FTML - Rapport

26 juin 2023

TABLE DES MATIÈRES

1 EXERCICE 1

1.1 Question 1

Bayes estimator and Bayes risk
For this question, we consider the following setting :

— $X = [0, 3]$

— $Y = \mathbb{R}$

— $X \sim U(X)$

— $Y = U[0, 1]$ if $X < 1$, $N(X)$ if $X > 1$

In this context, using the square loss, the Bayes predictor is given by :

$$f^*(x) = \begin{cases} \frac{1}{2} & \text{si } X < 1 \\ X & \text{si } X > 1 \end{cases}$$

The associated Bayes Risk is calculated as :

$$R^* = E[l(Y, f^*(x))] = \frac{1}{3} E\left[(U[0, 1] - \frac{1}{2})^2\right] + \frac{2}{3} E\left[(N(x, 1) - x)^2\right]$$

We calculate each of these expectations.

$$E[(U[-\frac{1}{2}, \frac{1}{2}])^2] = E[(U[0, \frac{1}{2}])]^2 = (\frac{1}{4})^2 = \frac{1}{16}$$

$$E[(N(x,1)-x)^2] = E[(N[0,1])^2] = V(N[0,1]) = 1$$

At the end, we have :

$$R^* = \frac{1}{3} \times \frac{1}{16} + \frac{2}{3} \times 1 = \frac{1}{48} + \frac{2}{3} \approx 0.70$$

1.2 Question 2

Using the same chosen setting, we propose the following estimator :

$$\tilde{f} = \begin{cases} X \to Y \\ x \to X \end{cases}$$

After running a simulation, we obtain a value for our estimator $\tilde{f}$ of approximately 0.70, which is logically slightly above the Bayes Risk.

Result of the simulation :

With statistical approximation of the tilde risk which is : 0.779019
With the bayes risk which is : 0.7083333

2 EXERCICE 2

2.1 Question 1

We want a configuration where the Bayes predictor is different depending on the use of the square loss or the absolute loss. We have seen that the Bayes predictor in the case of the square loss is the conditional expectation of Y given X. In the case of the absolute loss, it is the median of Y given X.

Therefore, we want to choose a probability distribution where the median is different from the expectation.

Let's consider the following model :

1. X follows a uniform distribution on [0.1, 0.5].
2. Y follows a geometric distribution with parameter X.

For the square loss :

$$f^*(x) = E[Y|X] = 1/X \tag{1}$$

$$\begin{aligned}
R(f^*(x)) &= E_{X,Y}[l(y, f^*(x))] \\
&= E_X[E_Y[l(y, f^*(x))|X = x]] \\
&= E_X[E_Y[(y - \frac{1}{x})^2|X = x]] \\
&= E_X[\sum_{y=1}^{+\infty} (y - \frac{1}{x})^2 (1-x)^{y-1} x] \\
&= \int_{0.1}^{0.5} (\sum_{y=1}^{+\infty} (y - \frac{1}{x})^2 (1-x)^{y-1} x) \frac{1}{0.5 - 0.1} dx
\end{aligned} \tag{2}$$

For absolute loss :

$$\begin{aligned} f^*(x) &= Mediane Y|X \\ &= \lceil \frac{-\log(2)}{\log(1-X)} \rceil \end{aligned} \tag{3}$$

$$\begin{aligned} R(f^*(x)) &= E_{X,Y}[l(y, f^*(x))] \\ &= E_X[E_Y[l(y, f^*(x))|X = x]] \\ &= E_X[E_Y[|y - \lceil \frac{-\log(2)}{\log(1-X)} \rceil||X = x]] \\ &= E_X[\sum_{y=1}^{+\infty} |y - \lceil \frac{-\log(2)}{\log(1-X)} \rceil|P(Y = y|X = x)] \\ &= E_X[\sum_{y=1}^{\lfloor(\frac{1}{x})} (\lceil \frac{-\log(2)}{\log(1-X)} \rceil - y)P(Y = y|X = x) + \sum_{y=\lfloor \frac{1}{x} \rfloor}^{+\infty} (y - \lceil \frac{-\log(2)}{\log(1-X)} \rceil)P(Y = y|X = x)] \\ &= \int_{0.1}^{0.9} (\sum_{y=1}^{\lfloor(\frac{1}{x})} (\lceil \frac{-\log(2)}{\log(1-X)} \rceil - y)(1-x)^{y-1}x + \sum_{y=\lfloor \frac{1}{x} \rfloor}^{+\infty} (y - \lceil \frac{-\log(2)}{\log(1-X)} \rceil)(1-x)^{y-1}x) \frac{1}{0.9 - 0.1} dx \end{aligned} \tag{4}$$

With the square loss :

- Bad estimator : 32.17904129531603
- Absolute Bayes predictor : 17.482005
- Squared Bayes predictor : 15.921225286059489

With the absolute loss :

- Bad estimator : 3.5219489510437 66
- Absolute Bayes predictor : 2.390873
- Squared Bayes predictor : 2.5562641799621604

After simulation :
Using the square loss, the empirical risk for the bad estimator is 32.556, indicating its poor performance compared to the true values. In contrast, the squared Bayes predictor achieves a significantly lower empirical risk of 16.135, demonstrating its superior performance. Similarly, the empirical risk for the absolute Bayes predictor is 17.738, also outperforming the bad estimator.

