**Instituto Superior Técnico**

**Machine Learning (Aprendizagem Automática)**

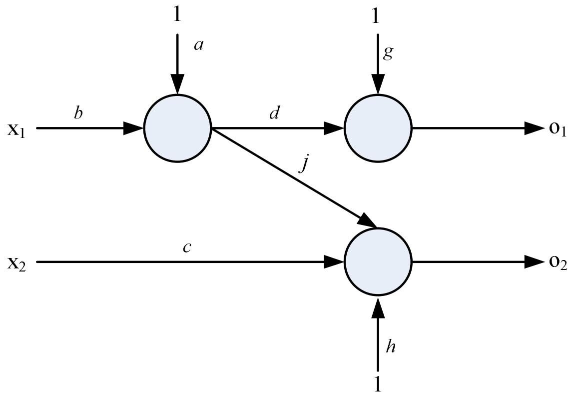
**Exam of 19/1/2013. Duration: 3 hours**

***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 following multilayer perceptron shown in the figure below-left, where all the units have as activation function the logistic function $f(s) = \frac{1}{1+e^{-s}}$ . Also consider the training set given on the table below-right.




| $x_1$ | $x_2$ | $d_1$ | $d_2$ |
|---|---|---|---|
| -1 | 1 | 0.5 | 0.5 |
| 1 | -2 | 0 | -1 |

a) Compute the values obtained at the multilayer perceptron's outputs for the **second** input pattern. Assume that all weight values are equal to 0.5.

b) Draw the backpropagation network. Don't forget to include the gains of all branches, as well as the input and output variables. If necessary, first draw the multilayer perceptron as a network with more detail, in order to be able to more easily specify some parameters of the backpropagation network.

c) Assume that, initially, all the multilayer perceptron's weights were equal to 0.4, and that, after a **first** update using backpropagation in **real-time mode,** all the weights were changed to 0.5. Compute the value of weight ***a*** after a **second** update, also in **real-time mode**. The cost function is the total squared error and the training is performed using non-adaptive step sizes with step size parameter $\eta$ = 0.1, without momentum.

d) Repeat 3) using the same step size parameter $\eta$ = 0.1, with momentum with parameter $\alpha$ = 0.1.

## Problem 2

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 |
|---|---|---|---|---|---|
| 0 | $\sqrt{2}$ | A | 1 | -3 | B |
| 1 | $\sqrt{2}$ | A | $\sqrt{3}$ | 2 | B |
| $\sqrt{2}$ | 0 | A | $\sqrt{6}$ | 1 | B |
| 0 | 1 | A | $-\sqrt{2}$ | $-\sqrt{6}$ | B |

a) Graphically sketch the positions of these patterns. Show that the two classes are not linearly separable. Briefly describe a technique that can be used, in support vector machines, to obtain a linear separation when the input data are not linearly separable.

b) Consider the following nonlinear mapping from input space to a two-dimensional feature space. Sketch the training patterns in feature space.

$$\varphi(x) = (x_1{}^2;\ x_2{}^2 - 2)$$

c) Find the kernel function that corresponds to the nonlinear mapping given above.

d) Find (by inspection) the widest-margin classifier in feature space. Indicate the support vectors and the equations of the classification boundary and of the margin boundaries.

e) Sketch the classification regions of classes A and B in feature space. You may draw them on the sketch of b) above.

f) Indicate, in input space, the support vectors and sketch the classification boundary and margins. You may draw them on the sketch of a) above.

g) Sketch the classification regions of classes A and B in input space.

**Problem 3**

Consider the training set $T_1$ = {–2, –1,1,4,5}.

a) Perform one iteration of the k-means algorithm using three centers with initial positions -1, 0, and 3.

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

Now consider a new training set $T_2$ = {–2, –1,1}.

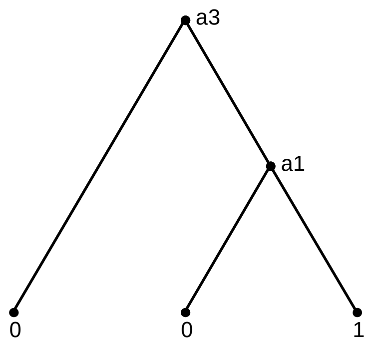
c) Perform one iteration of the EM algorithm to estimate the parameters of a mixture of three Gaussians. Assume the following initial conditions:

$$\begin{array}{lll} \mu_0 = -1 & \sigma_0 = 1 & w_0 = 0.6 \\ \mu_1 = 0 & \sigma_1 = 1 & w_1 = 0.2 \\ \mu_2 = 3 & \sigma_2 = 1 & w_2 = 0.2 \end{array}$$

**Problem 4**

Consider the training set and the tree given on the following table and figure, respectively. In the table, the $a_i$ represent attributes of the input patterns and $d$ represents the desired output for each pattern.

| $a_1$ | $a_2$ | $a_3$ | $d$ | $a_1$ | $a_2$ | $a_3$ | $d$ |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 |
| 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 |
| 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |




a) Find 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 5**

Consider the set of patterns $\{[6,4]^T, [-2,0]^T, [3,0]^T, [1,4]^T\}$ . The patterns are considered equiprobable.

a) Find the first and second principal directions of the distribution of these patterns. Indicate those directions by means of vectors.

b) Find the first principal component, the reconstruction with that principal component and the reconstruction error of the pattern $[3,5]^T$, based on the given distribution.
*If you didn't solve item a) above, assume that the first principal direction was given by the vector* $[-3,1]^T$.

c) Consider two random variables $X$ and $Y$ which are statistically independent from each other. Assume that they have the same principal directions, although possibly not in the same order (for example, the first principal direction of $X$ may be the second principal direction of $Y$). Show that, for any $a, b \in \mathbb{R}$, the random variable $Z = aX + bY$ also has the same principal directions (possibly in a different order from those of $X$ or $Y$).

If you're not able to give the proof as requested, you may make any additional assumptions that you find necessary (for example, that the principal directions of $X$ and $Y$ are in the same order, or that $a = b$). However, the value of this problem will be reduced, depending on the additional assumptions that you make.