Pattern Recognition
Reconhecimento de Padrões

2018/2019

Exame Normal 19 June 2019 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 - Minimum Distance Classifiers

□ **20 pts** Consider the fruits dataset composed of 58 images: 22 apples and 36 peaches. Two features were extracted: the shape ratio computed by dividing the fruit width by the fruit height; and the color ratio computed by dividing the red intensity by the green intensity for a region coincident with the center of the fruit. Fig.1 plots the image patterns according to shape ratio and color ratio. The sample statistics are:

- Mean Vector for apples is $\mathbf{m}_{\text{Apples}} = [0.97\ 0.90]^T$
- Mean Vector for peaches is $\mathbf{m}_{\text{Peaches}} = [1.44\ 1.02]^T$

Consider the Euclidean maximum distance classifier.

1. Write the linear decision function, $d(\mathbf{x})$.
2. Define the decision hyperplane, prove that the hyperplane is normal to the linear segment that links the points defined by the two mean vectors.
3. To which class belongs a fruit with a color ratio of 1 and a shape ratio of 1.

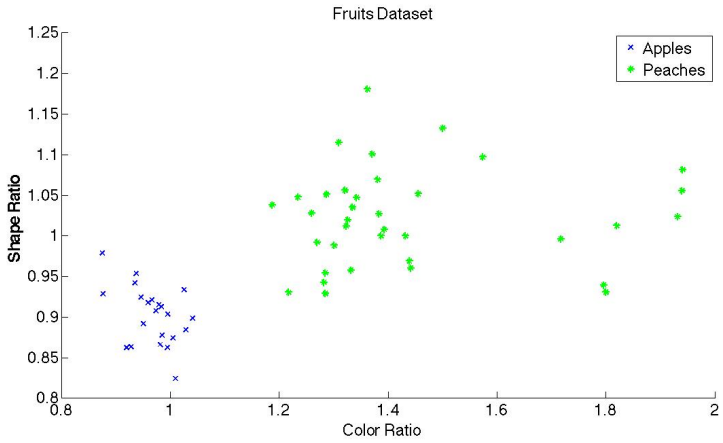


Figure 1: Fruits Dataset

Your answer to 1):

Cont. your answer to 1):

Question 2 - Non-Linear Support Vector Machines

□ **20 pts** Consider the data in Fig. 2. Suppose that a non-linear SVM is applied to the data. The SVM applies a transformation to the data characterized by $f(\mathbf{x}) = \begin{bmatrix} x_1^2 \\ x_2 \end{bmatrix}$.

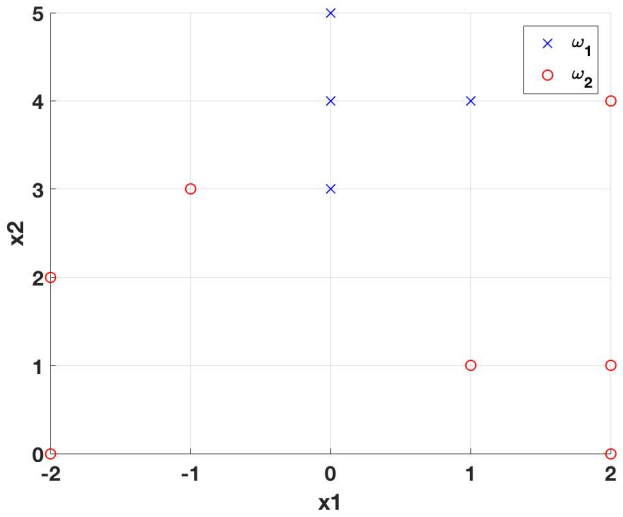


Figure 2: Binary classification problem described by features x_1 and x_2

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

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

(c) Identify possible support vectors in the transformed space and develop the decision function $d_{12}(\mathbf{x}) = \mathbf{w}^T f(\mathbf{x}) + b$. Sketch $d_{12}(\mathbf{x})$ in the original space.

Your answer to 2):

Cont. your answer to 2):

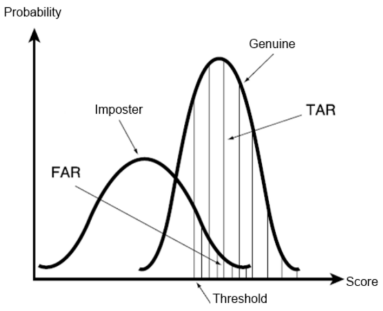
Name: **Number:**

Question 3 - Bayes Decision

□ **20pts** Suppose you have an image I and want to decide whether it is a Genuine G or an Imposter F. Figure 1 shows examples of the images and their probability distributions.

(a)

(b)



(c)

Figure 3: (a) Genuine (b) Imposter (c) Probability distribution

(a) Formulate this as a decision problem in terms of distributions $P(I|\omega)$ & $P(\omega)$, where $\omega \in \{G, F\}$.

(b) Define the loss function, the risk, the Bayes rule and the Bayes risk. Write down the loss function if all errors are weighted equally.

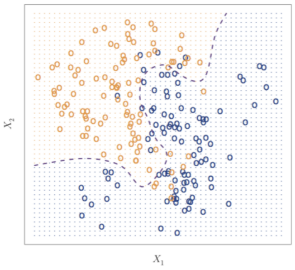
(c) Describe at least two strategies for obtaining a decision rule to label new images as Genuine or Imposter. Describe the difference between generalization and memorization.

Your answer to 3):

Cont. your answer to 3):

Question 4 - Hybrid Classification

□ **20 pts** A simulated data set consisting of 100 observations in each of two groups are indicated in orange (top) and blue (bottom) in figure below. The purple dashed line represents the Bayes decision boundary. The posterior probabilities obtained for a test data point $x_p = [5, 5]$ with 3 classifiers NB, KNN and SVM are shown in the table below.



Posteriori Probabilities

Algorithm	Orange class	Blue class
KNN	0.80	0.75
SVM	0.95	0.85
NB	0.91	0.90

(a) Write the Product Rule for designing a hybrid classifier

(b) Apply the Product Rule to the problem above and indicate whether the data point belongs to Orange (top) or Blue (bottom) class.

Your answer to 4):

Cont. your answer to 4):

Name: **Number:**

Question 5 - SVM Optimization

□ **20 pts** Write a function to search for the best C of a linear SVM classifier. The function should perform a grid-search looking 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):