Machine Learning LAB - Optimal binary Bayesian classifer

Mahamat HAMID   Yassine KADDAMI

15 octobre 2023

1 Theoretical computations

The purpose of this lab is to study the MPE test and assess it with respect to the Neyman-pearson test. Section 1 is dedicated to the theoretical computations that you are asked in Section 2 to implement so as to run simulations and verify these theoretical results. All the results needed to write your routines, in whatever language you wish to use, are given below; in Section 1.2, these results are framed. Hence, it is suggested that you begin by carrying out the simulations of Section 2 and make the theoretical computations later. You can return your theoretical computations in latex, word or even as a photo of your hand-writen notes (if the writing and the presentation are clear).

1.1 The MPE test and its probability of error

We consider the binary hypothesis testing problem :

$$\begin{cases} \mathcal{H}_0 : X \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_N) \\ \mathcal{H}_1 : X \sim \mathcal{N}(\theta, \sigma^2 \mathbf{I}_N) \end{cases} \quad \text{where } \sigma > 0 \text{ and } \theta \in \mathbb{R}^N$$

We assume the existence of prior probabilities of occurrence π_0 and π_1 for $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Alternatively, we can also pose :

$$X = \varepsilon X_1 + (1 - \varepsilon) X_0$$

where ε, X_0 and X_1 are random variables defined in same probability space $(\Omega, \Sigma, \mathbb{P})$ such that :

- $X_0 \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_N)$ and $X_1 \sim \mathcal{N}(\theta, \sigma^2 \mathbf{I}_N)$
- ε is independant of X_1 and X_0
- $\pi_0 = \mathbb{P}(\varepsilon = 0)$ and $\pi_1 = \mathbb{P}(\varepsilon = 1)$

> **Question 1**
>
> Compute the likelihood ratio $\Lambda = p_1/p_0$ of the two hypotheses where p_1 is the pdf of X_1 and p_0 that of X_0 (see slide 17) (**2 pts**).

L'expression des densités de probabilités p_1 et p_0 est la suivante :

$$\begin{cases} p_0(x) &= \dfrac{1}{(2\pi)^{N/2}\sqrt{|\det(\sigma^2\mathbf{I}_N)|}} \exp\left\{-\dfrac{1}{2\sigma^2}x^t \cdot \mathbf{I}_N \cdot x\right\} \\ p_1(x) &= \dfrac{1}{(2\pi)^{N/2}\sqrt{|\det(\sigma^2\mathbf{I}_N)|}} \exp\left\{-\dfrac{1}{2\sigma^2}(x-\theta)^t \cdot \mathbf{I}_N \cdot (x-\theta)\right\} \end{cases}$$

Ainsi en faisant le rapport on trouve :

$$\Lambda(x) = \frac{\exp\left\{-\dfrac{1}{2\sigma^2}(x-\theta)^t \cdot \mathbf{I}_N \cdot (x-\theta)\right\}}{\exp\left\{-\dfrac{1}{2\sigma^2}x^t \cdot \mathbf{I}_N \cdot x\right\}}$$

$$= \frac{\exp\left\{-\dfrac{1}{2\sigma^2}\|x-\theta\|^2\right\}}{\exp\left\{-\dfrac{1}{2\sigma^2}\|x\|^2\right\}}$$

$$= \exp\left\{-\frac{1}{2\sigma^2}\left(\|x-\theta\|^2 - \|x\|^2\right)\right\} = \exp\left\{-\frac{1}{2\sigma^2}\left(\|x\|^2 + \|\theta\|^2 - 2x^t \cdot \theta - \|x\|^2\right)\right\}$$

D'où :

$$\boxed{\Lambda(x) = \exp\left\{\frac{1}{\sigma^2}\left(x^t \cdot \theta - \frac{\|\theta\|^2}{2}\right)\right\}}$$

> **Question 2**
>
> Show that the MPE classifier (see slides 16 and 17) is given by :
>
> $$\forall x \in \mathbb{R}^N, \quad g_{\mathrm{MPE}}(x) = \begin{cases} 1 & \text{if } x^t \cdot \theta > \sigma^2 \ln(\pi_0/\pi_1) + \|\theta\|^2/2 \\ 0 & \text{otherwise} \end{cases}$$
>
> (**2 pts**)

$$\pi_1 p_1(x) > \pi_0 p_0(x) \iff \frac{p_1}{p_0}(x) > \frac{\pi_0}{\pi_1} \quad \text{si } \pi_1, p_0(x) \text{ non nuls}$$

$$\iff \Lambda(x) > \frac{\pi_0}{\pi_1}$$

$$\iff \exp\left\{\frac{1}{\sigma^2}\left(x^t \cdot \theta - \frac{\|\theta\|^2}{2}\right)\right\} > \frac{\pi_0}{\pi_1}$$

$$\pi_1 p_1(x) > \pi_0 p_0(x) \iff x^t \cdot \theta > \sigma^2 \ln(\pi_0/\pi_1) + \|\theta\|^2/2$$