When considering the absolute loss, the empirical risk for the bad estimator reduces to 3.535, but it still falls short compared to the true values. On the other hand, the squared Bayes predictor achieves a lower empirical risk of 2.563, showcasing its improved performance in minimizing loss. Remarkably, the absolute Bayes predictor achieves the lowest empirical risk of 2.399, indicating its excellent capability in predicting the true values.

These results highlight the significance of choosing appropriate loss functions and estimators for achieving accurate predictions. The empirical risks clearly demonstrate the superior performance of the Bayes predictors.

2.2 Question 2

$$
\begin{aligned}
f^*(x) &= argmin E[|y-z||X=x] \\
E[|y-z||X=x] &= \int_{-\infty}^{+\infty} |y-z| p_{Y|X=x} dy \\
&= \int_{-\infty}^{z} (z-y) p_{Y|X=x} dy + \int_{z}^{+\infty} (y-z) p_{Y|X=x} dy \\
&= \int_{-\infty}^{z} -(y-z) p_{Y|X=x} dy + \int_{z}^{+\infty} (y-z) p_{Y|X=x} dy
\end{aligned} \tag{5}
$$

To find the minimum, we look for when the derivative equals zero.

$$
\begin{aligned}
\frac{\partial E[|y-z||X=x]}{\partial z} &= \int_{-\infty}^{z} p_{Y|X=x} dy + \int_{z}^{+\infty} -p_{Y|X=x} dy = 0 \\
\int_{-\infty}^{z} p_{Y|X=x} dy &= \int_{z}^{+\infty} -p_{Y|X=x} dy \\
\implies 2 * \int_{-\infty}^{z} p_{Y|X=x} dy &= \int_{-\infty}^{+\infty} p_{Y|X=x} dy = 1 \\
\int_{-\infty}^{z} p_{Y|X=x} dy &= \frac{1}{2} \\
\text{so } F_{Y|X}(z) &= \frac{1}{2}
\end{aligned} \tag{6}
$$

Let's verify that it is indeed a minimum.

$$
\begin{aligned}
\frac{\partial E[|y-z||X=x]}{\partial z} = \int_{-\infty}^{z} p_{Y|X=x} dy - \int_{z}^{+\infty} p_{Y|X=x} dy &< 0 \\
F_{Y|X}(z) - (1 - F_{Y|X}(z)) &< 0 \\
2 * F_{Y|X}(z) &< 1 \\
F_{Y|X}(z) &< \frac{1}{2}
\end{aligned} \tag{7}
$$

3 EXERCICE 3

3.1 Question 1

Show that :

$$
E\left[R_n(\hat{\theta})\right] = E_\epsilon\left[\frac{1}{n}||(I_n - X(X^T X)^{-1} X^T)\epsilon||^2\right] \tag{8}
$$

Where E_ϵ means that the expected value is over ϵ.

Démonstration.

$$
R_n(\theta) = \frac{1}{n}||y - X\theta||_2^2 \tag{a}
$$

$$
\hat{\theta} = (X^T X)^{-1} X^T y \tag{b}
$$

$$
y = X\theta + \epsilon \tag{c}
$$

So according to (a) and (b)

$$\begin{aligned} R_n(\hat{\theta}) &= \frac{1}{n}||y - X\theta||^2 \\ &= \frac{1}{n}||y - X(X^TX)^{-1}X^Ty||^2 \\ &= \frac{1}{n}||(I_n - X(X^TX)^{-1}X^T)y||^2 \end{aligned}$$

Soit $A = I_n - X(X^TX)^{-1}X^T$ la matrice de projection orthogonale sur $E^{\perp}$.

$$\begin{aligned} R_n(\hat{\theta}) &= \frac{1}{n}||Ay||^2 \\ &= \frac{1}{n}y^TA^TAy \end{aligned}$$

$$\begin{aligned} E\left[R_n(\hat{\theta})\right] &= E\left[\frac{1}{n}y^TA^TAy\right] \\ &= \frac{1}{n}E\left[\epsilon^TA^TA\epsilon\right] \\ &= \frac{1}{n}E\left[||A\epsilon||^2\right] \end{aligned}$$

In conclusion, we find (8)

$$\begin{aligned} E\left[R_n(\hat{\theta})\right] &= \frac{1}{n}E\left[||(I_n - X(X^TX)^{-1}X^T)\epsilon||^2\right] \\ &= E\left[\frac{1}{n}||(I_n - X(X^TX)^{-1}X^T)\epsilon||^2\right] \end{aligned}$$

□

3.2 Question 2

Let $A \in \mathbb{R}^{n,n}$. Show that

$$\sum_{(i,j)\in[1,n]^2} A_{ij}^2 = \mathrm{Tr}(A^TA) \tag{9}$$

Démonstration.

