Modélisation statistique — Régression linéaire multiple

Exercice 1 (Régression linéaire multiple)
On considère le modèle de régression linéaire multiple

$$Y = X\beta + \epsilon,$$

avec $\beta \in \mathbb{R}^p$ inconnu et $\epsilon \sim \mathcal{N}(0, \Sigma)$. On suppose Σ connue et de rang plein (de rang n), mais pas nécessairement diagonale. On suppose de plus que $X \in \mathbb{R}^{n\times p}$ est de rang plein.

1) Considérons l'estimateur des moindres carrés ordinaire $\hat{\beta} = (X^\top X)^{-1}X^\top Y$. Calculer l'espérance et la matrice de variance-covariance de $\hat{\beta}$ sous les hypothèses données dans cet exercice.
2) Justifier l'existence d'une matrice Ω de taille $n \times n$ telle que $\Sigma = \Omega^\top \Omega$.
3) Montrer que $\Omega^{-1}X$ est de rang plein.
4) Soit $Y^* = \Omega^{-1}Y$, $X^* = \Omega^{-1}X$ et $\epsilon^* = \Omega^{-1}\epsilon$. Prouver que nous obtenons un nouveau modèle qui satisfait
   (a) $\mathbb{E}[\epsilon_i^*] = 0$,
   (b) $\mathrm{Var}[\epsilon_i^*] = \sigma^2 > 0$ (erreurs de variance constante),
   (c) $\mathrm{Cov}(\epsilon_i^*, \epsilon_j^*) = 0$ pour tout $i \neq j$.
5) Déduisez un "meilleur" estimateur fonction de X, Y, Σ. Nous désignons par $\hat{\beta}_G$ cet estimateur.
6) Calculer $\mathbb{E}(\hat{\beta}_G)$ et $\mathrm{Var}(\hat{\beta}_G)$.
7) Montrer que $\hat{\beta}_G$ est optimal parmi tous les estimateurs non biaisés T, c'est-à-dire que pour tout T
$$\mathrm{Var}(T) - \mathrm{Var}(\hat{\beta}_G)$$
   est définie positive.
8) Conclure par une comparaison avec l'estimateur ordinaire des moindres carrés.

Exercice 2 (Régression ridge)

On considère le modèle linéaire Gaussien

$$Y = X\beta + \epsilon,$$

où $Y \in \mathbb{R}^n$ est le vecteur de réponses, $X \in \mathbb{R}^{n\times p}$ est la matrice de design (prédicteurs), $\beta \in \mathbb{R}^p$ est le vecteur (inconnu) de coefficients de régression, et $\epsilon \sim \mathcal{N}(0, \sigma^2 I_n)$. Soit $\lambda > 0$, on définit l'estimateur de la régression ridge par :

$$\hat{\beta}_\lambda^R = \mathrm{argmin}_{\beta\in\mathbb{R}^p} \|Y - X\beta\|^2 + \lambda\|\beta\|^2.$$

1) Montrer que l'estimateur $\hat{\beta}_\lambda^R$ vérifie la relation

$$(X^\top X + \lambda I_p)\hat{\beta}_\lambda^R = X^\top Y.$$

2) Montrer que $X^\top X + \lambda I_p$ est inversible. En déduire une expression pour l'estimateur $\hat{\beta}_\lambda^R$.

3) Donner la limite de $\hat{\beta}_\lambda^R$ lorsque $\lambda \to 0$ et lorsque $\lambda \to \infty$.

4) Calculer le biais de l'estimateur ridge $\hat{\beta}_\lambda^R$.

5) Calculer la variance de l'estimateur ridge $\hat{\beta}_\lambda^R$.

6) Supposons que X est de rang plein (i.e. $X^\top X$ est de rang p). On souhaite comparer la variance de l'estimateur des moindres carrés classique $\hat{\beta}$ à l'estimateur ridge $\hat{\beta}_\lambda^R$.

   (a) Rappeler la formule donnant l'expression de l'estimateur des moindres carrés.

   (b) Définissons $P = \frac{1}{\lambda} X^\top X$. Montrer que

   $$\mathrm{Cov}(\beta) - \mathrm{Cov}(\hat{\beta}_\lambda^R) = \sigma^2 \frac{1}{\lambda}(P + I_p)^{-1}[I_p + P^{-1} - P(P + I_p)^{-1}].$$

   (c) Montrer que la matrice au membre de droite de l'équation ci-dessus est semi-définie positive (i.e. $x^\top A x \geq 0$ pour tout x).

   (d) Quel est selon vous l'avantage principal de l'estimateur ridge ?