Ainsi le test g_{MPE} defini par :

$$\forall x \in \mathbb{R}^N, \quad g_{\mathrm{MPE}}(x) = \begin{cases} 1 & \text{if } \Lambda(x) > \pi_0/\pi_1 \\ 0 & \text{otherwise} \end{cases}$$

s'écrit comme suit :

$$\forall x \in \mathbb{R}^N, \quad g_{\mathrm{MPE}}(x) = \begin{cases} 1 & \text{if } x^t \cdot \theta > \sigma^2 \ln(\pi_0/\pi_1) + \|\theta\|^2/2 \\ 0 & \text{otherwise} \end{cases}$$

Show that the probability of error of the MPE test (see slides p.15) is :

$$\mathbb{P}_e(g_{\mathrm{MPE}}) = \pi_0 \left(1 - \Phi\left(\frac{1}{\rho}\ln\frac{\pi_0}{\pi_1} + \frac{\rho}{2}\right)\right) + \pi_1\Phi\left(\frac{1}{\rho}\ln\frac{\pi_0}{\pi_1} - \frac{\rho}{2}\right)$$

with $\rho = \|\theta\|/\sigma$ and Φ is the cumultaive distribution function (cdf) of the normal distribution $\mathcal{N}(0,1)$.

(**4 pts**)

Montrons d'abord que :

$$\text{Si} \quad Z \sim \mathcal{N}(a\theta, \sigma^2\mathbf{I}_N) \quad \text{alors} \quad Z^T\theta \sim \mathcal{N}(a\|\theta\|^2, \|\theta\|^2\sigma^2)$$

Soit $Z \sim \mathcal{N}(a\theta, \sigma^2\mathbf{I}_N)$. $Z^T\theta$ est gaussien puisque c'est une combinaison linéaire d'éléments d'un vecteur gaussien.

Il suffit donc de montrer que : $\begin{cases} \mathbb{E}[Z^T\theta] = a\|\theta\|^2 \\ \mathbb{V}[Z^T\theta] = \|\theta\|^2\sigma^2 \end{cases}$

$$\mathbb{E}[Z^T\theta] = \mathbb{E}\left[\sum_{i=1}^N Z_i\theta_i\right] = \sum_{i=1}^N \theta_i\mathbb{E}[Z_i] = \sum_{i=1}^N a\theta_i^2 = a\|\theta\|^2$$

$$\begin{aligned}
\mathbb{V}[Z^T\theta] &= \mathbb{E}[(Z^T\theta - a\|\theta\|^2)^2] \\
&= \mathbb{E}[(\sum_{i=1}^N (Z_i\theta_i - a\theta_i^2))^2] \\
&= \mathbb{E}[(\sum_{i=1}^N (Z_i\theta_i - a\theta_i^2))(\sum_{j=1}^N (Z_j\theta_j - a\theta_j^2))] \\
&= \mathbb{E}[(\sum_{i=1}^N \sum_{j=1}^N (Z_i\theta_i - a\theta_i^2)(Z_j\theta_j - a\theta_j^2))] \\
&= \sum_{i=1}^N \sum_{j=1}^N \theta_i\theta_j\mathbb{E}[((Z_i - a\theta_i)(Z_j - a\theta_j))]
\end{aligned}$$

Puisque la matrice de covariance de Z vaut $\sigma^2\mathbf{I}_N$, on obtient :

$$\mathbb{V}[Z^T\theta] = \sum_{i=1}^N \sum_{j=1}^N \theta_i\theta_j\sigma^2\delta_{ij} = \sigma^2\sum_{i=1}^N \theta_i^2 = \sigma^2\|\theta\|^2$$

On peut maintenant répondre à la question on utilisant le résultat démontré ci-dessus. Par définition, la probabilité d'erreur est donnée par :

$$\mathbb{P}_e(g_{\mathrm{MPE}}(X)) = \pi_0\mathbb{P}(g_{\mathrm{MPE}}(X_0) = 1) + \pi_1\mathbb{P}(g_{\mathrm{MPE}}(X_1) = 0)$$

