Instituto Superior Técnico

Machine Learning (Aprendizagem Automática)

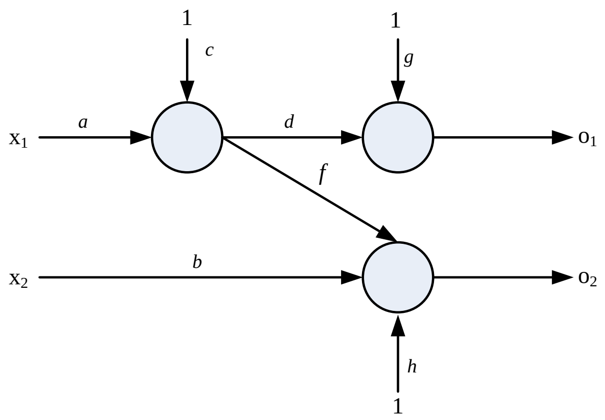
Exam of 27/1/2014. Duration: 3 hours

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 digits after the decimal point in all calculations.*

Problem 1

Consider the multilayer perceptron shown in the following figure, in which all units have as activation function $\frac{1}{1+e^{-s}}$, where s is the input sum. Also consider the training set shown in the following table.



x_1	x_2	d_1	d_2
1	1	-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. Assume that, initially, all weights were equal to 0.9, and that after the first weight update in **real time mode**, all weights became equal to 1.

b) Compute the value of weight ***c*** after **the second weight update in real time mode**. The training is performed with fixed step sizes, with step size parameter $\eta = 0.1$ and no momentum. The cost function is the total squared error.

c) Repeat the update you performed in b), but now using the momentum acceleration technique with momentum parameter $\alpha = 0.1$.

Problem 2

Note: In this problem, indicate enough data, so that the sketches that you make are uniquely defined.

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

x_1	x_2	class
0	2	A
2	0	A
$-\sqrt{7}$	$-\sqrt{7}$	B
3	3	B

a) Graphically sketch the positions of these patterns. Are the two classes linearly separable?

b) Consider the following nonlinear mapping from input space to a two-dimensional feature space:

$$\varphi(\bar{x}) = [(x_1)^2, (x_2)^2].$$

Find the kernel function that corresponds to this mapping.

c) Sketch the training patterns in feature space.

d) Find (by inspection) the widest-margin classifier in feature space. Indicate the support vectors, and sketch the classification boundary and the margin boundaries in that space; you may draw them on the sketch of c) above. Find the equations of the classification boundary and of the margin boundaries.

e) Find and sketch the classification regions of classes A and B in feature space; you may draw them on the sketch of c) and d) above.

f) Find the support vectors and the expressions of the classification boundary and of the margin boundaries in input space, and sketch them; you may draw them on the sketch of a) above.

g) Find and sketch the classification regions of classes A and B in input space; you may draw them on the sketch of a) and f) above.

h) Indicate the condition, in terms of the components of an input pattern $\bar{x}$, for that pattern to be classified in class A by the support vector machine that you have just developed.

Problem 3

Consider the training set given in the following table, for which we wish to build a decision tree.

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

a) Assume that the root of the decision tree has already been chosen, and that it tests attribute a_3. Perform the building operation regarding both of the children of the root node. Add the correct leaf children to any attribute nodes that you create. Use the MML criterion with full-depth tree construction, with regularization parameter $\lambda = 1$.

b) Compute the error rate in the training set for the tree you obtained in a). If you didn't solve a), assume that the child corresponding to $a_3 = 0$ tests attribute a_2, and that the child corresponding to $a_3 = 1$ tests attribute a_1.

Problem 4

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

a) Consider the training set $\{(-5,-1), (-3,1), (0,0), (3,1)\}$. Perform one iteration of the k-means algorithm using two centers with initial positions $(-5, 0)$ and $(4, 0)$.

b) Compute the value of the cost function of the k-means algorithm for the result that you found in a). If you didn't solve a), perform the computation for the initial positions of the centers.

c) Consider the training set $\{-5, -3, 0, 3\}$. Perform one iteration of the EM algorithm to estimate the parameters of a mixture of two Gaussians. Assume the initial conditions

$$\mu_1 = -2, \qquad \sigma_1 = 1, \qquad w_1 = 0.5,$$
$$\mu_2 = 0, \qquad \sigma_2 = 2, \qquad w_2 = 0.5\,.$$

d) 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 a system, with a scalar output o in the open interval $(0,1)$, which is trained in supervised mode. Assume that the system has a single input pattern (similarly to what we have done when studying the statistical properties of supervised systems). Also assume that the desired values d are binary with probabilities $P(d = 1) = p$ and $P(d = 0) = 1 - p$, with $p \in (0,1)$, and that the cost function is

$$C(o,d) = \begin{cases} o & \text{if } d = 0 \\ o - \log o & \text{if } d = 1. \end{cases}$$

Find, as a function of p, the value of the output that minimizes the expected cost. Show that that value actually yields the absolute minimum of the expected cost, and not some other stationary point.