$$\mathrm{Tr}(A) = \sum_{i\in[1,n]} a_{i\,i} \tag{a}$$

$$\mathrm{Tr}(A^TA) = \sum_{i\in[1,n]} (A^TA)_{ii}$$

Because A is a square matrix

$$\begin{aligned} (A^TA)_{ii} &= \sum_{j\in[1,n]} (A^T)_{ij}A_{ji} \\ &= \sum_{j\in[1,n]} A_{ji}A_{ji} \\ &= \sum_{j\in[1,n]} (A_{ji})^2 \end{aligned}$$

$$\begin{aligned} \mathrm{Tr}(A^TA) &= \sum_{i\in[1,n]}\sum_{j\in[1,n]} (A_{ji})^2 \\ &= \sum_{(i,j)\in[1,n]^2} A_{ij}^2 \end{aligned}$$

□

3.3 Question 3

Show that

$$E_\epsilon\left[\frac{1}{n}\|A\epsilon\|^2\right] = \frac{\sigma^2}{n}\mathrm{Tr}(A^TA) \tag{10}$$

Démonstration.

$$E_\epsilon\left[\frac{1}{n}\|A\epsilon\|^2\right] = \frac{1}{n}\|A\|^2E_\epsilon\left[\|\epsilon\|^2\right]$$

Since

$$Var(X) = E(X^2) - E^2(X) \tag{a}$$

and

$$E_\epsilon(\epsilon) = 0 \tag{b}$$

$$\begin{aligned} E_\epsilon\left[\frac{1}{n}\|A\epsilon\|^2\right] &= \frac{1}{n}\|A\|^2E_\epsilon\left[\epsilon^2\right] \\ &= \frac{\sigma^2}{n}\|A\|^2 \end{aligned}$$

Because of (9)

$$E_\epsilon\left[\frac{1}{n}\|A\epsilon\|^2\right] = \frac{\sigma^2}{n}\mathrm{Tr}(A^TA)$$

□

3.4 Question 4

We note

$$A = I_n - X(X^TX)^{-1}X^T \tag{11}$$

Show that

$$A^TA = A \tag{12}$$

Démonstration.

$$\begin{aligned} A^T &= (I_n - X(X^TX)^{-1}X^T)^T \\ &= I_n^T - ((X(X^TX)^{-1})X^T)^T \\ &= I_n - X(X(X^TX)^{-1})^T \\ &= I_n - X(X^TX)^{-T}X^T \\ &= I_n - X(X^TX)^{-1}X^T \\ &= A \end{aligned}$$

$$\begin{aligned} A^TA &= (I_n - X(X^TX)^{-1}X^T)^2 \\ &= I_n - X(X^TX)^{-1}X^T - X(X^TX)^{-1}X^T + X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T \\ &= I_n - X(X^TX)^{-1}X^T - X(X^TX)^{-1}X^T + X(X^TX)^{-1}X^T \\ &= I_n - X(X^TX)^{-1}X^T \\ &= A \end{aligned}$$

□

3.5 Question 5

Conclude that

$$\mathbb{E}\left[R_n(\hat{\theta})\right] = \frac{n-d}{n}\sigma^2 \tag{13}$$

Démonstration. We proved that (8) and (10) so

$$\mathbb{E}\left[R_n(\hat{\theta})\right] = \frac{\sigma^2}{n}\mathrm{Tr}(A^T A)$$

And with (11)

$$\begin{aligned}
\mathbb{E}\left[R_n(\hat{\theta})\right] &= \frac{\sigma^2}{n}\mathrm{Tr}(A) \\
&= \frac{\sigma^2}{n}\mathrm{Tr}(I_n - X(X^T X)^{-1} X^T)
\end{aligned}$$

$$\mathrm{Tr}(A+B) = \mathrm{Tr}(A) + \mathrm{Tr}(B) \tag{a}$$

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA) \tag{b}$$

$$\begin{aligned}
\mathbb{E}\left[R_n(\hat{\theta})\right] &= \frac{\sigma^2}{n}(\mathrm{Tr}(I_n) - \mathrm{Tr}(X(X^T X)^{-1} X^T)) \\
&= \frac{\sigma^2}{n}(n - \mathrm{Tr}(X^T X(X^T X)^{-1})) \\
&= \frac{\sigma^2}{n}(n - \mathrm{Tr}(I_d)) \\
&= \frac{\sigma^2}{n}(n-d) \\
&= \frac{n-d}{n}\sigma^2
\end{aligned}$$

□

3.6 Question 6

Still in the same setting, what is the expected value of $\frac{\|y - X\hat{\theta}\|_2^2}{n-d}$

$$\begin{aligned}
\mathbb{E}\left[\frac{\|y - X\hat{\theta}\|_2^2}{n-d}\right] &= \frac{n}{n-d}\mathbb{E}\left[R_n(\hat{\theta})\right] \\
&= \frac{n}{n-d}\frac{n-d}{n}\sigma^2 \\
&= \sigma^2
\end{aligned}$$

4 EXERCICE 4

For this particular topic, you will find all the explanations in the corresponding notebooks.

5 EXERCICE 5

For this particular topic, you will find all the explanations in the corresponding notebooks.