
QUIZ FINAL : Modèles linéaires

- 1) Soit $Y_1 = X_1^\top \beta + \epsilon_1$ avec $X_1 \in \mathbb{R}^p$ et $\epsilon_1 \in \mathbb{R}$ des variables aléatoires indépendantes, $\beta \in \mathbb{R}^p$. Que vaut $\text{Cov}(X_1, Y_1)$? Exprimer β lorsque $\text{Cov}(X_1)$ est inversible.
- 2) Que vaut le biais de l'estimateur de $\text{Cov}(x_1, y_1)$ défini par $\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_n)(x_i - \bar{x}_n)$, où l'on note \bar{y}_n, \bar{x}_n les moyennes empiriques, pour $(x_i, y_i)_{i=1, \dots, n} \in \mathbb{R}^2$ i.i.d.
- 3) On suppose que l'on observe $(x_1, y_1), \dots, (x_n, y_n)$, des variables réelles i.i.d., centrées et de variance σ^2 . Quel est le risque quadratique de l'estimateur $\frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n x_i y_j$ de $\mathbb{E}[x_1] \mathbb{E}[y_1]$? Quel est le biais de l'estimateur $\bar{x}_n \bar{y}_n$?
- 4) Soit x_1, x_2, \dots, x_n i.i.d. tel que $\mathbb{E}[x_1^2] < \infty$. Quel estimateur $\hat{\mu}$ minimise $\sum_{i=1}^n w_i (x_i - \mu)^2$ pour des $w_i > 0$ quelque soit $i \in \llbracket 1, n \rrbracket$? Donner son biais et sa variance, pour tout $n > 1$ (les w_i étant fixés et déterministes).
- 5) La fonction $(\theta_0, \theta_1) \rightarrow \frac{1}{2} \sum_{i=1}^n y_i^2 \exp(\theta_0 - \theta_1 x_i)$ est elle convexe ou concave?

Moindres carrés : $\mathbf{y} = (y_1, \dots, y_n)^\top \in \mathbb{R}^n$ et ici $X \in \mathbb{R}^{n \times p}$ est déterministe. Le modèle est $Y = X\boldsymbol{\theta}^* + \boldsymbol{\varepsilon}$.

- 6) Décrire en fonction de $\ker(X)$, l'ensemble des solutions possible du problème des moindres carrés $\arg \min_{\boldsymbol{\theta} \in \mathbb{R}^p} \|\mathbf{y} - X\boldsymbol{\theta}\|^2$. Même question pour le problème pondéré $\arg \min_{\boldsymbol{\theta} \in \mathbb{R}^p} (\mathbf{y} - X\boldsymbol{\theta})^\top \Omega (\mathbf{y} - X\boldsymbol{\theta})$, où $\Omega = \text{diag}(w_1, \dots, w_n)$ est définie positive.
- 7) Pour le cas pondéré décrit précédemment, donner une formule explicite lorsque X est de rang plein. Donner la matrice de covariance de l'estimateur des moindres carrés (dans l'hypothèse d'un bruit $\boldsymbol{\varepsilon} = \mathbf{y} - X\boldsymbol{\theta}^*$ centré et de matrice de covariance $\sigma^2 \text{Id}_n$).
- 8) Toujours dans le cas pondéré, donner la valeur du risque de prédiction.

On note ici $\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^p} \frac{1}{2n} \|\mathbf{y} - X\boldsymbol{\theta}\|_2^2 + \frac{\lambda}{2} \|\boldsymbol{\theta}\|_2^2$ l'estimateur Ridge.

- 9) Sous quels conditions l'estimateur ridge est-il uniquement défini? Donner une formule explicite pour l'estimateur Ridge en fonction de X , y et λ .
- 10) Donner le biais de l'estimateur Ridge sous l'hypothèse que le bruit $\varepsilon = \mathbf{y} - X\boldsymbol{\theta}^*$ est centré.
- 11) Lorsque X est de rang plein, et que $\text{Var}(\varepsilon) = I_n$, donner la variance de l'estimateur Ridge. La variance de l'estimateur Ridge d'un coefficient donné, est-elle plus grande ou plus petite que celle de l'estimateur des moindres carrés de ce même coefficient?
- 12) Exprimer $\eta_\lambda(z) = \arg \min_{x \in \mathbb{R}} x \mapsto \frac{1}{2}(z - x)^2 + \lambda|x|$ en utilisant le signe de z et la fonction partie positive $(\cdot)_+$.
- 13) Donner en tout point la sous-différentielle de la fonction réelle $x \mapsto (x)_+ = \max(x, 0)$.
- 14) Donner l'étape de mise à jour principale en descente par coordonnée pour résoudre le problème de l'Elastic Net : $\hat{\boldsymbol{\theta}}_\lambda = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^p} \left[\frac{1}{2} \|\mathbf{y} - X\boldsymbol{\theta}\|_2^2 + \lambda \left(\alpha \|\boldsymbol{\theta}\|_1 + (1 - \alpha) \frac{\|\boldsymbol{\theta}\|_2^2}{2} \right) \right]$.
- 15) Donner l'étape de mise à jour principale en descente par coordonnée pour résoudre le problème du Lasso Positif : $\hat{\boldsymbol{\theta}}_\lambda = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}_+^p} \frac{1}{2} \|\mathbf{y} - X\boldsymbol{\theta}\|_2^2 + \lambda \|\boldsymbol{\theta}\|_1$.
- 16) On suppose que l'on dispose d'un solveur Lasso(X, \mathbf{y}, λ) qui résout le problème du Lasso $\hat{\boldsymbol{\theta}}_\lambda = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^p} \frac{1}{2} \|\mathbf{y} - X\boldsymbol{\theta}\|_2^2 + \lambda \|\boldsymbol{\theta}\|_1$. En utilisant ce solveur comment résoudre le problème suivant : $\hat{\boldsymbol{\theta}}_\lambda = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^p} \frac{1}{2} (\mathbf{y} - X\boldsymbol{\theta})^\top \Gamma (\mathbf{y} - X\boldsymbol{\theta}) + \lambda \sum_{j=1}^p w_j |\theta_j|$, pour des $w_j > 0$ et $\Gamma = \text{diag}(\gamma_1, \dots, \gamma_n)$ définie positive?
- 17) Pour des X_1, \dots, X_n identiquement distribuées à valeur dans $\{0, 1\}$, décrire une procédure de test de l'hypothèse $p = P(X_1 = 1) = 1/2$ contre son contraire.
- 18) On considère un modèle de régression où le design est aléatoire et $\mathbb{E}[X_1 X_1^\top]$ est inversible. Décrire la loi asymptotique de $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$, où $\hat{\boldsymbol{\theta}}$ est l'estimateur des moindres carrés. En déduire un intervalle de confiance à 95%. Si maintenant, $\varepsilon = \mathbf{y} - X\boldsymbol{\theta}^*$ est un vecteur Gaussien, donner un autre intervalle de confiance.