Aprendizagem Computacional em Biologia
Inteligência Geoespacial
Reconhecimento de Padrões

2023/2024

Exame Recurso 27 June 2024 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)	25		
3)	20		
4)	20		
5)	15		

Graded by:

Question 1 - Dimensionality Reduction

□ **20 pts**

a) Highlight the main differences between feature selection and feature transformation by PCA/LDA?

b) Highlight the LDA advantages and disadvantages?

c) Consider the data in Figure 1, and two projection vectors related to the most important projection directions obtained with PCA and LDA:

(a) $w_1 = [0.08 \quad -1.00]^T$

(b) $w_2 = [0.41 \quad 0.91]^T$

Identify the vector associate with PCA and LDA. Justify your selection.

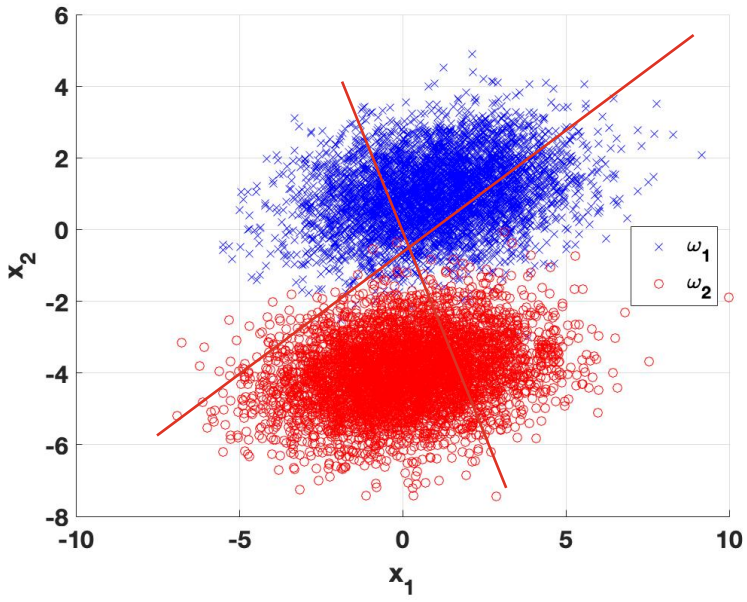

Figure 1: Binary classification problem described by features x1 and x2

Your answer to 1):

Cont. your answer to 1):

Question 2 - Minimum Distance Classifiers

□ **25 pts**

(a) Describe the main difference between the Euclidean and Mahalanobis minimum distance classifiers.

(b) Given the data available in Figure 2, define a Mahalanobis minimum distance classifier, knowing that all classes have the same covariance matrix and that it is $\mathbf{C} = \frac{1}{2}\mathbf{I}$, being $\mathbf{I}$ the identity matrix. To which class the sample $\mathbf{x} = [1 \;\; 1]^T$ bellongs to? Justify with calculations.

(c) Develop the decision functions and draw the decision hyperplanes.

(d) Can the classifier in (b) be related to the multiclass *one-vs-all* strategy? Justify.

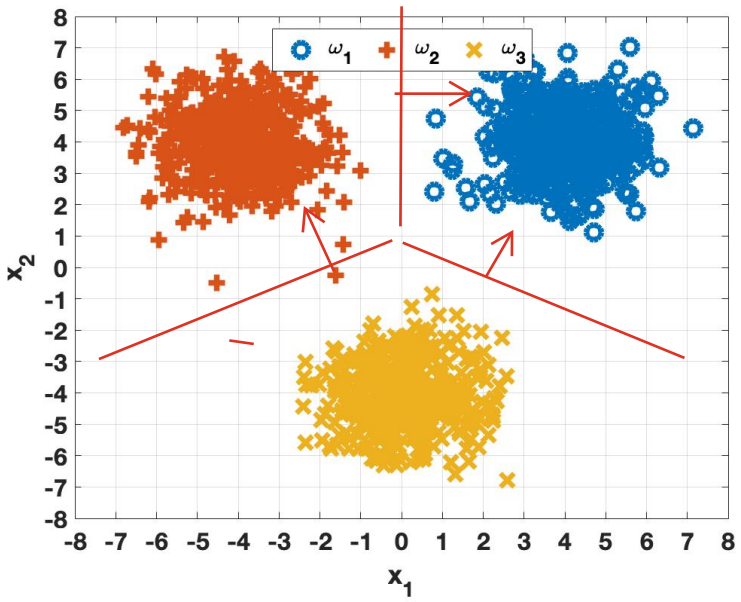

Figure 2: Multiclass classification problem described by features x1 and x2

Calculus support:

