FTML

Corentin Pion

June 2022

1 Exercice 1

1.1 Question 1

1.2 Question 2

2 Exercice 2

We consider a supervised regression problem, Y = R. We have seen that when the loss used is the square loss l2(y, z) = (y z), then the Bayes predictor is the conditional expectation

2.1 Question 1

Subject : Propose a setting where the Bayes predictor is different for the square loss and for the absolute loss.

We could use the continuous distribution of Exp() with the a theta parameter

First, let's calculate a bayes estimator with the squared loss function

$$f*(x) = E[Y|X = x] = \frac{1}{\theta}$$

When we use the squared error we get the Esp() of the exponential
Now if we switch to an absolute loss function we have

$$f*(x) = E[|y - z||X = x] = \frac{ln(2)}{\theta}$$

We can conclude that it's possible that the bayes estimator can change if the loss function is changed

2.2 Question 2

3 Exercice 3

We want to show that in the linear model, fixed design we have

$$E[R_x(\widehat{\theta})] = \frac{n-d}{n}\sigma^2$$

In this expression, both y and ˆ are random variables, that are not independent, as theta is the OLS estimator. The expectation is over the distribution of both variables.

3.1 step 1

Show that :

$$E[R_n(\widehat{\theta})] = E_\epsilon[\frac{1}{n}\|I_n - X(X^tX)^{-1}X^t)\epsilon\|^2]$$

$$E[R_n(\widehat{\theta})] = E[\frac{1}{n}\|y - X\widehat{\theta}\|^2]$$

$$E[R_n(\widehat{\theta})] = E[\frac{1}{n}\|X\theta + \epsilon - X\widehat{\theta}\|^2]$$

$$E[R_n(\widehat{\theta})] = E[\frac{1}{n}\|X\theta + \epsilon - X(X^tX)^{-1}X^ty\|^2]$$

$$E[R_n(\widehat{\theta})] = E[\frac{1}{n}\|X\theta + \epsilon - X(X^tX)^{-1}X^t(X\theta + \epsilon)\|^2]$$

$$E[R_n(\widehat{\theta})] = E_\Theta[\frac{1}{n}\|X\theta - X(X^tX)^{-1}X^tX\theta\|^2] + E_\epsilon[\frac{1}{n}\|\epsilon - X(X^tX)^{-1}X^t\epsilon\|^2]$$

$$E[R_n(\widehat{\theta})] = E_\Theta[\frac{1}{n}\|0\|^2] + E_\epsilon[\frac{1}{n}\|(I_n - X(X^tX)^{-1}X^t)\epsilon\|^2]$$

$$E[R_n(\widehat{\theta})] = E_\epsilon[\frac{1}{n}\|(I_n - X(X^tX)^{-1}X^t)\epsilon\|^2]$$

3.2 step 2

Let A Rn,n. Show that

$$\sum_{i=1}^{n} A_{ij}^2 = tr(A^tA)$$

To show this demonstration wxe can use the definition of the scalaire and the trace function

$$tr(A^tA) =< A, A >= \sum_{i=1}^{n} A_{ij}^2$$

3.3 step 3

Show that

$$E[\frac{1}{n}\|A\epsilon\|^2] = \frac{\sigma^2}{n}tr(A^tA)$$

$$E[\frac{1}{n}\|A\epsilon\|^2] = E[\frac{1}{n}\sum_{i=1}^{n}(A_{ij}\epsilon)^2]$$

$$E[\frac{1}{n}\|A\epsilon\|^2] = \frac{1}{n}tr(A^tA)E[\epsilon^2]$$

But

$$E[\epsilon^2] = E[\epsilon^2 - E[\epsilon]^2] = Var = \sigma^2$$

Since in the subject its said that E(epsilon) is actualy equal to 0 So in the end we have

$$E[\frac{1}{n}\|A\epsilon\|^2] = \frac{\sigma^2}{n}tr(A^tA)$$

3.4 step 4

We note

$$A = I_n - X(X^tX)^{-1}X^t$$

Show that

$$A^tA = A$$

$$A^tA = (I_n - X(X^tX)^{-1}X^t)^t(I_n - X(X^tX)^{-1}X^t)$$

$$A^tA = A$$

3.5 step 5

We can now show that our proposition in the start is true

$$E[R_x(\widehat{\theta})] = E[[E[R_n(\widehat{\theta})]]$$

we have

$$[E[R_n(\widehat{\theta})] = \frac{\sigma^2}{n}tr(I_n - X(X^tX)^{-1}X^t)$$

we now use trace proprieties tr(a+b) = tr(a) + tr(b):

$$[E[R_n(\widehat{\theta})] = \frac{\sigma^2}{n}tr(I_n) - tr(X(X^tX)^{-1}X^t)$$

we now use trace proprieties tr(abc) = tr(cab)):

$$[E[R_n(\widehat{\theta})] = \frac{\sigma^2}{n}tr(I_n) - tr(X^tX(X^tX)^{-1})$$

$$[E[R_n(\widehat{\theta})] = \frac{\sigma^2}{n} tr(I_n) - tr(I_d)$$

$$[E[R_n(\widehat{\theta})] = \frac{\sigma^2}{n}(n-d)$$

if we take back our first equation:

$$E[R_x(\widehat{\theta})] = E[[E[R_n(\widehat{\theta})]] = \frac{\sigma^2}{n}(n-d)$$