Donc

$$\mathbb{P}_e(g_{\mathrm{MPE}}(X)) = \pi_0\mathbb{P}\left(X_0^t \cdot \theta > \sigma^2\ln(\pi_0/\pi_1) + \|\theta\|^2/2\right) + \pi_1\mathbb{P}\left(X_1^t \cdot \theta \leq \sigma^2\ln(\pi_0/\pi_1) + \|\theta\|^2/2\right)$$

Or $\begin{cases} X_0^t \cdot \theta \sim \mathcal{N}(0, \sigma^2 \|\theta\|^2) \\ X_1^t \cdot \theta \sim \mathcal{N}(\|\theta\|^2, \sigma^2 \|\theta\|^2) \end{cases} \implies \begin{cases} \dfrac{X_0^t \cdot \theta}{\sigma \|\theta\|} \qquad \sim \mathcal{N}(0,1) \\ \dfrac{X_1^t \cdot \theta - \|\theta\|^2}{\sigma \|\theta\|} \quad \sim \mathcal{N}(0,1) \end{cases}$

Donc :

$$\begin{cases} X_0^t \cdot \theta > \sigma^2 \ln(\pi_0/\pi_1) + \|\theta\|^2/2 \\ X_1^t \cdot \theta \le \sigma^2 \ln(\pi_0/\pi_1) + \|\theta\|^2/2 \end{cases} \iff \begin{cases} \dfrac{X_0^t \cdot \theta}{\sigma \|\theta\|} > \dfrac{1}{\rho} \ln \dfrac{\pi_0}{\pi_1} + \dfrac{\rho}{2} \\ \dfrac{X_1^t \cdot \theta - \|\theta\|^2}{\sigma \|\theta\|} \le \dfrac{1}{\rho} \ln \dfrac{\pi_0}{\pi_1} - \dfrac{\rho}{2} \end{cases}$$

Donc :

$$\begin{cases} \mathbb{P}\left(X_0^t \cdot \theta > \sigma^2 \ln(\pi_0/\pi_1) + \|\theta\|^2/2 \right) &= \mathbb{P}\left(\dfrac{X_0^t \cdot \theta}{\sigma \|\theta\|} > \dfrac{1}{\rho} \ln \dfrac{\pi_0}{\pi_1} + \dfrac{\rho}{2} \right) = 1 - \Phi\left(\dfrac{1}{\rho} \ln \dfrac{\pi_0}{\pi_1} + \dfrac{\rho}{2} \right) \\ \mathbb{P}\left(X_1^t \cdot \theta \le \sigma^2 \ln(\pi_0/\pi_1) + \|\theta\|^2/2 \right) &= \mathbb{P}\left(\dfrac{X_1^t \cdot \theta - \|\theta\|^2}{\sigma \|\theta\|} \le \dfrac{1}{\rho} \ln \dfrac{\pi_0}{\pi_1} - \dfrac{\rho}{2} \right) = \Phi\left(\dfrac{1}{\rho} \ln \dfrac{\pi_0}{\pi_1} - \dfrac{\rho}{2} \right) \end{cases}$$

Ce qui donne finalement :

$$\mathbb{P}_e(g_{\mathrm{MPE}}) = \pi_0 \left(1 - \Phi\left(\frac{1}{\rho} \ln \frac{\pi_0}{\pi_1} + \frac{\rho}{2} \right) \right) + \pi_1 \Phi\left(\frac{1}{\rho} \ln \frac{\pi_0}{\pi_1} - \frac{\rho}{2} \right)$$

1.2 A detour by the Neyman-Pearson theory

In this section, we apply the Neymal-Pearson (NP) test to the classification problem considered so far. We make the computations to state the results that you are asked to use in the next section to carry out simulations. It is recommended that you take some time at home to fully understand the following reasoning and calculations. For the lab session, admit the formulas given below and try to proce them later. They are not so difficult to prove and if necessary contact me for further explanations. According to your course on statistics, the NP test g_{NP}^γ with size $\gamma \in]0,1[$ to test $\mathcal{H}_0$ against $\mathcal{H}_1$ when we ignore the priors π_0 and π_1 is given by

$$\forall x \in \mathbb{R}^N, \quad g_{\mathrm{NP}}^\gamma(x) = \begin{cases} 1 & \text{if } \Lambda(x) > \lambda \\ 0 & \text{otherwise} \end{cases}$$

where, as above, Λ is the likelihood ration and λ satisfies the equation $\mathbb{P}(\Lambda(X_0) > \lambda) = \gamma$

> **Question 4**
>
> Prove the inequality $\forall A > 0, \forall \gamma \in]0;1[, \quad 1 - \Phi(A/2) \le 0{,}5 \cdot (\gamma + \Phi(\Phi^{-1}(1-\gamma) - A))$ (**1 pt**).

