# Machine Learning LAB - Optimal binary Bayesian classifer

Mahamat HAMID Yassine KADDAMI

12 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 :

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

We assume the existence of prior probabilities of occurrence $\pi_0$ and $\pi_1$ for $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Alternatively, we can also pose :

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

where $\varepsilon, X_0$ 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 $X_1 \sim \mathcal{N}(\theta, \sigma^2 \mathbf{I}_N)$
- $\varepsilon$ 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 :

$$\begin{aligned}
\Lambda(x) &= \frac{\exp\left\{-\dfrac{1}{2\sigma^2}(x-\theta)^t \cdot \mathbf{I}_N \cdot (x-\theta)\right\}}{\exp\left\{-\dfrac{1}{2\sigma^2}x^t \cdot \mathbf{I}_N \cdot x\right\}} \\
&= \frac{\exp\left\{-\dfrac{1}{2\sigma^2}\|x-\theta\|^2\right\}}{\exp\left\{-\dfrac{1}{2\sigma^2}\|x\|^2\right\}} \\
&= \exp\left\{-\frac{1}{2\sigma^2}\left(\|x-\theta\|^2 - \|x\|^2\right)\right\} = \exp\left\{-\frac{1}{2\sigma^2}\left(\cancel{\|x\|^2} + \|\theta\|^2 - 2x^t\cdot\theta - \cancel{\|x\|^2}\right)\right\}
\end{aligned}$$

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 :

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

(**2 pts**)

$$\begin{aligned}
\pi_1 p_1(x) > \pi_0 p_0(x) &\iff \frac{p_1}{p_0}(x) > \frac{\pi_0}{\pi_1} \qquad \text{si } \pi_1, p_0(x) \text{ non nuls} \\
&\iff \Lambda(x) > \frac{\pi_0}{\pi_1} \\
&\iff \exp\left\{\frac{1}{\sigma^2}\left(x^t\cdot\theta - \frac{\|\theta\|^2}{2}\right)\right\} > \frac{\pi_0}{\pi_1} \\
\pi_1 p_1(x) > \pi_0 p_0(x) &\iff x^t\cdot\theta > \sigma^2\ln(\pi_0/\pi_1) + \|\theta\|^2/2
\end{aligned}$$

Ainsi le test $g_{\text{MPE}}$ defini par :

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

s'écrit comme suit :

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

**Question 3**

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 $\Phi$ is the cumultaive distribution function (cdf) of the normal distribution $\mathcal{N}(0,1)$.
**(4 pts)**

Par définition, la probabilité d'erreur est donnée par :

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

Donc

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

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

Donc :

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

Donc :

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

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 $g_{\mathrm{NP}}^{\gamma}$ with size $\gamma \in ]0,1[$ to test $\mathcal{H}_0$ against $\mathcal{H}_1$ when we ignore the priors $\pi_0$ and $\pi_1$ 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, $\Lambda$ is the likelihood ration and $\lambda$ satisfies the equation $\mathbb{P}(\Lambda(X_0) > \lambda) = \gamma$

**Question 4**

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

$$\mathbb{P}_e(g_{\text{MPE}}) \leqslant \mathbb{P}_e(g) \tag{$*$}$$

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 $\gamma$ défini une pfa comme ci-dessus. D'après ($*$), en posant $g = g_{\text{NP}}^{\gamma}$ on a :

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

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

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

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.

```
# -*- coding: utf-8 -*-
"""
Created on Wed Oct 11 01:02:35 2023

@author: hamid & kaddami
"""

import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as st

sig = 1 # std
def data_generator(M, p1, theta, N) :
    """
    Parameters
    ----------
    M : int
        number of realisations.
    p1 : float between (0,1)
```

```python
        prior probability for the hypothesis H_1.
    theta : numpy float array of size Nx1
        mean of data under H_1.
    N : int
        dimension of a data vector.

    Returns
    -------
    M realization of the vector X in numpy array form of size NxM.
    """
    X_0 = np.random.normal(loc = 0, scale = sig, size = (N,M))
    X_1 = X_0 + theta
    eps = np.random.binomial(1, p1, M)

    return eps * X_1 + (1-eps) * X_0, eps

A = 2
N = 2
theta = A/np.sqrt(2) * np.ones((N,1))
p1 = .2
M = 100_000

data = data_generator(M, p1, theta, N)
X = data[0]
eps = data[1]

### MPE test error rate ###
MPE_test = X.T @ theta > sig**2 * np.log((1-p1)/p1) + np.sum(theta*theta)/2
err_rate_MPE = np.sum(np.abs(MPE_test.T - eps))/M
print(f"MPE test error rate : {err_rate_MPE:.4f}")

### NP error rate ###
gamma = 1e-3
NP_test = X.T @ theta > sig * np.linalg.norm(theta) * st.norm.ppf(1-gamma)
err_rate_NP = np.sum(np.abs(NP_test.T - eps))/M
print(f"NP test error rate : {err_rate_NP:.4f}")

### Ananlytical expression of MPE & NPE proba of error ###
rho = np.linalg.norm(theta)/sig
Pe_MPE = (1-p1) * (1 - st.norm.cdf(1/rho * np.log((1-p1)/p1) + rho/2)) + p1 * st.
    norm.cdf(1/rho * np.log((1-p1)/p1) - rho/2)
Pe_NP = (1-p1) * gamma + p1 * st.norm.cdf(st.norm.ppf(1-gamma)-rho)
print(f"MPE test error probability : {Pe_MPE:.4f}")
print(f"NP test error probability : {Pe_NP:.4f}")



#%% plots
M = [10**x for x in range(2,7)]
err_rates_MPE = np.empty(len(M))
err_rates_NP = np.empty(len(M))
```

```python
for idx, iM in enumerate(M) :
    data = data_generator(iM, p1, theta, N)
    X = data[0]
    eps = data[1]

    ### MPE test error rate ###
    MPE_test = X.T @ theta > sig**2 * np.log((1-p1)/p1) + np.sum(theta*theta)/2
    err_rates_MPE[idx] = np.sum(np.abs(MPE_test.T - eps))/iM

    ### NP error rate ###
    NP_test = X.T @ theta > sig * np.linalg.norm(theta) * st.norm.ppf(1-gamma)
    err_rates_NP[idx] = np.sum(np.abs(NP_test.T - eps))/iM

fig = plt.figure()
ax = fig.add_subplot(111)
ax.semilogx(M, err_rates_MPE, marker='o', label = 'MPE')
ax.semilogx(M, err_rates_NP, marker='o', label = 'NP')
ax.axhline(y = Pe_MPE, color = 'r', label = '$\mathbb{P}_e(g_{MPE})$')
ax.axhline(y = Pe_NP, color = 'b', label = '$\mathbb{P}_e(g_{NP})$')
plt.xlabel("Monte carlo number")
plt.ylabel("Error rate")
plt.grid(linestyle = '--')
ax.legend()
plt.savefig(fname = 'Error_rate.pdf', bbox_inches='tight')
plt.show()
```

lab_function_tests.py