Instituto Superior Técnico

Machine Learning (Aprendizagem Automática)

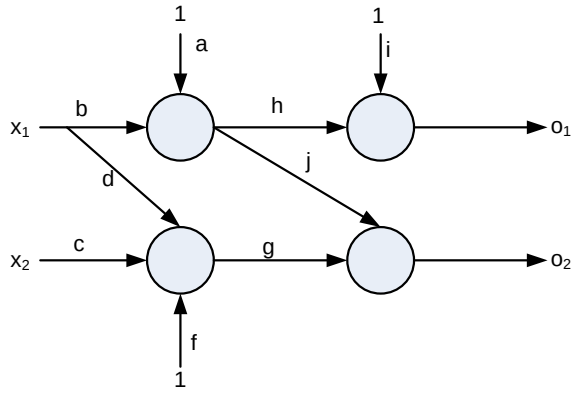
Exam of 9/1/2015. Duration: 3 hours

Notes:

- *Present all responses in a clear, ordered and detailed manner, with a brief justification of each step.*
- *Present all reasonings and calculations.*
- *Keep at least three digits after the decimal point in all calculations.*

Problem 1

Consider the multilayer perceptron (MLP) shown in the following figure, in which all units have as activation function the hyperbolic tangent. Also consider the training set shown in the following table. The MLP is to be trained using backpropagation in **batch mode** with fixed step sizes, with step size parameter $\eta = 0.1$. The cost function is the total squared error. Assume that in a certain iteration all weights were equal to 1, and that in the next iteration all were equal to 0.5.



x_1	x_2	d_1	d_2
–1	1	–1	0
1	–1	1	1

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

b) Compute the value of weight ***b*** after the next iteration, using backpropagation without momentum.

c) Assume that, after several iterations performed in the conditions indicated above, you found that the algorithm was not converging. What alternatives would you propose? If there are several possible alternatives, indicate them all; also indicate which one you would consider preferable, and why.

d) Repeat b), but now using the momentum acceleration technique with momentum parameter $\alpha = 0.1$.

Problem 2

a) Consider two-class *linear* classification problems. Explain what is the criterion that is used, in support vector machines, to choose the classification boundary for problems of this kind. Also explain why that is a good criterion for the choice of the classification boundary. Be as precise as possible in your explanation, defining the exact meaning of the terms that you use, whenever possible.

Note: You are requested to explain what is the criterion that is used to choose the boundary, but you are not requested to obtain the final form of the optimization problem that is solved to find that boundary.

b) Consider a classification problem in two dimensions, with two classes, in which the training set is given by

X_1	X_2	Class	X_1	X_2	Class
2	2	A	2	4	B
4	2	A	0	2	B
2	0	A	0	4	B
4	0	A	0	6	B

c) Obtain a linear support vector machine for this problem. Graphically sketch the training patterns, the classification boundary and the regions that correspond to the classification into classes A and B. Indicate which patterns are support vectors. You may obtain the support vector machine by inspection, but you should explain what you have done.

Problem 3

Note: In this problem, keep four digits after the decimal point in all calculations.

Consider the following training set: $T = \{1,3,7\}$.

a) Perform one iteration of the EM algorithm to estimate the parameters of a mixture of two Gaussians for these data. Use the following initial conditions:

$$\mu_1 = 2 \quad \sigma_1 = 2 \quad w_1 = 0.7$$
$$\mu_2 = 5 \quad \sigma_2 = 1 \quad w_2 = 0.3$$

b) Write the expression of the probability density function obtained after the iteration performed in a). If you have not solved that item, use the initial values of the parameters.

c) Use the result of a) to cluster the data from the training set into two clusters. If you have not solved item a), use the initial values of the parameters.

Problem 4

Consider the following training and pruning sets, to be used for inferring a decision tree with the ID3 algorithm:

Training

a_1	a_2	a_3	d	a_1	a_2	a_3	d
0	0	0	*A*	1	0	0	A
0	0	0	*A*	1	0	0	A
0	0	0	*A*	1	0	1	B
0	0	1	*B*	1	0	1	B
0	0	1	*B*	1	1	0	B
0	1	0	*A*	1	1	0	C
0	1	0	*A*	1	1	0	C
0	1	0	*C*				

Pruning

a_1	a_2	a_3	d	a_1	a_2	a_3	d
0	0	0	*A*	1	0	0	A
0	0	0	*A*	1	0	0	A
0	0	0	*B*	1	0	1	B
0	0	1	*B*	1	0	1	C
0	0	1	*B*	1	1	0	B
0	1	0	*A*	1	1	0	B
0	1	0	*B*	1	1	0	C
0	1	0	*C*				

a) Perform the construction phase of the algorithm.

b) Perform the pruning phase of the algorithm.

Problem 5

Consider the training set $\{[3,-1]^T, [2,-2]^T, [1,0]^T\}$.

a) Compute the first and second principal directions of the distribution of these patterns. Indicate those directions by means of vectors. Also find the variance of the data along both of those directions.

b) Compute the first principal component, the reconstruction with that principal component, and the reconstruction error of the pattern $[2,0]^T$. If you didn't solve item a) above, assume that the first principal direction is given by the vector $[1,3]^T$.

Problem 6

Assume that there are random data from two classes, C_0 and C_1, with $P(C_0) = 1/3$. The data are observed with errors: for the data from class C_0, the probability of error is p, meaning that the observation $\hat{C}_1$ is obtained with probability p and the observation $\hat{C}_0$ is obtained with probability $1 - p$; for the data from class C_1, the probability of error is q, with a similar meaning.

We wish to make a classifier of the data, based on the observations. The cost of classifying data from C_0 into C_1 is 2, and the cost of classifying data from C_1 into C_0 is 1. The costs of correct classifications are zero. Find the classifier with minimum expected cost, as a function of p and q.

Note that you should indicate the classifier in an explicit form, such as: "When $\hat{C}_0$ is observed, classify into C_0 if $p < 2q$; otherwise classify into C_1. When $\hat{C}_1$ is observed, classify into C_1 if $p > 1 - q$; otherwise classify into C_0."