TELECOM ParisTech **MDI-720**
2010/2020 **François Portier, Anne Sabourin**

QUIZZ : Modèle linéaire

---

1 Général :

1) Que vaut $Cov(X + \mu)$ pour tout $\mu \in \mathbb{R}^p$ déterministe, et tout vecteur aléatoire $X \in \mathbb{R}^p$?
2) Que vaut $Cov(AX)$, pour toute matrice $A \in \mathbb{R}^{m\times p}$ et tout vecteur aléatoire $X \in \mathbb{R}^p$?
3) Donner un modèle naturel pour "un lancer de dé" (non-nécéssairement équilibré) ?
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 $\sum_{i=1}^n (x_i - \mu)^2$? Donner son biais et sa variance, pour tout $n > 1$.
5) Que vaut le biais de $\dfrac{1}{n}\displaystyle\sum_{i=1}^n (y_i - \overline{y}_n)^2$ ($\overline{y}_n$ est la moyenne empirique) pour des y_i i.i.d, gaussiens, centrés et de variance σ^2 ?
6) On suppose que l'on observe $y_1, \dots, y_n$, des variables réelles i.i.d., gaussiennes, centrées et de variance σ^2. Quel est le risque quadratique de l'estimateur $\dfrac{1}{n}\displaystyle\sum_{i=1}^n (y_i - \overline{y}_n)^2$ de σ^2 ($\overline{y}_n$ est la moyenne empirique) ?
7) Quelle est la projection orthogonal du vecteur $\mathbf{y} \in \mathbb{R}^n$ sur $\mathrm{Vect}(1_n)$, avec $1_n = (1, \dots, 1)^\top \in \mathbb{R}^n$?
8) Quels sont les vecteurs $\mathbf{y} \in \mathbb{R}^n$ tels que $\mathrm{var}_n(\mathbf{y}) = 0$ (var_n est la variance empirique) ?

2 Moindres carrés unidimensionnels :

On observe $\mathbf{y} = (y_1, \dots, y_n)^\top$ et $\mathbf{x} = (x_1, \dots, x_n)^\top$

1) La fonction $(\theta_0, \theta_1) \to \frac{1}{2}\sum_{i=1}^n (y_i - \theta_0 - \theta_1 x_i)^2$ est elle convexe ou concave ?
2) Donner la formule $(\hat{\theta}_0, \hat{\theta}_1)$ des estimateurs des moindres carrés où $\hat{\theta}_0$ correspond au coefficient des constantes et $\hat{\theta}_1$ correspond à l'influence de $\mathbf{x}$ sur $\mathbf{y}$. On les exprimera en fonction des $x_i, y_i, \overline{x}_n, \overline{y}_n$

3 Moindres carrés :

On note ici $\mathbf{y} = (y_1, \dots, y_n)^\top \in \mathbb{R}^n$ et $X \in \mathbb{R}^{n\times p}$ (sauf mention contraire).

1) Écrire un pseudo-code de descente de gradient pour résoudre le problème des moindres carrés.
2) Pour une matrice $X \in \mathbb{R}^{n\times p}$, que vaut $\mathrm{Ker}(X^\top X)$?
3) Pour une matrice $X \in \mathbb{R}^{n\times(p+1)}$, $n > 1$ et $p \geqslant 1$, qui possède comme première colonne une colonne de 1, notons $(1, \tilde{X}_i^\top) \in \mathbb{R}^{p+1}$, les lignes de X. Montrer que $X^\top X$ non-inversible est équivalent à
$$\sum_{i=1}^n (\tilde{X}_i - \hat{\mu}_n)(\tilde{X}_i - \hat{\mu}_n)^T \quad \text{non-inversible},$$
où $\hat{\mu}_n = \sum_{i=1}^n \tilde{X}_i$.

4) Si la matrice $X \in \mathbb{R}^{n\times p}$ est de plein rang, donner une formule exacte de l'estimateur des moindres carrés.
5) Si la matrice $X \in \mathbb{R}^{n\times p}$ n'est pas de plein rang, donner une formule pour un estimateur des moindres carrées.
6) Décrire les 2 cas possibles quant à la définition de l'estimateur des moindre carrées (existence et unicité).
7) Quand l'estimateur des moindre carrées est non-unique, quel est l'ensemble des solutions du problème d'optimisation associé ?
8) Si la matrice $X \in \mathbb{R}^{n\times p}$ est de plein rang, donner la matrice de covariance de l'estimateur des moindres carrés (dans l'hypothèse d'un bruit $\varepsilon = \mathbf{y} - X\theta^*$ centré et de matrice de covariance $\sigma^2\mathrm{Id}_n$).
9) Donner un estimateur sans biais du niveau du bruit σ^2 (dans le cas où le X est déterministe).
10) On suppose que X est de rang plein et on note $\hat{\boldsymbol{\theta}}_n$ l'estimateur OLS. On note $\tilde{X} = (\tilde{X}_1, \dots, \tilde{X}_p)$. On change l'échelle d'une des variables : $\tilde{X}_k$ est remplacé par $\tilde{X}_k b$, où $b > 0$.
    (a) Soit $X_b = (1, X_1, \dots, X_k b, \dots, X_p)$. 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 une formule explicite du problème $\arg\min_{\boldsymbol{\theta}} \frac{1}{2}(\mathbf{y}-X\boldsymbol{\theta})^\top \Omega(\mathbf{y}-X\boldsymbol{\theta})$ pour une matrice $\Omega = \mathrm{diag}(w_1, \dots, w_n)$ définie positive, dans le cas où X est de plein rang.
12) Dans le cas du modèle de régression avec design aléatoire, décrire l'asymptotique de l'estimateur des moindre carrées. On donnera la loi asymptotique de $\sqrt{n}(\hat{\beta} - \beta^*)$.
13) Dans le cas du modèle de régression avec design déterministe et bruit gaussien centré de variance σ^2, donner la loi de l'estimateur des moindre carrées $\hat{\beta}$.
14) Dans le cas du modèle de régression avec design déterministe où X est de plein rang p, donner la valeur du risque de prédiction.

