FTML 2025

Antoine Malmezac
Cyprien Deruelle
Lucas Collemare
Charles Prioux

3 Partie 3 : Valeur espérée du risque empirique pour OLS

3.1 Question 1

La Proposition 1 dit que $E[R_X(\hat{\theta})] = \frac{n-d}{n}\sigma^2$.

Le risque de Bayes dans ce contexte est σ^2 (la variance du bruit).

Donc l'estimateur OLS a un risque plus petit que le risque de Bayes ! C'est normal car on travaille en « design fixe » : on conditionne sur X, ce qui réduit l'incertitude.

Quand n devient grand, le facteur $\frac{n-d}{n}$ tend vers 1, et on retrouve le risque de Bayes.

3.2 Questions 2-6 : Démonstration de la Proposition

Question 2 : On part de

$$R_n(\hat{\theta}) = \frac{1}{n}\|y - X\hat{\theta}\|^2.$$

Avec $y = X\theta^* + \epsilon$ et $\hat{\theta} = (X^TX)^{-1}X^Ty$, on obtient :

$$y - X\hat{\theta} = (I_n - X(X^TX)^{-1}X^T)\epsilon.$$

Donc

$$R_n(\hat{\theta}) = \frac{1}{n}\|(I_n - X(X^TX)^{-1}X^T)\epsilon\|^2.$$

Question 3 : Pour une matrice A, on a

$$\sum_{i,j} A_{ij}^2 = \mathrm{tr}(A^TA)$$

par définition de la trace.

Question 4 : Si $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$, alors :

$$E[\|A\epsilon\|^2] = E[\epsilon^TA^TA\epsilon] = \sigma^2\mathrm{tr}(A^TA).$$

Question 5 : La matrice $A = I_n - P$ où $P = X(X^TX)^{-1}X^T$ est la projection orthogonale.

Comme P est idempotente et symétrique, on a $A^TA = A$.

Question 6 : En combinant tout :

$$E[R_n(\hat{\theta})] = \frac{\sigma^2}{n}\mathrm{tr}(A) = \frac{\sigma^2}{n}(n-d) = \frac{n-d}{n}\sigma^2.$$

3.3 Question 7

On a montré que $E[\|y - X\hat{\theta}\|^2] = (n-d)\sigma^2$.
Donc : $E\left[\frac{\|y-X\hat{\theta}\|^2}{n-d}\right] = \sigma^2$
C'est pourquoi $\hat{\sigma}^2 = \frac{\|y-X\hat{\theta}\|^2}{n-d}$ est un estimateur sans biais de σ^2.