Instituto Superior Técnico

Machine Learning (Aprendizagem Automática)

Exam of 13/1/2017. Duration: 3 hours

Number:	**First and last names:**

Important notes:

- **Present all answers in a very clear, ordered and detailed manner, with all steps of the calculations, and with a brief explanation of each one.**

  If you don't present your answers as indicated, your marks will be reduced, and can be very low, even if your answers are correct.
- Justify all answers, except in multiple-choice questions.
- Keep at least four significant digits in all calculations.
- For problems 1 to 4, the instructions on how to answer are given in the problems themselves. Solve problems 5 to 8 in separate sheets. You may solve several items of each of those problems in the same sheet.

Answer table for problems 1 to 3:

	a)	**b)**	**c)**	**d)**	**e)**
Problem 1					
Problem 2					
Problem 3					
Problem 4					

Problem 1 (0.8 marks for each item; wrong answers will be marked -0.5)

Answer this problem by writing, in the answer table above, in the box corresponding to each item, "T" for "True", or "F" for "False".

For each of the following learning techniques, indicate whether the corresponding objective functions can have local optima that are not global optima.

a) Support vector machines. b) ADALINEs. c) Multilayer perceptrons.

Problem 2 (0.8 marks for each item; wrong answers will be marked -0.5)

Answer this problem by writing, in the answer table above, in the box corresponding to each item, "T" for "True", or "F" for "False".

Indicate whether each of the following assertions is true or false.

a) The optimal value of the objective function for the estimation of a probability density through the EM algorithm, using a mixture of $p + 1$ Gaussians, cannot be higher than the objective function for the same estimation, using a mixture of p Gaussians.

b) The naive Bayes classifier that we have used, in our course, for language recognition is a special case of the Bayes classifier.

c) The objective function of multilayer perceptrons trained with gradient descent always stops improving after a finite number of iterations.

d) The objective function of the k-means algorithm always stops improving after a finite number of iterations.

e) The objective function for the estimation of a probability density by means of a mixture of Gaussians, through the EM algorithm, always stops improving after a finite number of iterations.

Problem 3 (0.8 marks; wrong answers will be marked -0.2)

Answer this problem by writing, in the answer table at the beginning of this exam, an "X" in the box corresponding to the sentence that you chose.

Indicate a true sentence:

a) The EM algorithm minimizes the posterior probability function.

b) The EM algorithm minimizes the likelihood function.

c) The EM algorithm maximizes the posterior probability function.

d) The EM algorithm maximizes the likelihood function.

Problem 4 (0.8 marks for each item; wrong answers will be marked -0.5)

Answer this problem by writing, in the answer table at the beginning of this exam, in the box corresponding to each item, "+1" or "−1".

The following table gives a training set for a classification problem. The x_i represent components of the input patterns, and d represents the desired classification of each pattern.

x_1	x_2	d	x_1	x_2	d
-2	7	-1	1	0	$+1$
2	6	-1	6	6	$+1$
4	12	-1	6	10	$+1$

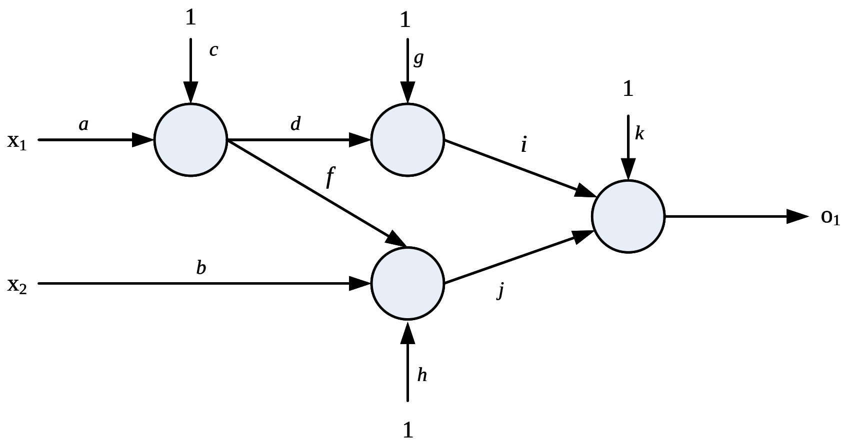
Consider a hard-threshold linear support vector machine (SVM) developed with this training set. Indicate the classification ($+1$ or -1) that the SVM gives to each of the following input patterns:

