Instituto Superior Técnico

Machine Learning (Aprendizagem Automática)

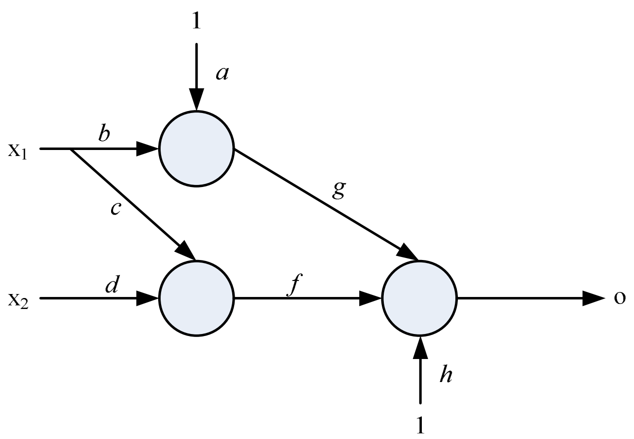
Exam of 28/1/2013. Duration: 3 hours

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, where all units of the first layer have as activation function the hyperbolic tangent and the unit of the second layer is linear. Also consider the training set shown in the following table.



x_1	x_2	d
-1	1	1
1	-1	0

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 values obtained at the multilayer perceptron's outputs for the **first** input pattern. Assume that all weights are equal to 0.5.

c) 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 $\eta = 0.2$. Assume that initially all weights are equal to 0.5. The cost function is

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

d) 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.

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:

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

In these expressions, the constant e is always the base of the natural logarithm, even when it occurs in the exponent.

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

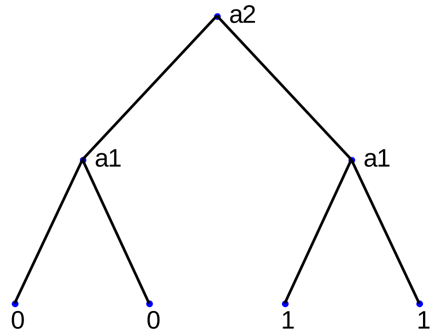
b) Assume that we wished to train a supervised system with input x and output $o(x)$, the desired output being $d = 0$ for data from class C_0 and $d = 1$ for data from class C_1. Assume, furthermore, that the training was performed so as to minimize the expected value of the quadratic error between $o(x)$ and d. Find the expression of the optimal output function $o^*(x)$. Graphically sketch that function for $x \in [0,10]$ (compute a few values, to be able to make a reasonably good sketch).

c) Assume now that we wished to make a classifier, with input x and output $y \in \{C_0, C_1\}$, that would 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.

Problem 3

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

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



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

b) Perform the pruning of the tree using the given training set.

Problem 4

Consider the training set $T = \{-5, -3, 1, 5\}$.

a) Perform one iteration of the k-means algorithm using two centers with initial positions –4 and 0.

b) Find the value of the cost function for the position of the centers obtained in a). Is this a fixed point of the algorithm?

c) Perform one iteration of the EM algorithm to estimate the parameters of a mixture of three Gaussians with this training set. Assume the following initial conditions:

$$\mu_0 = -4 \quad \sigma_0 = 1 \quad w_0 = 0.6$$
$$\mu_1 = 0 \quad \sigma_1 = 2 \quad w_1 = 0.4$$

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

Problem 5

a) Assume that you are training a support vector machine (SVM) with a bidimensional input, and that you have found that there are only two support vectors, $\bar{x}^1 = [0,0]^T$ and $\bar{x}^2 = [2,2]^T$. Assuming that the SVM performs a linear classification in input space, find the expressions of the maximum-margin classification boundary and of the two margin boundaries, and graphically sketch the two support vectors and the boundaries.

   Note: In this problem you are asked to find the classification boundary through an exact and well justified reasoning, and not by inspection. Finding the boundary by inspection will yield a lower grade in this item, and won't help you in item b), below.

b) Consider now an SVM that makes the linear classification in feature space with the kernel $k(\bar{x}, \bar{y}) = (\bar{x} \cdot \bar{y})^2$, where the dot denotes the inner product. Assume that the only support vectors are the ones given in a) above. Find the equations of the maximum-margin classification boundary and of the margin boundaries in terms of the components of the input pattern, and graphically sketch the two support vectors and the boundaries in input space.

   Suggestion: Start by completely expressing, in terms of inner products, the procedure that you used in a) to find the boundaries. This may help you to better understand which procedure should be used here.