Problème – Minimisation de l'erreur de prédiction

Dans ce problème, on s'intéresse à l'erreur de prédiction dans le modèle linéaire. On travaille ici avec un modèle linéaire, non supposé gaussien, et on note σ^2 la variance (constante) du bruit.

Première partie : cas à une variable explicative. On cherche à apprendre le modèle linéaire suivant pour prédire le réel Y_i en fonction d'une variable explicative réelle z_i :

$$\forall 1 \leq i \leq n, \quad Y_i = \beta_0 + \beta_1 z_i + \varepsilon_i \,.$$

On notera $\overline{y}$ et $\overline{z}$ les moyennes empiriques de $z = (z_1, \ldots, z_n)^T$ et de $Y = (Y_1, \ldots, Y_n)^T$ (attention, elles dépendent de n). On suppose que le modèle est identifiable et on note $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ l'estimateur des moindres carrés de β.

Sous le même modèle, on observe une nouvelle valeur z_{n+1} de la variable explicative et on cherche à prédire la variable réponse Y_{n+1} avec l'estimateur

$$\hat{y}_{n+1} := \hat{\beta}_0 + \hat{\beta}_1 z_{n+1} = x_{n+1}^T \hat{\beta}, \quad \text{où } x_{n+1} := \begin{bmatrix} 1 \\ z_{n+1} \end{bmatrix} \,.$$

L'erreur de prédiction est définie par :

$$\mathrm{err}(z_{n+1}) := \mathbb{E}\left[(Y_{n+1} - \hat{y}_{n+1})^2\right] \,.$$

Notons que dans cette espérance, l'aléa vient du bruit ε_{n+1} dans Y_{n+1} ainsi que de l'aléa dans le $\hat{\beta}$.

1. Redémontrer que si S est un vecteur aléatoire de $\mathbb{R}^m$ de matrice de covariance (finie) C, alors pour tout vecteur $u \in \mathbb{R}^m$, $\mathrm{Var}(u^T S) = u^T C u$.

2. Que vaut $\mathbb{E}\left[Y_{n+1} - \hat{y}_{n+1}\right]$?

3. Ecrire la matrice X du plan d'expérience dans ce modèle (on mettra l'intercept dans la première colonne), et montrer que

$$(X^T X)^{-1} = \frac{1}{n v(z)} \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n z_i^2 & -\overline{z} \\ -\overline{z} & 1 \end{bmatrix},$$

   avec $v(z)$ la variance empirique de z.

4. En utilisant les résultats des questions 1, 2 et 3, montrer que

$$\mathrm{err}(z_{n+1}) = \sigma^2 \left(1 + \frac{1}{n} + \frac{(z_{n+1} - \overline{z})^2}{\sum_{i=1}^n (z_i - \overline{z})^2}\right) \,.$$

5. Pour quelle(s) valeur(s) de z_{n+1} l'erreur de prédiction est-elle minimale ? Interpréter ce résultat.

6. Quelle est la limite de l'erreur de prédiction minimale lorsque $n \to \infty$? Interpréter cette valeur limite.

Deuxième partie : cas général. Nous considérons cette fois-ci le cas général à plusieurs covaraibles. Ici, pour tout $1 \leq i \leq n$,

$$Y_i = \beta_0 + \beta_1 z_{1,i} + \ldots + \beta_p z_{p,i} + \varepsilon_i \,.$$

On note

$$Z := \left[\begin{array}{c|c|c} z_1 & \dots & z_p \end{array}\right] \in \mathbb{R}^{n\times p},$$

avec pour tout $1 \leq \ell \leq p$, $z_\ell := \begin{bmatrix} z_{\ell,1} & \dots & z_{\ell,n}\end{bmatrix}^T \in \mathbb{R}^n$. On note $\overline{z}$ le vecteur des moyennes empiriques $\overline{z} := \begin{bmatrix}\overline{z_1} & \dots & \overline{z_p}\end{bmatrix}^T \in \mathbb{R}^p$ (dont les entrées dépendent de n).

7. En notant $\beta = \begin{bmatrix}\beta_0 & \beta_1 & \dots & \beta_p\end{bmatrix}^T$, écrire le modèle matriciellement sous la forme $Y = X\beta + \varepsilon$ en précisant X.

8. Ecrire la matrice X^TX sous forme de 4 blocs faisant intervenir Z, $\overline{z}$ et n.

Dans toute la suite, on suppose que le modèle est identifiable.

9. On donne la formule d'inversion matricielle par blocs : soit A une matrice inversible s'écrivant par blocs $A = \begin{bmatrix} T & U \\ V & W \end{bmatrix}$ avec T inversible. Alors $Q = W - VT^{-1}U$ est inversible et l'inverse de A est :

$$A^{-1} = \begin{bmatrix} T^{-1} + T^{-1}UQ^{-1}VT^{-1} & -T^{-1}UQ^{-1} \\ -Q^{-1}VT^{-1} & Q^{-1} \end{bmatrix}.$$

   On note $\Gamma := \frac{1}{n}Z^TZ - \overline{z}\,\overline{z}^T$. Ecrire la matrice $(X^TX)^{-1}$ sous la forme d'une matrice par blocs en fonction de $n, \overline{z}$ et Γ^{-1}.

10. Comme dans la première partie, on observe désormais un vecteur

$$x_{n+1} = (1, z_{n+1}) = (1, z_{1,n+1}, \dots z_{\ell,n+1})$$

   et l'on cherche à prédire la variable réponse Y_{n+1} avec l'estimateur $\hat{y}_{n+1} := x_{n+1}^T\hat{\beta}$. L'erreur de prédiction est définie comme précédemment.

   Exprimer $\mathrm{err}(z_{n+1})$ en fonction de n, de σ^2, du vecteur $(z_{n+1} - \overline{z})$ et de la matrice Γ^{-1}.

11. On admet dans cette question que Γ est symétrique définie positive. Pour quelle(s) valeur(s) de z_{n+1} l'erreur de prédiction est-elle minimale ?

12. ($\star$) Montrer que $\Gamma = \frac{1}{n}Z^TZ - \overline{z}\,\overline{z}^T$ est bien symétrique définie positive. *On pourra chercher à l'écrire sous la forme $\frac{1}{n}WW^T$ avec W une matrice bien choisie.*