A. Durmus alain.durmus@polytechnique.edu

6 Discriminant analysis

Exercise 6.1. Let $\mu^\star \in \mathbb{R}^d$, $\mathbf{\Sigma}^\star$ be a symmetric positive definite matrix and $\{X_1, \ldots, X_n\}$ be a i.i.d. samples with distribution $\mathrm{N}(\mu^\star, \mathbf{\Sigma}^\star)$. We assume $n \geq d$.

(1) Show that $\hat{\mu}_n = n^{-1}\sum_{i=1}^n X_i$ is the Maximum Likelihood Estimator (MLE) of $\mu^\star$.

(2) Recall that for any square matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\nabla\psi_1(\mathbf{A}) = \mathbf{B}^\top \text{ and } \nabla\psi_2(\mathbf{A}) = (\mathbf{A}^{-1})^\top \quad , \tag{30}$$

where $\psi_1(\mathbf{A}) = \mathrm{Trace}(\mathbf{AB})$ and $\psi_2(\mathbf{A}) = \log\det\mathbf{A}$, assuming that $\mathbf{A}$ is invertible. Show that

$$\hat{\mathbf{\Sigma}}_n = \frac{1}{n}\sum_{i=1}^n (X_i - \hat{\mu})(X_i - \hat{\mu})^\top$$

is the MLE of $\mathbf{\Sigma}^\star$.

Exercise 6.2. Let (X, Y) be a pair of random variables valued in $\mathbb{R}^d \times \{0, 1\}$, such that

$$\begin{cases} \mathbb{P}(Y = 1) = \pi \in (0, 1) \\ X|Y = 0 \sim \mathrm{N}(\mu_0, \mathbf{\Sigma}_0) \\ X|Y = 1 \sim \mathrm{N}(\mu_1, \mathbf{\Sigma}_1), \end{cases}$$

with $\mu_0 \in \mathbb{R}^d$, $\mu_1 \in \mathbb{R}^d$, $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$ two symmetric positive definite matrices of size $d \times d$.

(1) What is the joint distribution of (X, Y) (the density of which is denoted $p(x, y)$)? What is the distribution of X?

(2) For all $x \in \mathbb{R}^d$, give an expression of $\eta(x) = \mathbb{E}[Y|X = x]$.

(3) Let $\mathscr{C}^\star : x \in \mathbb{R}^d \mapsto \mathbb{1}_{\eta(x)\geq 1/2}$ be the Bayes classifier and assume that $\mathbf{\Sigma}_0 = \mathbf{\Sigma}_1 = \mathbf{\Sigma}$. Show that, $\mathscr{C}^\star(x) = 1$ if and only $w^\top x + b \geq 0$, where

$$\begin{aligned} w &= \mathbf{\Sigma}^{-1}(\mu_1 - \mu_0) \\ b &= \frac{1}{2}(\mu_0^\top \mathbf{\Sigma}^{-1}\mu_0 - \mu_1^\top \mathbf{\Sigma}^{-1}\mu_1) + \log\left(\frac{\pi}{1-\pi}\right) . \end{aligned}$$

(4) Show that, $\mathscr{C}^\star(x) = 1$ if and only

$$(\mu_1 - \mu_0)^\top \mathbf{\Sigma}^{-1}\left(x - \frac{\mu_1 + \mu_0}{2}\right) \geq \log\left(\frac{\pi}{1-\pi}\right),$$

and interpret this result geometrically.

(5) In this question only, we assume that $\pi = \frac{1}{2}$. Prove that

$$\mathbb{P}(\mathscr{C}^\star(X) = 1|Y = 0) = \Phi\left(-\frac{\sqrt{(\mu_1 - \mu_0)^\top \mathbf{\Sigma}^{-1}(\mu_1 - \mu_0)}}{2}\right),$$

where Φ is the cumulative distribution function of $\mathrm{N}(0, 1)$.

Deduce the value of $\mathbb{P}(\mathscr{C}^\star(X) \neq Y)$.

(6) Using Exercise 6.1, give the MLEs of the unknown parameters.

(7) Show that, when $\mathbf{\Sigma}_0^\star \neq \mathbf{\Sigma}_1^\star$, $\mathscr{C}^\star(x) = 1$ if and only $\frac{1}{2}x^\top Q x + w^\top x + b \geq 0$, where

$$
\begin{aligned}
Q &= \mathbf{\Sigma}_0^{-1} - \mathbf{\Sigma}_1^{-1} \\
w &= \mathbf{\Sigma}_1^{-1}\mu_1 - \mathbf{\Sigma}_0^{-1}\mu_0 \\
b &= \frac{1}{2}(\mu_0^\top \mathbf{\Sigma}_0^{-1}\mu_0 - \mu_1^\top \mathbf{\Sigma}_1^{-1}\mu_1) - \frac{1}{2}\log\left(\frac{\det(\mathbf{\Sigma}_1)}{\det(\mathbf{\Sigma}_0)}\right) + \log\left(\frac{\pi}{1-\pi}\right).
\end{aligned}
$$

What is the shape of the decision boundary in this case?

Exercise 6.3. Let (X, Y) be a pair of random variables with values in $\mathbb{R}^2 \times \{0, 1\}$, such that

$$
\begin{cases}
\mathbb{P}(Y = 1) = \pi \in (0, 1) \\
X^{(1)}|Y = 0 \sim \mathrm{N}(\mu_0, \sigma_0^2) \text{ is independent from } X^{(2)}|Y = 0 \sim \mathbf{Ber}(p_0) \\
X^{(1)}|Y = 1 \sim \mathrm{N}(\mu_1, \sigma_1^2) \text{ is independent from } X^{(2)}|Y = 0 \sim \mathbf{Ber}(p_1),
\end{cases}
$$

with $\sigma_0 > 0$, $\sigma_1 > 0$, $p_0 \in (0, 1)$, $p_1 \in (0, 1)$, $\mu_0 \in \mathbb{R}$ and $\mu_1 \in \mathbb{R}$.

1. What is the joint distribution of (X, Y) (the density of which is denoted $p(x, y)$)?
2. Let $\mathscr{C}^\star : x \in \mathbb{R}^d \mapsto \mathbb{1}_{\eta(x) \geq 1/2}$ be the Bayes classifier. Show that, $\mathscr{C}^\star(x) = 1$ if and only $\left(\frac{1}{2\sigma_0^2} - \frac{1}{2\sigma_1^2}\right) {x^{(1)}}^2 + w^\top x + b \geq 0$, where

$$
\begin{aligned}
w &= \left[\left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_0}{\sigma_0^2}\right), \log\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)\right] \\
b &= \frac{\mu_0^2}{2\sigma_0^2} - \frac{\mu_1^2}{2\sigma_1^2} + \log\left(\frac{\pi}{1-\pi}\right) + \log\left(\frac{\sigma_0}{\sigma_1}\right) + \log\left(\frac{1-p_1}{1-p_0}\right).
\end{aligned}
$$