Discriminant analysis

1 Classification error

Linear discriminant analysis assumes that the random variables $(X, Y) \in \mathbb{R}^p \times \{0, 1\}$ has the following distribution. For all $A \in \mathcal{B}(\mathbb{R}^p)$ and all $y \in \{0, 1\}$,

$$\mathbb{P}\left(X \in A; Y = y\right) = \pi_y \int_A g_y(x)\mathrm{d}x\,,$$

where π_0 and π_1 are positive real numbers such that $\pi_0 + \pi_1 = 1$ and g_0 (resp. g_1) is the probability density of a Gaussian random variable with mean $\mu_0 \in \mathbb{R}^d$ (resp. μ_1) and positive definite covariance matrix $\Sigma_0 \in \mathbb{R}^{d\times d}$ (resp. Σ_1). Define the classifier $h_* : \mathbb{R}^p \to \{0, 1\}$ by

$$h_* : x \mapsto \mathbb{1}_{\{\pi_1 g_1(x) > \pi_0 g_0(x)\}}\,.$$

1. Give the distribution of the random variable X and prove that

$$\mathbb{P}(h_*(X) \neq Y) = \min_{h:\mathbb{R}^p\to\{0,1\}} \left\{\mathbb{P}(h(X) \neq Y)\right\}\,.$$

2. Assume that $\mu_0 \neq \mu_1$. Prove that when $\Sigma_0 = \Sigma_1 = \Sigma$, for all $x \in \mathbb{R}^p$,

$$h_*(x) = 1 \Leftrightarrow (\mu_1 - \mu_0)^T\Sigma^{-1}\left(x - \frac{\mu_1 + \mu_0}{2}\right) > \log(\pi_0/\pi_1)\,.$$

   Provide a geometrical interpretation.

3. Prove that when $\pi_1 = \pi_0$,

$$\mathbb{P}(h_*(X) = 1|Y = 0) = \Phi(-d(\mu_1, \mu_0)/2)\,,$$

   where Φ is the cumulative distribution function of a standard Gaussian random variable and

$$d(\mu_1, \mu_0)^2 = (\mu_1 - \mu_0)^T\Sigma^{-1}(\mu_1 - \mu_0)\,.$$

4. Assume now that $\Sigma_1 \neq \Sigma_0$. What is the nature of the frontier between $\{x\,;\, h_*(x) = 1\}$ and $\{x\,;\, h_*(x) = 0\}$?

2 Maximum likelihood estimation

We assume that the joint distribution of (X, Y) belongs to a family of distributions parametrized by a vector θ with real components. For $k \in \{-1, 1\}$, write $\pi_k = \mathbb{P}(Y = k)$. Assume that conditionally on the event $\{Y = k\}$, X has a Gaussian distribution with mean $\mu_k \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$, whose density is denoted g_k. In this case, the parameter $\theta = (\pi_1, \mu_1, \mu_{-1}, \Sigma)$ belongs to the set $\Theta = [0, 1] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{d\times d}$. The parameter π_{-1} is not part of the components of θ since $\pi_{-1} = 1 - \pi_1$. In this case, the parameter $\theta = (\pi_1, \mu_1, \mu_{-1}, \Sigma)$ belongs to the set $\Theta = [0, 1] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{d\times d}$. The parameter π_{-1} is not part of the components of θ since $\pi_{-1} = 1 - \pi_1$.

When Σ and μ_1 and μ_{-1} are unknown, the discriminant analysis classifier cannot be computed explicitely. Assume that $(X_i, Y_i)_{1\leqslant i\leqslant n}$ are independent observations with the same distribution as (X, Y).

1. Write the joint loglikelihood of the observations.
2. Provide the maximum likehood estimator of θ.
3. How do you suggest to use this estimator to build a classifier ?