Aprendizagem Computacional em Biologia
Inteligência Geoespacial
Reconhecimento de Padrões

2022/2023

Exame Recurso 28 June 2023 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)	25		
5)	15		

Graded by:

Question 1 - Feature Selection

□ 20pts

Figure 1 represent scatter plots of three best features to solve a given problem. The three best features were selected using AUC.

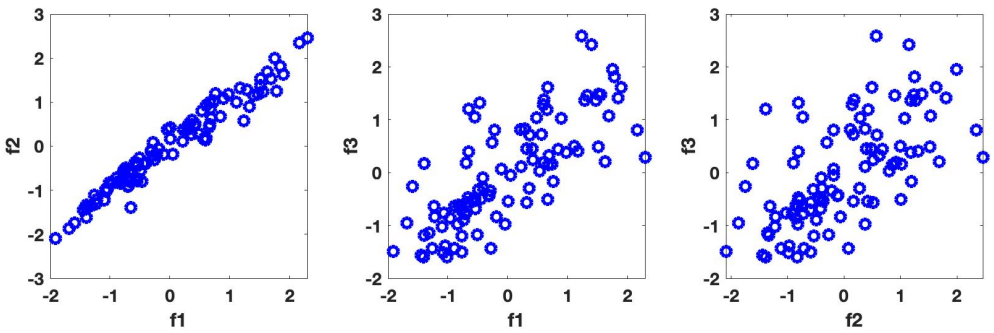

Figure 1: Data Distributions

a) Describe why and how ROC curves can be used for feature selection?

b) If a given feature present an AUC of 1, what this mean?

c) Based on Figure 1, and assuming that features have approximately the same AUC, which best two features would you keep? Justify.

Your answer to 1):

Cont. your answer to 1):

Question 2 - Multiclass Linear Classifiers

□ **20 pts**

(a) Describe the most common strategies to develop linear classifiers for multiclass problems.

(b)Consider the classification pipeline presented in Figure 2. The LDA transformation matrix is on calculus support bellow, as well as the mean vectors in the original space.

- What is the dimensionality of the original data? And what are the number of classes? Justify.
- Develop a simple and straitforward classifier, that receive LDA transformed data ($\mathbf{x}'$) and classify it. To which class the classifier will assign the sample $\mathbf{x} = [2 \;\; 2 \;\; 3]^T$?
- Compute the decision boundary $d_{12}(\mathbf{x})$, where $\mathbf{x}$ is a feature vector in the original space.

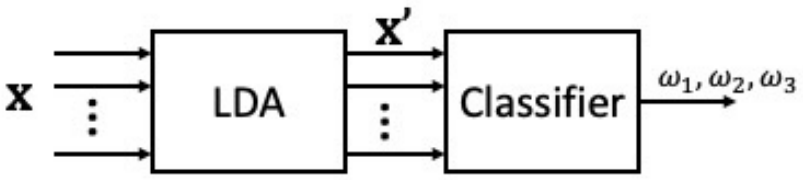

Figure 2: Multiclass Linear Classification Pipeline

Calculus support:

Mean vectors: $\mu_1 = \begin{bmatrix} -0.93 \\ -0.96 \\ -0.89 \end{bmatrix}$ $\mu_2 = \begin{bmatrix} -0.26 \\ -0.19 \\ -0.27 \end{bmatrix}$ $\mu_3 = \begin{bmatrix} 1.19 \\ 1.14 \\ 1.16 \end{bmatrix}$

LDA transformation Matrix: $\mathbf{W} = \begin{bmatrix} 0.26 & -0.75 \\ -0.77 & 0.65 \\ -0.58 & 0.12 \end{bmatrix}$

Your answer 2):

Cont. your answer to 2):

Cont. your answer to 2):