Mean vectors: $\mathbf{m}_1 = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$ $\mathbf{m}_2 = \begin{bmatrix} -4 \\ 4 \end{bmatrix}$ $\mathbf{m}_3 = \begin{bmatrix} 0 \\ -4 \end{bmatrix}$

Covariance matrices:

$$\mathbf{C}\omega_1 = \mathbf{C}\omega_2 = \mathbf{C}\omega_3 = \begin{bmatrix} 0.5 & 0.0 \\ 0.0 & 0.5 \end{bmatrix}$$

Your answer 2):

Cont. your answer to 2):

Cont. your answer to 2):

Question 3 - kNN

□ **20 pts** Consider the multiclass data in Figure 3.

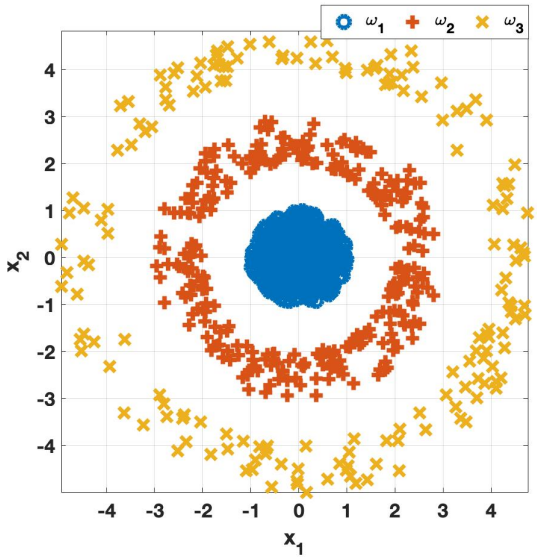

Figure 3: Multiclass data in a space spanned by features x1 and x2.

a) Is the data linearly separable? Justify.
b) kNN is a lazy learner; will it be able to classify the data correctly? Discuss the implications of the k value for this dataset.
c) Draw approximately the decision surfaces for k=1 and k=3.

Your answer to 3):

Cont. your answer to 3):

Cont. your answer to 3):

Question 4 - Ensemble Classifiers

□ **20 pts** Consider an AdaBoost classifier trained for a binary problem: the positive class is ω_1 (+1), and the negative class ω_2 (-1). Consider that each weak learner is a binary tree and that each sample is represented by a feature vector $\mathbf{x} = [x_1 \;\; x_2]^T$. The training results are listed in the following table:

α_i	0.76	0.75	0.63
Weak Learner ($H_j(\mathbf{x})$)	$\begin{cases} \omega_1 & x_2 \geq 6.00 \\ \omega_2 & x_2 < 6.00 \end{cases}$	$\begin{cases} \omega_2 & x_1 \geq 7.71 \\ \omega_1 & x_1 < 7.71 \end{cases}$	$\begin{cases} \omega_1 & x_2 \geq 9.00 \\ \omega_2 & x_2 < 9.00 \end{cases}$

(a) Explain the difference between Bagging and Boosting?
(b) What is the generic decision function for the given Adaboost classifier?
(c) To which class the sample $\mathbf{x} = [6 \;\; 1]^T$ belongs? Write out all the steps.

Your answer 4):

Cont. your answer to 4):

Cont. your answer to 4):

Question 5 - SVM Optimization

□ **15 pts** Write a function to search for the best C of a linear SVM classifier. The function should perform a grid search for the appropriate parameter value for a given dataset. You can choose your language of choice. In Matlab the function should have the following prototype:

- **function [MeanF1,StdF1]=optimize_svm(Xtr,Ttr,Cs,n_runs)**

Where:

- **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 with dimensiona 1 x Ptr, and with"1" labeling positive patterns and "2" labeling negative patterns;
- **Cs** is a vector containing the C values to be considered.
- **n_runs** defines the number of runs. This parameter defines the number of times that grid-search should be performed aiming to have appropriate statistics about the influence of the different parameter combinations.
- **MeanF1** is a vector containing the average F1 measure values for each element in Cs.
- **StdF1** is a vector containing the standard deviation of the F1 measure values for each element in Cs.

Your answer to 5):

Cont. your answer to 5):