31

TELECOM ParisTech — MDI720
2018/2019 — Anne Sabourin, François Portier

EXAM : Modèle linéaire

La durée de l'examen est 3 heures. Les calculatrices, téléphones portables et ordinateurs sont interdits. Pour chaque étudiant, une feuille A4 recto-verso est autorisée. Chaque réponse doit être justifiée. Des points seront attribués pour la présentation.

Questions générales

X 1) Soit X et Y deux vecteurs Gaussiens indépendants de dimension respective p et q. Donner la loi de $AX + BY$ en fonction des moyennes (μ_X et μ_Y) et covariances (Σ_X et Σ_Y) de X et Y, où $A \in \mathbb{R}^{2\times p}$ et $B \in \mathbb{R}^{2\times q}$ sont des matrices déterministes.

X 2) Soit $y_1, y_2, \ldots, y_n$ i.i.d. tel que $\mathbb{E}[y_1^2] < \infty$. Soit $(w_i)_{1\leqslant i\leqslant n}$ un vecteur de poids positifs et déterministes. Quel estimateur $\hat{\mu}$ minimise $\sum_{i=1}^{n} w_i(y_i - \mu)^2$? Donner son biais et sa variance, pour tout $n > 1$.

X 3) Soit $y_1, y_2, \ldots, y_n$ i.i.d. tel que $\mathbb{E}[y_1^2] < \infty$. Soit $(w_i)_{1\leqslant i\leqslant n}$ un vecteur de poids positifs et déterministes. Quel estimateur $\hat{\mu}$ minimise $\sum_{i=1}^{n} w_i(y_i - \mu)^2 + \lambda\mu^2$? Donner son biais et sa variance, pour tout $n > 1$.

4) On observe un échantillon i.i.d. $(X_i)_{1\leqslant i\leqslant n}$. On note $\sigma^2 = \mathrm{Var}(X_1)$. Donner un estimateur sans biais de σ^2. Justifier.

X 5) La fonction $x \mapsto e^{-x^2}$, définie su $\mathbb{R}$, est-elle convexe ? Est-elle concave ?

✓ 6) On note A^+ la pseudo-inverse de A. En utilisant la SVD de X (dont on introduira les notations), montrer que $X(X^TX)^+X^T$ est un projecteur orthogonal. On précisera son espace image.

✓ 7) Donner le projeté orthogonal du vecteur $Y \in \mathbb{R}^n$ sur $\mathrm{Vect}(e_k)$, avec e_k le k-ème vecteur de la base canonique de $\mathbb{R}^n$.

X 8) Montrer que pour toute matrice A, $\ker(A) = \ker(A^TA)$.

Moindres carrés : $Y = (y_1, \ldots, y_n)^T \in \mathbb{R}^n$ **et** $X = (1_n, \tilde{X}) \in \mathbb{R}^{n\times(p+1)}$, $1_n = (1, \ldots, 1)^T \in \mathbb{R}^n$.

9) Soit $\mu_n = \tilde{X}^T 1_n/n$ et $X_c = \tilde{X} - 1_n\mu_n^T$. Montrer que l'assertion "$\exists u \neq 0$ dans $\mathbb{R}^{p+1}$ telle que $Xu = 0$" est équivalente à "$\exists v \neq 0$ dans $\mathbb{R}^p$ telle que $X_c v = 0$". En admettant que pour toute matrice A, $\ker(A) = \ker(A^T A)$, on exprimera la condition d'unicité des moindres-carrées à l'aide de la matrice de covariance empirique.

10) On suppose que X est de rang plein et on note $\hat{\boldsymbol{\theta}}_n$ l'estimateur OLS. On note $\tilde{X} = (\tilde{X}_1, \ldots, \tilde{X}_p)$. On change l'échelle d'une des variables : $\tilde{X}_k$ est remplacé par $\tilde{X}_k b$, où $b > 0$.
   (a) Soit $X_b = (1, X_1, \ldots, X_k b, \ldots, X_p)$. Montrer que $X_b = XD$ où D est une matrice diagonale que l'on précisera.
   (b) Soit $\hat{\boldsymbol{\theta}}_{b,n}$ l'estimateur OLS associé à X_b. Exprimer $\hat{\boldsymbol{\theta}}_{b,n}$ en fonction de $\hat{\boldsymbol{\theta}}_n$ et D.
   (c) Donner la variance de $\hat{\boldsymbol{\theta}}_{b,n}$.
   (d) On a vu que l'estimateur $\hat{\boldsymbol{\theta}}_n$ était affecté par un changement d'échelle. Qu'en est-il de la valeur prédite par le modèle ?