Considérons le cas où les probabilités à priori sont telles que : $\pi_0 = \pi_1 = 1/2$. Par définition du MPE, on a que :
$$\mathbb{P}_e(g_{\mathrm{MPE}}) \le \mathbb{P}_e(g) \tag{$*$}$$
et ce $\forall g \in \mathcal{F}(\mathbb{R}^N, \{0,1\})$.
Soit $A, \gamma \in \mathbb{R}_+^* \times]0;1[$ tels que $A = \rho = \|\theta\|/\sigma$ et γ défini une pfa comme ci-dessus. D'après $(*)$, en posant $g = g_{\mathrm{NP}}^\gamma$ on a :

$$\mathbb{P}_e(g_{\mathrm{MPE}}) \le \mathbb{P}_e(g_{\mathrm{NP}}^\gamma) \iff \frac{1}{2}\left(1 - \Phi\left(\frac{A}{2} \right) \right) + \frac{1}{2}\Phi\left(-\frac{A}{2} \right) \le \frac{1}{2} \cdot \left(\gamma + \Phi\left(\Phi^{-1}(1-\gamma) - A \right) \right)$$

Or : $\Phi(-x) = 1 - \Phi(x)$

$$\Longleftrightarrow \frac{1}{2}\left(1 - \Phi\left(\frac{A}{2}\right)\right) + \frac{1}{2}\left(1 - \Phi\left(\frac{A}{2}\right)\right) \leqslant \frac{1}{2} \cdot \left(\gamma + \Phi\left(\Phi^{-1}(1-\gamma) - A\right)\right)$$

D'où

$$1 - \Phi\left(\frac{A}{2}\right) \leqslant \frac{1}{2} \cdot \left(\gamma + \Phi\left(\Phi^{-1}(1-\gamma) - A\right)\right) \qquad \forall A \in \mathbb{R}_+^*, \gamma \in]0, 1[.$$

2 Numerical simulations

The purpose of these numerical simulations is to verify numerically the theoretical results stated above for the MPE and the NP tests.

> ### Question 5
>
> 1. Write a function with input parameters M, π_1, θ, N to generate M realizations of X as defined in (1).
> 2. Calculate the error rate of the MPE test when $N = 2$, $\theta = A\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$, A varies in $[1, 10]$, $\pi_1 = 0.2$, $\sigma = 1$. Choose M at will but don't choose a too small value, otherwise, your simulations will not fit well the theoretical results.
> 3. Calculate the error rate of the Neyman-Pearson test with size $\gamma = 10^{-3}$ when, as above, $N = 2$, $\theta = A\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$, A varies in $[1, 10]$, $\pi_1 = 0.2$.
> 4. Calculate numerically the probabilities of error of the MPE test and the NP with size $\gamma = 10^{-3}$ (cf. questions 2 and (4)) when, as above, $N = 2$, $\theta = A\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$, A varies in $[1, 10]$, and $\pi_1 = 0.2$.
> 5. On the same figure, plot the error rates and probabilities of error of these tests to verify that the theoretical results are experimentally verified by your Monte-Carlo simulations and compare the performance of the MPE and NP tests. In particular, verify numerically the inequality stated in question 4.
> 6. Can you explain the difference in performance of the Neyman-Pearson and Bayesian classifiers ?

D'après la figure obtenue ci-dessous, on voit que le test MPE a tendance à avoir un taux d'erreur et une probabilité d'erreur plus faibles par rapport au test NP. Dans le cadre de notre problème, le test MPE semble être le meilleur choix pour réduire les erreurs de classification.

6. On peut expliquer cette différence de performances par le fait que le test MPE cherche à maximiser la probabilité a posteriori pour choisir l'hypothèse la plus probable compte tenu des données et des probabilités a priori. En revanche, le test NP vise à contrôler strictement le taux d'erreur de Type I (faux positifs) avec le seuil tout en maximisant la puissance du test.

Error rates and probabilities of error of NP and MPE tests