1 El classico

Consider a binary classification problem. Each class $\mathcal{C}_k, k = 1, 2$ is characterized by a prior probability $P(\mathcal{C}_k)$ and a conditional density.

$$p(\boldsymbol{x}|\mathcal{C}_k) = \frac{1}{(2\pi)^{d/2}\sigma_k^d} \exp^{-\frac{\|\boldsymbol{x}-\boldsymbol{\mu}_k\|^2}{2\sigma_k^2}}, \quad \boldsymbol{x}, \boldsymbol{\mu}_k \in \mathbb{R}^d, \sigma_k \in \mathbb{R} \tag{1}$$

1. Let $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$ be the training data. We aim to estimate the parameters σ_k and $\boldsymbol{\mu}_k$ of class $\mathcal{C}_k, k = 1, 2$ by maximum likelihood estimation using the training data.
   (a) To estimate σ_k and $\boldsymbol{\mu}_k$, which samples in $\mathcal{D}$ will you consider ?
   (b) Give the expression of the likelihood.
   (c) Deduce σ_k and $\boldsymbol{\mu}_k$ by maximum likelihood estimation.

   We want to design a discrimination function of the samples using Bayesian theory. The cost of a good decision is 0 and a bad decision costs λ_s.

2. Give the expression of the conditionals risks $R(\mathcal{C}_k/\boldsymbol{x}), k = 1, 2$.
3. Deduce that the minimum risk is attained by deciding $\mathcal{C}_k$ if $P(\mathcal{C}_k|\boldsymbol{x}) > P(\mathcal{C}_\ell|\boldsymbol{x}) \ \ \forall \ell \neq k$.
4. Give the explicit expression of the decision function knowing that $P(\mathcal{C}_k) = 1/2 \ \forall k = 1, 2$ and $p(\boldsymbol{x}|\mathcal{C}_k)$ is given by Equation (1).
5. Now, we consider the reject option with cost λ_r.
   Show that we shall assign a sample $\boldsymbol{x}$ to class $\mathcal{C}_k$ if

   $$P(\mathcal{C}_k|\boldsymbol{x}) > P(\mathcal{C}_\ell|\boldsymbol{x}) \quad \forall k \neq \ell \quad \text{and} \quad P(\mathcal{C}_k|\boldsymbol{x}) > 1 - \frac{\lambda_r}{\lambda_s}$$

   What happens if $\lambda_r = 0$? Same question if $\lambda_r > \lambda_s$.

2 Bayes with log-normal conditional distribution

Let us have a classification problem of K classes. Each class $\mathcal{C}_k$ has a prior probability $P(\omega_k)$ and a conditional density function $p(x|\mathcal{C}_k)$. Suppose that each sample of $\mathcal{C}_k$ follows a log-normal distribution with $x \in \mathbb{R}^+ - \{0\}$

$$p(x|\mathcal{C}_k) = \frac{1}{x\sigma_k\sqrt{2\pi}} \exp\left(-\frac{(\ln(x) - \mu_k)^2}{2\sigma_k^2}\right) \tag{4}$$

Answer to the same questions as in the exercise 1 with this conditional distribution.

3 The distribution of networks

Let be some computer networks $\mathcal{C}_k, k = 1, \cdots, K$ with $K = 3$ that we aim to automatically classify based on their characteristics. Each class $\mathcal{C}_k$ is defined by a prior probability $P(\mathcal{C}_k)$ and a conditional density $p(x|\mathcal{C}_k)$. Here, the samples x represent event occurrence dates. The conditional distribution of each class $\mathcal{C}_k$ is an Erlang's distribution.

$$p(x|\mathcal{C}_k) = \theta_k^2 \, x \, e^{-x \, \theta_k} \, \Gamma(x) \tag{5}$$

where $\Gamma(x)$ is the function equal to 1 if $x > 0$ and 0 otherwise.

1. We have the training data $\{(x_i, y_i)\}_{i=1}^N$. For each given class we aim to estimate the parameter θ_k using maximum likelihood estimation.
   (a) Give the expression of the likelihood function.
   (b) Deduce the estimation of θ_k as a solution of the likelihood maximization problem.

2. We want to classify our data using the Bayes rule. The The cost of a good decision is 0 and a bad decision costs λ_s.
   Derive the decision function assuming that classes $\mathcal{C}_k$ are equally likely (their prior probabilities are equal).

4 Bayes' little gallop

Lets consider a binary classification problem with the following conditional probabilities :

$$\begin{aligned} p(x|C_1) &= \frac{1}{2} e^{-|x|} \\ p(x|C_2) &= e^{-2|x|} \end{aligned}$$

and the following costs : $\ell_{11} = 0$, $\ell_{22} = 0$, $\ell_{12} = 2$ and $\ell_{21} = 1$

1. Determine Bayes' decision rule and its associated risk if the prior probability of C_1 is $P(C_1) = \frac{2}{3}$.
2. Remake the same computations for $P(C_1) = \frac{1}{2}$.