a) $(3,7)$. b) $(4,3)$. c) $(0,4)$.

Note: It may be useful to carefully plot the given patterns in square-lined paper before answering this problem.

Problem 5

Consider the following multilayer perceptron (MLP), in which all the units have as activation function $S(s) = \frac{s}{1+s^2}$. Also consider the training set given in the following table.



x_1	x_2	d_1
-1	-1	-0.5
-1	1	0.5

a) (1.6 marks) Draw the backpropagation network. Don't forget to include the gains of all branches, as well as the input variables.

b) (2 marks) Compute the value of the weight a after the second update using backpropagation in **batch mode**, assuming that, initially, all the weights were equal to 0.25, and that, after the first iteration, all weights were equal to 0.5. Training is performed with the fixed step-size parameter $\eta = 0.1$ and with momentum, with parameter $\alpha = 0.2$. The cost function is the mean squared error.

Problem 6 (1.5 marks)

Consider the training and pruning sets given in the following tables. The x_i represent attributes of the input patterns, d represents the desired output for each pattern, and "number" indicates the number of times that the corresponding pattern occurs in the set.

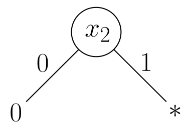
Training

x_1	x_2	x_3	d	number	x_1	x_2	x_3	d	number
0	0	A	0	2	1	0	B	0	1
0	0	B	0	2	1	1	A	1	2
0	1	A	1	1	1	1	B	0	1
0	1	B	0	2	1	1	B	1	2
0	1	B	2	1	1	1	B	2	2
1	0	A	0	2					

Pruning

x_1	x_2	x_3	d	number	x_1	x_2	x_3	d	number
0	0	A	0	2	0	1	B	0	1
0	0	A	1	1	1	0	A	0	2
0	0	B	0	2	1	0	B	2	1
0	0	B	2	1	1	1	A	0	1
0	1	A	1	2	1	1	B	0	2
0	1	A	2	1	1	1	B	1	1

The figure on the right shows a partially-constructed decision tree obtained with the ID3 algorithm, using the given training set.



Complete the construction phase of the ID3 algorithm, i.e., construct the node marked with "*" in the figure, and any nodes below it.

Problem 7

Consider the following data set: $\{(0,0),(2,-3),(4,-3)\}$.

a) (1.4 marks) Find its principal directions. You may indicate those directions by means of vectors.

b) (0.8 marks) Find the variances of the principal components of the data set.

c) (1.1 marks) Find the first principal component of the point $(2,1)$, and also find that point's reconstruction using the first principal component.

Note: If you haven't solved item a), assume that the two first principal directions of the given data set are the directions of the vectors $(5.736, 6.454)$ *and* $(-6.454, 5.736)$*, respectively*

Problem 8 (2 marks)

This problem is intended to distinguish the students that deal best with the topics studied in our course. In this problem, you must provide a very detailed response, with a very careful justification of each step.

Consider a classification support vector machine (SVM) that uses a second-degree polynomial kernel, i.e., a kernel of the form $K(\bar{x},\bar{y}) = (\bar{x}\cdot\bar{y}+a)^2$, where a is a scalar. Show that the classification boundary of the SVM in the input space is described by an equation of the form $\bar{x}^T A\bar{x} + \bar{v}^T\bar{x} + c = 0$, in which A is a matrix, $\bar{v}$ is a vector, and c is a scalar.