CLASSIFICATION EXAM - 3TSI

Monday, November 28, 2016

Lecture notes and slides authorized

Exercice 1 : Classification using logistic regression (9 points)

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

$$f(x|\omega_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x - m_i)^2\right] \quad i = 1, 2 \tag{1}$$

with $x \in \mathbb{R}$, $\sigma > 0$ and $m_1 > m_2$.

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 $P(\omega_1) = P_1$ and $P(\omega_2) = P_2$. Interpret this result using the centroid distance rule when $P_1 = P_2$ and $P_1 > P_2$. Express the probability of error of this rule as a function of m_1, m_2, σ^2 and the cumulative distribution function of the $\mathcal{N}(0,1)$ Gaussian distribution denoted as F.

2. (2 pts) Show that the Bayesian decision rule can be written as

$$d^*(x) = \omega_1 \Leftrightarrow g\left[a(x)\right] = \frac{1}{1 + \exp\left[a(x)\right]} \leq \frac{1}{2}$$

   where
$$a(x) = \ln\left[\frac{f(x|\omega_1)P(\omega_1)}{f(x|\omega_2)P(\omega_2)}\right].$$

   For the example of the previous question, derive the function $a(x)$ and prove that is is affine, i.e., $a(x) = a_1x + a_2$, where a_1 and a_2 are two functions of $m_1, m_2, \sigma^2, P_1, P_2$ that you will determine.

3. (4 pts) Based on the results of the previous question, we can define a so-called logistic regression classifier defined as

$$d_{\text{LR}}(x) = \omega_1 \Leftrightarrow g_{\boldsymbol{a}}(x) = \frac{1}{1 + \exp(-a_1x - a_2)} \leq \frac{1}{2}.$$

   where $\boldsymbol{a} = (a_1, a_2)^T$. In a practical application, the parameter vector $\boldsymbol{a}$ can be determined using training data from the two classes ω_1 and ω_2 denoted as $\chi = \{(x_1, y_1), (x_2, y_2),..., (x_n, y_n)\}$ where $y_i = 0$ if x_i belongs to class ω_1 and $y_i = 1$ else.

   - A first idea is to determine the vector $\boldsymbol{a}$ that minimizes the cost function

$$C_1(\chi, \boldsymbol{a}) = \frac{1}{n}\sum_{i=1}^{n}\left[g_{\boldsymbol{a}}(x_i) - y_i\right]^2.$$

     Why do you think that this cost function is not appropriate for estimating the vector $\boldsymbol{a}$?

   - Another idea is to minimize the cost function

$$C_2(\chi, \boldsymbol{a}) = \frac{1}{n}\sum_{i=1}^{n}\left\{-y_i \ln[g_{\boldsymbol{a}}(x_i)] - (1 - y_i)\ln[1 - g_{\boldsymbol{a}}(x_i)]\right\}$$

     with respect to $\boldsymbol{a}$. By considering samples from the class ω_1 (such that $y_i = 0$), analyze the value of the ith term of the cost function when $g_{\boldsymbol{a}}(x_i)$ is close to 1 or close to 0 and explain why this cost function is appropriate. Calculate the gradient of this cost function and show that the steepest descent rule can be expressed as

$$a_1^{n+1} = a_1^n - \frac{\mu}{n}\sum_{i=1}^{n}[g_{\boldsymbol{a}}(x_i) - y_i]x_i, \text{ and } a_2^{n+1} = a_2^n - \frac{\mu}{n}\sum_{i=1}^{n}[g_{\boldsymbol{a}}(x_i) - y_i].$$

Questions related to the working paper (11 points)

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

1. (1 pt) Explain why higher-order statistics (HOS) are resistant to additive colored Gaussian noise.
2. (1 pt) Express the 4th order cumulant C_{40} of the signal $y(n)$ as a function of $E[y^4(n)$ and $E[y^2(n)]$.
3. (1 pt). What is a BPSK constellation? Demonstrate that $C_{40} = -2$ for this constellation.
4. (1 pt). What is a PAM(4) constellation? Demonstrate that $C_{40} = -1.36$ for this constellation.
5. (1 pt) Explain why C_{42} is unaffected by a (deterministic) phase rotation.
6. (1 pt) ? Demonstrate Eq. (15).
7. (1 pt) Justify the decision rule (18).
8. (1 pt) In Example 3, explain why the pdf $f(g) = (1-\epsilon)f_N(g)+\epsilon f_I(g)$ corresponds to the presence of outliers in the data. What is the outlier probability for this pdf?
9. (1 pt) In Example 7, explain why the presence of frequency offset generates symbol points that are smeared along arcs.
10. (1pt) In Example 13, where does the statistics q_{LLR} come from?
11. (1pt) What kind of methods do the authors recommend when the observed data are drawn from an unknown symbol set?