Instituto Superior Técnico

Machine Learning (Aprendizagem Automática)

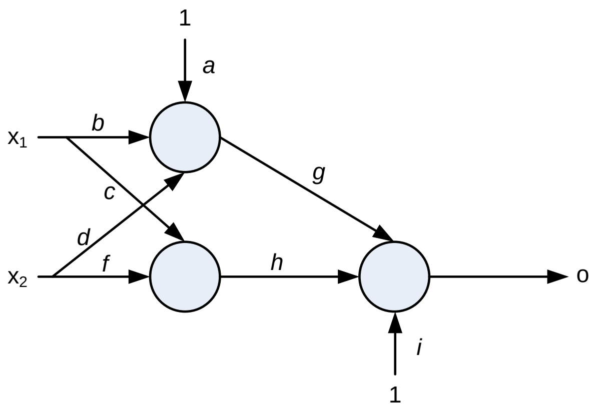
Exam of 11/1/2014. Duration: 3 hours

Notes:

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

Problem 1

Consider the multilayer perceptron 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.



x_1	x_2	d
-1	1	-1
1	1	1

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

b) Compute the value of weight $\boldsymbol{c}$ after the **first** update, using backpropagation in **batch mode**. The training is performed with fixed step sizes, with step size parameter $\eta = 0.1$. Assume that initially all weights are equal to 0.5. The cost function is

$$E = \frac{1}{2}\sum_{k=1}^{K}(o^k - d^k)^2$$

c) Repeat b), but now using backpropagation in **real time**, and using as cost function the previous one plus a regularization term for exponential weight decay with parameter $\lambda = 0.1$. Assume again that, initially, all weights are equal to 0.5.
Note: If you don't know how to use weight decay, perform the calculations without regularization. This will have a lower grade.

Problem 2

Consider data that are generated in the following manner: first, one of two classes, C_0 and C_1, is randomly chosen, both having probability 1/2. Then, a value $x \in \mathbb{R}$ is randomly generated following one of two distributions, depending on the class that was chosen:

$$p(x|C_0) = \frac{1}{2}e^{-|x|}$$
$$p(x|C_1) = e^{-2|x|}.$$

a) Find the expression of the posterior probability of class C_0.

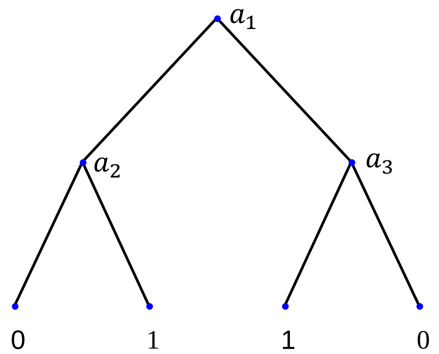
b) Find the classifier, with input x and output $y \in \{C_0, C_1\}$, that will classify the data x with the minimum possible expected error rate relative to the classes that the data originally came from. Indicate the classification regions, i.e., the set of values of x that would be classified in class C_0 and the set of values that would be classified in class C_1. Find the boundary point(s) between the two regions with the accuracy indicated above, of at least three decimal places.

c) Repeat b) above, but now assuming that the cost of classifying data that belongs to C_0 as belonging to C_1 is equal to 1, that the cost of classifying data that belongs to C_1 as belonging to C_0 is equal to 1.5, and that we wish to find the classifier that minimizes the expected cost of classification errors.

Problem 3

Consider the training set and the decision tree given in the following table and figure, respectively.

a_1	a_2	a_3	d	a_1	a_2	a_3	d
0	0	0	0	0	1	1	1
0	0	0	0	0	1	1	1
0	0	0	0	1	0	0	1
0	0	0	0	1	0	1	1
0	0	0	0	1	1	0	1
0	1	1	1	1	1	1	0
0	1	1	1	1	1	1	0



a) Compute the number of errors that the tree yields in the training set.

b) Compute the total coding length, for the given tree and training set, according to the MML criterion.

c) Perform the pruning of the tree, regarding the node marked "a_2", with the given training set.

Problem 4

Note: In this problem, keep four decimal places in all calculations.

Consider the following training set: T={-4, -2,4,5}.

a) Perform one iteration of the k-means algorithm using two centers with initial positions -3 and 3.

b) Perform one iteration of the EM algorithm to estimate the parameters of a mixture of two Gaussians. Assume the following initial conditions:

$$\mu_1 = -3 \qquad \sigma_1 = 1 \qquad w_1 = 0.7$$
$$\mu_2 = 3 \qquad \sigma_2 = 2 \qquad w_2 = 0.3$$

c) Write the expression of the probability density function of the Gaussian mixture model that you obtained in b). If you didn't solve b), use the initial values of the parameters given above.

Problem 5

Consider the training set $\{[3,-3]^T, [0,0]^T, [4,-2]^T, [1,1]^T\}$.

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 $[3,2]^T$.

If you didn't solve item a) above, assume that the first principal direction is given by the vector $[1,2]^T$.

c) Consider a bidimensional random variable $\bar{X}$ whose components, X_1 and X_2, are independent from each other and have variances 1 and 2, respectively. A new random variable $\bar{Y}$ is defined by

$$Y_1 = X_1 + X_2$$
$$Y_2 = -2X_1 + X_2.$$

Find the principal directions of $\bar{Y}$ and the variance of $\bar{Y}$ along both of those directions.