Instituto Superior Técnico

Machine Learning (Aprendizagem Automática)

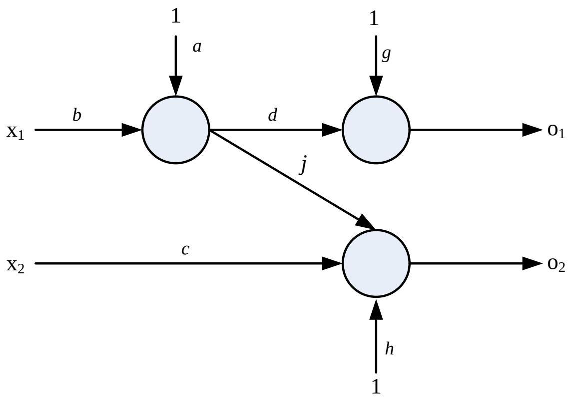
Exam of 8/1/2016. Duration: 3 hours

Important notes:

- Present all answers in a clear, ordered and detailed manner.
- **Justify all answers.** In calculations, present all steps, with a brief explanation of each one.
- Keep at least four significant digits in all calculations.
- Solve the problems in separate sheets. You may solve the various items of each problem in the same sheet.

Problem 1

Consider the following multilayer perceptron, where all the units have as activation function the hyperbolic tangent.



Consider also the following training set.

x_1	x_2	d_1	d_2
-1	1	0.5	0.5
1	-1	0	-1

a) (1 point) Draw the backpropagation network. Include the gains of all branches, as well as the input variables.

b) (1.8 points) Compute the value of weight ***b*** after the first update, using backpropagation **in real-time mode**. The cost function is the total squared error, and the training is performed using the fixed step-size parameter $\eta = 0.2$. Assume that, initially, all the weights are equal to 0.1.

c) (1.2 points) Consider a multilayer perceptron with a single output, and the two cost functions $C_1 = \sum_k (e^k)^2$ and $C_2 = (\sum_k e^k)^2$, where e^k is the output error corresponding to the input pattern $\bar{x}^k$. Which of the cost functions is the most appropriate one to train the multilayer perceptron?

Problem 2

Assume that we wish to estimate a mixture of two Gaussians for the following data points:

$$\bar{x}^1 = [-4,0]^T, \quad \bar{x}^2 = [4,-2]^T, \quad \bar{x}^3 = [-2,2]^T.$$

The initial parameter values are $\bar{\mu}_1 = [-3,0]^T$, $\bar{\mu}_2 = [1,2]^T$, $V_1 = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$, $V_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$, $w_1 = 0.4$, $w_2 = 0.5$ (V_i are the covariance matrices of the Gaussian distributions). The values of the two Gaussian densities at the data points, for the initial parameter values, are supplied:

	$g_1(x)$	$g_2(x)$
$\bar{x}^1$	0.0351	0.0001
$\bar{x}^2$	0.0001	0.0002
$\bar{x}^3$	0.0213	0.0084

a) (1.2 points) Compute the log-likelihood of the initial parameter values.

b) (1.8 points) Perform one iteration of the EM algorithm.

c) (1 point) How do you expect the log-likelihood value to have changed after the iteration that you just performed? Respond without computing the new log-likelihood.

Problem 3

Consider the following training set, in which the a_i represent attributes of the input patterns and d represents the desired classification 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	1	1	0	0	1
0	1	0	0	1	0	1	0
0	1	0	0	1	1	1	1
0	1	1	0				

a) (1.8 points) Using this training set, build a decision tree with the ID3 algorithm, without pruning.

b) (1 point) Compute the confusion matrix for the decision tree that you obtained.

Problem 4

Consider the bidimensional patterns with binary components, $\bar{x} = (x_1, x_2)$ with $x_1, x_2 \in \{0,1\}$, and two classes, C_1 and C_2. The joint probabilities of the patterns and the classes are given in the following table.

$\bar{x}$	$P(\bar{x}, C_1)$	$P(\bar{x}, C_2)$
(0,0)	0.1	0.1
(0,1)	0	0.1
(1,0)	0.1	0.3
(1,1)	0.2	0.1

a) (1.7 points) Find the probability of error of a Bayes classifier for these patterns.

b) (1.7 points) Find the classification of the pattern (1,1) according to the Naïve Bayes classifier for these data, using x_1 and x_2 as features.

Problem 5

Consider a bidimensional random variable, $\bar{X} = (X_1, X_2)$. The mean and variance of X_1 are 2 and 4, respectively. X_2 is given by $X_2 = X_1/2 + \varepsilon$, where ε is a random variable that is independent from X_1 and that has mean 0 and variance 2.

a) (1.6 points) Find the principal directions of the distribution of $\bar{X}$. You may indicate those directions by means of vectors. *If you are unable to solve this item, use, for the next items,* $\bar{v}_1 = [1.224, 1]^T, \bar{v}_2 = [0.8170, -1]^T, \lambda_1 = 4.793,$ *and* $\lambda_2 = 1.239$.

b) (1 point) Find the variances of the two principal components of the distribution of $\bar{X}$.

c) (1.2 points) Find the reconstruction of the pattern $[3,4]^T$ with the first principal component, and also find its reconstruction error.

Problem 6 (2 points)

Consider two multidimensional random variables, X and Y, that have the same principal directions. Prove the truth, or the falsehood, of the following assertion:

For any X and Y in these conditions, the random variable $Z = X + Y$ has the same principal directions as X and Y.

Note: If you're not able to solve the problem as stated above, you may make additional assumptions to simplify it. However, in that case, your score will be reduced by an amount that depends on the assumptions that you make.