11) Donner la formulation de la pseudo inverse de X connaissant sa SVD (réduite) : $X = \sum_{i=1}^{r} s_i \mathbf{u}_i \mathbf{v}_i^\top$, avec $r = \mathrm{rg}(X)$ et $s_1 \geqslant \cdots \geqslant s_r > 0$. A l'aide de cette dernière et du vecteur Y, exprimer l'estimateur OLS. On notera que X n'est pas nécessairement de rang plein ici.

12) Soit n un entier pair. Donner une formulé explicite du problème $\arg\min_{\boldsymbol{\theta}\in\mathbb{R}^{p+1}} \frac{1}{2}(Y - X\boldsymbol{\theta})^\top \Omega (Y - X\boldsymbol{\theta})$ pour une matrice $\Omega = \mathrm{diag}(w_1, \ldots, w_n)$ telle que $w_i = 1$ si i est pair et 0 si i est impair. Donner une condition équivalente à l'unicité des solutions.

Tests, intervalle de confiance et Bootstrap. On suppose toujours $Y = (y_1, \ldots, y_n)^T \in \mathbb{R}^n$ et $X = (1_n, \tilde{X}) \in \mathbb{R}^{n\times(p+1)}$. On note $\boldsymbol{\theta}^\star = (\theta_0^\star, \ldots, \theta_p^\star)^T \in \mathbb{R}^{p+1}$.

13) Dans le cas du modèle de régression (avec design deterministe) et bruit Gaussien centré de variance σ^2, donner la loi de l'estimateur des moindres carrés $\hat{\boldsymbol{\theta}}_n$ en supposant que X est de plein rang. Soit $x \in \mathbb{R}^{p+1}$ et l'hypothèse nulle $H_0 : x^T\theta^\star = 10$. En déduire une statistique de test et une région critique à 95%.

14) Dans la cadre du modèle linéaire Gaussien avec matrice X de rang plein et bruits de variance σ^2 **connue**, donner la loi de la statistique $T = (\hat{\theta}_k - \theta_k^\star)/\sqrt{\sigma^2 s_k}$, avec $s_k = e_k^T (X^\top X)^{-1} e_k$, $k \in \{0, \ldots p\}$ et e_{k-1} le k-ème vecteur de la base canonique. En déduire un intervalle de confiance à 95% pour le paramètre $\theta_k^\star$.

15) Soient $Z_1, \ldots, Z_n$ des variables aléatoires i.i.d. d'espérance μ et de variance finie. Écrire un pseudo code de bootstrap pour le test sur la moyenne $\mu = 1$. On définira la racine $\hat{R}$ et la racine bootstrap $\hat{R}^*$.

Ridge. A partir de maintenant $X \in \mathbb{R}^{n\times p}$. On note ici $\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \frac{1}{2}\|Y - X\boldsymbol{\theta}\|_2^2 + \frac{\lambda}{2}\|\boldsymbol{\theta}\|_2^2$ l'estimateur Ridge, où $\lambda > 0$.

16) Exprimer $\hat{\boldsymbol{\theta}}_n$.

17) On rappelle la SVD complète $X = USV^T$ où $U \in \mathbb{R}^{n\times n}$ et $V \in \mathbb{R}^{p\times p}$.

(a) Montrer que

$$(X^TX + \lambda I_p)^{-1}X^T = X^T(XX^T + \lambda I_n)^{-1}$$

(hint : on utilisera ici la SVD complète de X)

(b) Si $n \ll p$, donner une formule alternative pour calculer l'estimateur Ridge avec moins d'opération. Comparer le nombre d'opérations nécessaires.

18) Donner une formule explicite de

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} (Y - X\boldsymbol{\theta})^T\Omega(Y - X\boldsymbol{\theta}) + \lambda\|\boldsymbol{\theta}\|_2^2,$$

pour une matrice diagonale $\Omega = \mathrm{diag}(w_1, \ldots, w_n)$ dont les coefficients sont strictement positifs.

19) Donner en tout point la sous-différentielle de la fonction réelle $x \mapsto \max(-x, 0)$.

20) Donner l'étape de mise à jour principale en descente par coordonnée pour résoudre le problème de l'Elastic Net : $\hat{\boldsymbol{\theta}}_\lambda = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \left[\frac{1}{2}\|Y - X\boldsymbol{\theta}\|_2^2 + \lambda\left(\alpha\|\boldsymbol{\theta}\|_1 + (1-\alpha)\frac{\|\boldsymbol{\theta}\|_2^2}{2}\right)\right]$.