CLASSIFICATION EXAM - 3TSI

Friday, January 22, 2016

Lecture notes and slides authorized

Exercice 1 (11 points)

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

$$f(x|\omega_1) = 2x\mathbb{I}_{]0,1[}(x) \quad \text{and} \quad f(x|\omega_2) = 2(1-x)\mathbb{I}_{]0,1[}(x) \tag{1}$$

where $\mathbb{I}_{]0,1[}(x)$ is the indicator function on the interval $]0,1[$ (such that $\mathbb{I}(x) = 1$ if $x \in]0,1[$ and $\mathbb{I}(x) = 0$ if $x \notin]0,1[$).

1. (3 pts) Derive the Bayesian classification rule associated with this problem when we use the $0-1$ cost function and when the two classes have the prior probabilities $P(\omega_1) = 1/4$ and $P(\omega_2) = 3/4$. How does this rule modify when the two classes are equiprobable? Explain this result. Determine the probability of error in the equiprobable case.

2. (3 pts) Assume that we have a learning set composed of two elements of class ω_1 denoted as $x_1 = 3/4$ and $x_2 = 7/8$ and two elements of class ω_2 denoted as $x_3 = 1/8$ and $x_4 = 3/8$. What is the classification rule associated with the nearest neighbor rule. The asymptotic error probability of the nearest neighbor rule is known to be

$$P_1 = \int_{-\infty}^{+\infty} \left[1 - \sum_{i=1}^{2} P^2(\omega_i|x)\right] f(x)dx.$$

   Compute this error probability in the equiprobable case. Check that this result is in good agreement with the Cover and Hart inequality.

3. (3 pts) We assume now that the probability density function $f(x|\omega_1)$ is unknown and estimate it using the following estimator

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} \phi\left(\frac{x - x_i}{h_n}\right)$$

   where $x_1, ..., x_n$ are training samples from the class ω_1 (i.e., distributed according to $f(x|\omega_1)$) and

$$\phi(u) = \begin{cases} e^{-u} & \text{if } u > 0 \\ 0 \text{ sinon} \end{cases}$$

   Provide some motivations for the estimator $f_n(x)$ introduced above. Determine $E[f_n(x)]$ as a function of x and h_n. Determine also

$$\lim_{h_n \to 0} E[f_n(x)].$$

   What can we conclude about the estimator $f_n(x)$?

4. (2 pts) Create by hand a dendrogram for the following 6 points in one dimension: $x_1 = -5.5$, $x_2 = -4.0$, $x_3 = -3.0$, $x_4 = 5.0$, $x_5 = 6.1$ and $x_6 = 7.3$, when the distance between two clusters X_i and X_j is defined as

$$d(X_i, X_j) = \min_{x \in X_i, y \in X_j} d(x, y).$$

Questions related to the working paper (10 points)

Remark: please make sure to justify all your responses very carefully.

1. (1 pt) Explain how we can classify a feature vector with the decision tree displayed in Fig. 8.1.
2. (0.5 pt) Does CART belong to the class of supervised or un-supervised classification methods?
3. (1 pt) Can we always build a tree with binary decisions? Justify your response by means of an example.
4. (1 pt) Build a branch of a decision tree which leads to the decision region R_1 displayed in the left figure of Fig. 8.3.
5. (1 pt) What is the value of the entropy $i(N)$ defined in (1) for equally likely (equiprobable) classes? for two classes with respective probabilities $P(\omega_1) = 0$ and $P(\omega_2) = 1$?
6. (0.5 pt) Why might we prefer to use the Gini impurity index rather than the misclassification impurity?
7. (1 pt) Explain the term "cross-validation" (appearing page 11 of the paper).
8. (2 pts) As explained in the paper, we can use hypothesis testing to decide whether we have to stop the growing of a tree at a given node. Explain how this strategy is working. What is the distribution of χ^2 defined in (9)? Provide a mathematical expression of the test threshold as a function of the confidence level α (probability if false alarm of the test) and the inverse cumulative distribution function of χ^2.
9. (1 pt) Explain how the first threshold 0.6 has been obtained in the top right tree of Example 1.
10. (1pt) Explain "Preprocessing by principal components can be effective" in Section 3.7.1. In particular, explain how these principal components can be computed.