Question 3 - kNN Classification

□ **20 pts** Consider the dataset in Figure 3

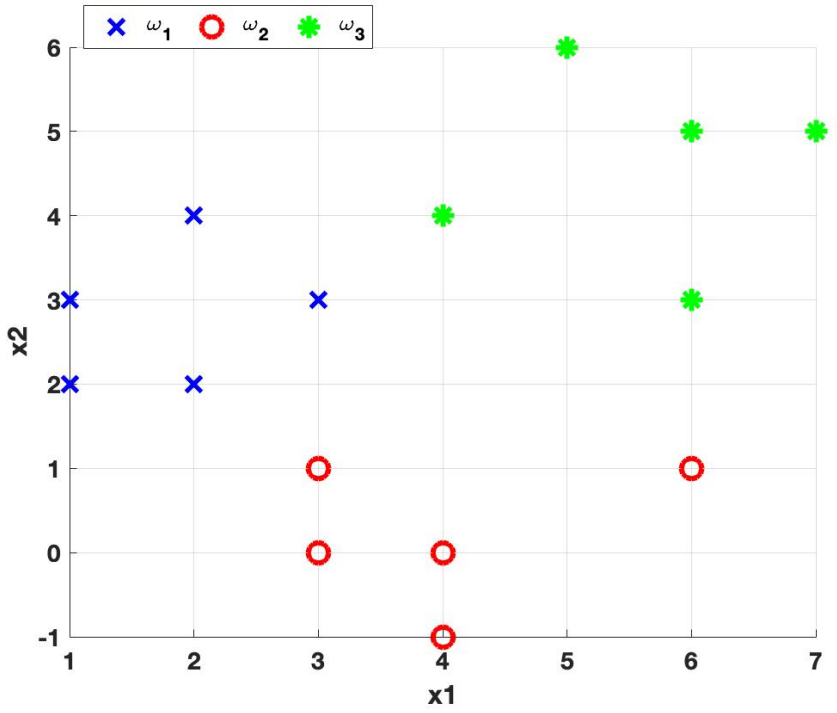

Figure 3: Sample distribution in a space spanned by features x1 and x2.

a) Why kNN is considered a Lazy Learner?
b)Train a kNN for the data. Describe the algorithm.
c) Schetch the decision boundaries for k=3. What is the training accuracy?
d) What is the accuracy if the following testing samples $\mathbf{x} = [x1 \;\; x2]^T$ are considered? Consider k=1 and k=3.

$x1$	3.5	3.0	4
$x2$	4.0	1.8	5
True class(ω_k)	1	2	3

Your answer to 3):

Cont. your answer to 3):

Cont. your answer to 3):

Question 4 - Multiclass Support Vector Machines

□ **25 pts**

Consider a three class linear SVM classification strategy based on the One-vs-All 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	ω_1 (+) vs All (-)	$[-3.0 \quad 1.0]^T$	7.0
SVM2	ω_2 (+) vs All (-)	$[0.4 \quad -1.6]^T$	1.4
SVM3	ω_3 (+) vs All (-)	$[1.0 \quad 1.3]^T$	-8.3

(a) Compute the separation margin for SVM3.
(b) What is the coding design matrix? If training One-vs-One was selected what should be the coding design matrix?
(c) Is $\mathbf{x} = [3 \quad 1]^T$ a support vector? If yes, what is its class? Justify your answers.
(d) What is the class of sample $\mathbf{x} = [4 \quad 2]^T$? Is this a special case? Justify?
Note: Consider the Hamming distance in the decoding phase.

Your answer 4):

Cont. your answer to 4):

Cont. your answer to 4):

Question 5 - Bayes classifiers

□ **15 pts**

Write two functions to training and testing Bayes classifiers. You can choose your language of choice. In Matlab the functions should have the following prototypes:

- **function model=bayes_training(Xtr,Ttr)**

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

Where:

- **model** is a structure with fields that contain the Bayes classifier 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;
- **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;
- **out** is the classifier output in the testing data with dimensiona 1 x Pte, and with "1" classifying positive patterns and "2" classifying negative patterns;
- **ss** is the sensitivity on the testing data;
- **sp** is the specificity on the testing data;

Important note: Your code should list all the important steps in Bayes classifier training and testing.

Your answer to 5):

Cont. your answer to 5):