the rest:

$$g_1(x) = -1, -1 \quad g_2(x) = 5, +1 \quad g_3(x) = 2, +1$$

Output code: $(-1, +1, +1)$

Hamming distance 0 with the code for w_2
Classify as w_2 , missed!

$\begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix}$