**Instituto Superior Técnico**

**Machine Learning (Aprendizagem Automática)**

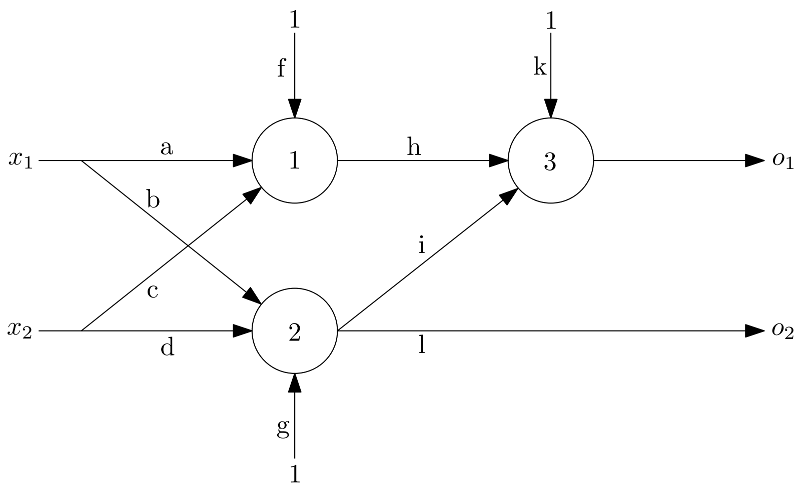
**Exam of 26/1/2015. Duration: 3 hours**

***Notes:***

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

## Problem 1

Consider the multilayer perceptron (MLP) shown in the following figure, in which all units have the activation function $1/(1+e^{-s})$. Also consider the training set shown in the following table. The MLP is to be trained using backpropagation in **real-time mode** with fixed step sizes and without momentum, with step size parameter $\eta = 0.2$. The cost function is the total squared error. The initial values of the weights are all equal to 0.5.




| $x_1$ | $x_2$ | $d_1$ | $d_2$ |
|---|---|---|---|
| –1 | 1 | 1 | 2 |
| 1 | –1 | 2 | 1 |

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

b) Compute the value of weight ***d*** after the first weight update, using backpropagation without weight decay.

c) Repeat b), but now using exponential weight decay with parameter $\lambda = 0.1$.

d) Explain what can be the advantage of using exponential weight decay when training a multilayer perceptron.

## 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 | 3 | A | 0 | 0 | B |
| 0 | –3 | A | 0 | –2 | B |
| 4 | 1 | A | 1 | 1 | B |
| 4 | –2 | A | | | |

a) Sketch the points of the training set in the input space. Are the classes linearly separable in that space?

b) Consider the mapping to feature space $(x_1, x_2) \rightarrow (x_1, x_2^2)$. Indicate the expression of the kernel that corresponds to that mapping.

c) Sketch the points of the training set in the feature space. Also sketch the classification boundaries and the margin boundaries of the linear support vector machine (SVM) optimized for the training set in the feature space. Indicate which are the support vectors. You can visually find the boundaries and the support vectors, but you should describe, in a precise way, the criterion that you have used to find them.

d) Sketch the classification boundaries and the margin boundaries in the input space, on the plot made in a). Find the equations of those boundaries. Indicate which are the support vectors, and which are the classification regions for the two classes, in the input space.

e) Indicate the condition that the nonlinear SVM that you have developed uses for classifying patterns into class A. Express the condition in terms of the input coordinates.

**Problem 3**

*In this problem you may compare distances visually, as long as the result of the comparison is clear.*

Consider the following training set: $T = \{(0,-3),(2,5),(4,1),(6,0),(8,2)\}$.

a) Iterate the k-means algorithm, until you reach a fixed point, using two centers with initial positions (0,0) and (0,7). Compute the value of the algorithm's cost function at the fixed point that you found.

b) Indicate a position of the centers that corresponds to a fixed point of the algorithm with a value of the cost function that is lower than the one found in a). You don't need to show how you found the fixed point. You only need to show that it is a fixed point, and that the corresponding value of the cost function is lower than the one found in a).

**Problem 4**

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

Consider the following training set: $T = \{-3,-1,6,10\}$. Assume that you wish to use the EM algorithm to estimate a mixture of two Gaussian distributions from these data, with the following initial values of the parameters:

$$\mu_1 = -2 \quad \sigma_1 = 2 \quad w_1 = 0.2$$
$$\mu_2 = -1 \quad \sigma_2 = 1 \quad w_2 = 0.8$$

a) Write the expression of the probability density function that corresponds to these values of the parameters.

b) Perform one iteration of the EM algorithm, starting from the parameter values indicated above.

c) Indicate the approximate values that the parameters would converge to, if the iterations of the algorithm were continued.

**Problem 5**

Consider the training set $\{[-3,0]^T,[0,4]^T,[2,0]^T,[5,4]^T\}$.

a) Find the two principal directions of the distribution of these patterns. Indicate those directions by means of vectors. Also find the variance of the training set along both of those directions.

b) Find the two principal components of the pattern $[4,1]^T$, using the result of a). Also find the reconstruction of that pattern using the first principal component, and find the corresponding reconstruction error. If you didn't solve item a), assume that the first principal direction is given by the vector $[3,1]^T$.

c) In problems with many dimensions, finding eigenvectors by the usual algebraic method can be computationally heavy. Sometimes, one wishes to find only one principal direction of a set of data, and then the following iterative method for finding an eigenvector of a matrix $A$ (of size $n \times n$) can be more efficient:

   1. Randomly choose an initial vector $\overline{u}^{(0)}$, of size $n$.
   2. Iterate:

   $$\overline{u}^{(n+1)} = A\overline{u}^{(n)}.$$

   Assume, for simplicity, that $A$ is symmetrical and positive semi-definite, and that its eigenvalues are all different from one another. Show that the sequence of directions of $\overline{u}^{(n)}$ converges to the direction of one of the eigenvectors of $A$.

   For reasonable choices of the probability distribution from which $\overline{u}^{(0)}$ is drawn, which is, with high probability, the eigenvalue corresponding to the direction found by this method? And what is the value of that "high probability"?

   *Suggestion: Express the vectors in the basis formed by the eigenvectors of $A$.*