EXAMEN CLASSIFICATION - 3TSI

Mercredi 7 Janvier 2015 (14h-15h45)

Partiel avec polycopié de cours et copie des transparents autorisés

Exercice 1: Bayesian Classifier

We consider a classification problem with two equiprobable classes ω_1 and ω_2 whose densities are

$$f(x|\omega_i) = \frac{\frac{1}{\pi b}}{1 + \left(\frac{x-a_i}{b}\right)^2}, \quad x \in \mathbb{R}, \tag{1}$$

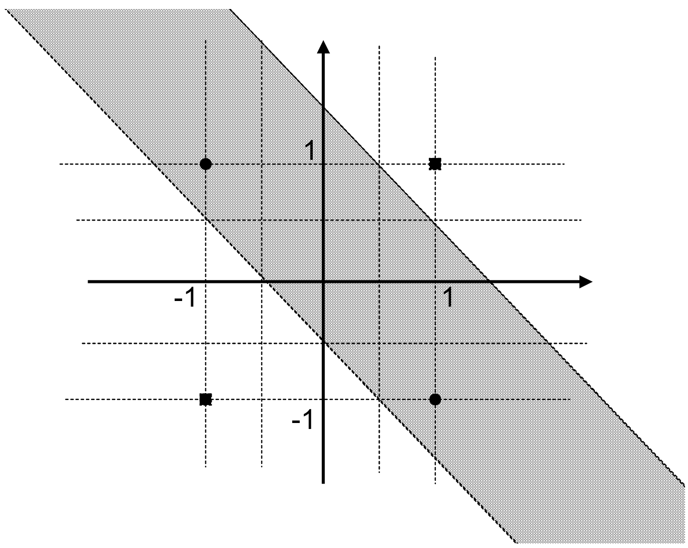
with $i \in \{1, 2\}$, $b > 0$ and $a_2 > a_1$.

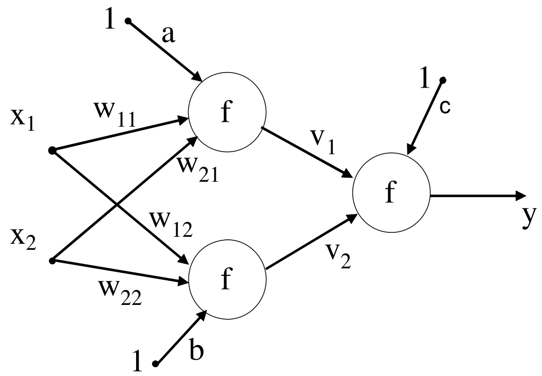
1. Derive the Bayesian classification rule associated with this problem.
2. Express the error probability of the Bayesian classifier as functions of $a_2 - a_1$ and b. Compute the limit of this probability when $a_2 - a_1$ tends to $+\infty$ and explain this result. Same question when b tends to $+\infty$.

Exercice 2: Neural Network

We consider a classification problem with two classes ω_1 and ω_2 and the training samples $\boldsymbol{x}_1 = (1, 1)^T$, $\boldsymbol{x}_2 = (-1, -1)^T$, $\boldsymbol{x}_3 = (1, -1)^T$ and $\boldsymbol{x}_4 = (-1, 1)^T$. We know that $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are associated with class ω_1 whereas $\boldsymbol{x}_3$ and $\boldsymbol{x}_4$ are associated with class ω_2.

1. We build a neural network with one hidden layer with two nodes and one output y such that $y = 1$ if $\boldsymbol{x}$ belongs to the dark gray area displayed in the figure below and $y = 0$ else. All non-linearities used in this neural network are Heaviside functions such that $f_s(u) = 1$ if $u > 0$ and $f_s(u) = 0$ else. We also assume that the output of node 1 of the first layer is equal to 1 when $\boldsymbol{x}$ falls on one side of one of the two straight lines delimiting the dark gray area (the side corresponding to the region inside the dark gray area) and is equal to 0 else. The output of node 2 is defined similarly with the other straight line delimiting the dark gray area. Provide the values of the different weights of this neural net which is displayed in the next page of this exam.





2. The neural net investigated in the first question is displayed in the figure above where $\boldsymbol{w} = (w_{11}, w_{12}, w_{22}, w_{22})^T$ is the weight vector of the first layer with offsets a and b, and $\boldsymbol{v} = (v_1, v_2)^T$ is the weight vector of the output with offset c. We recall the following relations

$$y_1 = f(w_{11}, x_1 + w_{21}x_2 - a), \; y_2 = f(w_{12}, x_1 + w_{22}x_2 - b) \text{ and } y = f(v_1 y_1 + v_2 y_2 - c)$$

where y_1 and y_2 are the outputs of the nodes of the hidden layer and where we have assumed that $f(t) = \frac{\exp(\alpha t)}{1+\exp(\alpha t)}$ (contrary to the first question). What are the backpropagation update rules associated with this neural network?

Questions related to the working paper

1. What are the advantages of hyperspectral sensors with respect to multispectral sensors?
2. Explain what the authors mean by the "curse of dimensionality" (Second column of page 1778).
3. Explain the differences between supervised and unsupervised classification rules.
4. Explain the principles of the sequential forward selection (SFS) (mentioned in the first column of page 1779) and an example of feature selection criterion that can be used for the SFS method.
5. How can we obtain Eq. (5) from (4)?
6. What is the leave-one-out (LOO) rule mentioned in the first column of page 1782?
7. Provide the mathematical expressions of the discriminant functions used for the one-against-all (OAA) rule mentioned after Eq. (16)?
8. What are the minimum and maximum value of the score function $S_i(x)$? When do we obtain these minimum and maximum values?
9. Explain how the tree of Fig. 6 (a) has been obtained and how a vector $\boldsymbol{x}$ is classified using this tree.
10. Explain how the tree of Fig. 6 (b) has been obtained. In particular, justify the order of the different classes (i.e., ω_7 first, ω_1 second, ...) and explain how a vector $\boldsymbol{x}$ is classified using this tree.
11. Assume that we have $n = 200$ test samples ($n_0 = 100$ from class ω_0 and $n_1 = 100$ from class ω_1). A classifier correctly classifies 70 samples from class ω_0 (called true negatives) and 80 samples from class ω_1 (called true positives). What is the overall accuracy for this problem? (the overall accuracy is indicated in Table V for the classification of hyperspectral pixels).
12. How do the authors of this paper justify the poor performance of k-nearest neighbor rule for their problem?