4 Ridge :

On note $\hat{\theta} = \arg\min_\theta \frac{1}{2}\|\mathbf{y} - X\theta\|_2^2 + \frac{\lambda}{2}\|\theta\|_2^2$ l'estimateur Ridge

1) Donner une formule explicite pour l'estimateur Ridge en fonction de y et λ quand $X = \mathrm{Id}_n$.
2) Donner une formule explicite pour l'estimateur Ridge en fonction de X, y et λ.
3) Donner la variance de l'estimateur Ridge sous l'hypothèse que le bruit $\mathbf{y} - X\boldsymbol{\theta}^*$ est centré et de variance $\sigma^2\mathrm{Id}_n$.
4) Donner en fonction de X, y, $D \in \mathbb{R}^{p\times p}$ et λ, une formule explicite de

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \frac{1}{2}\|\mathbf{y} - X\boldsymbol{\theta}\|_2^2 + \frac{\lambda}{2}\|D\boldsymbol{\theta}\|_2^2,$$

5 Lasso :

1) Exprimer $\eta_\lambda(z) = \arg\min_{x\in\mathbb{R}} x \mapsto \frac{1}{2}(z-x)^2 + \lambda|x|$ en fonction du signe de x et de la partie positive $(\cdot)_+$.

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

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

4) Donner l'étape de mise à jour principale en descente par coordonnée pour résoudre le problème du *Lasso Positif* : $\hat{\boldsymbol{\theta}}_\lambda = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}_+^p} \frac{1}{2}\|\mathbf{y} - X\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_1$.

5) Donner l'étape de mise à jour principale en descente par coordonnée pour résoudre le problème :

$$\hat{\boldsymbol{\theta}}_\lambda = \underset{\boldsymbol{\theta}\in\mathbb{R}^p, s.c. \Omega\boldsymbol{\theta}\geqslant 0}{\arg\min} \frac{1}{2}\|\mathbf{y} - X\boldsymbol{\theta}\|_2^2,$$

avec $\Omega = \mathrm{diag}(w_1, \dots, w_n)$.

6) On suppose que l'on dispose d'un solveur Lasso(X, 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 : $\hat{\boldsymbol{\theta}}_\lambda = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \frac{1}{2}\|\mathbf{y} - X\boldsymbol{\theta}\|_2^2 + \lambda\sum_{j=1}^p w_j|\theta_j|$, pour des $w_j \geqslant 0$?

6 ACP/SVD :

1) Donner la formulation de la pseudo inverse de X connaissant sa SVD : $X = \sum_{i=1}^r s_i \mathbf{u}_i \mathbf{v}_i^\top$, avec $r = \mathrm{rg}(X)$ et $s_1 \geqslant \cdots \geqslant s_r > 0$.

2) Que vaut $\begin{cases} \max\limits_{u\in\mathbb{R}^n, v\in\mathbb{R}^p} u^\top X v \\ \text{s.c. } \|u\|_2^2 = 1 \text{ et } \|v\|_2^2 = 1 \end{cases}$?

7 Test :

1) 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.

2) Soient $X_1, \dots, X_n$ des variables aléatoires i.i.d selon des lois gaussiennes de moyenne (inconnue) μ et de variance connue σ^2, *i.e.*,$X_i \sim \mathcal{N}(\mu, \sigma^2)$. Décrire une procédure de test de l'hypothèse $\mu = 1$ contre son contraire.

3) Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes et distribuées selon des lois gaussiennes de moyenne (inconnue) μ et de variances connues σ_i^2, *i.e.*,$X_i \sim \mathcal{N}(\mu, \sigma_i^2)$. Décrire une procédure de test de l'hypothèse $\mu = 1$ contre son contraire.

8 Bootstrap :

1) Soient $\epsilon_1, \dots, \epsilon_n$ des variables aléatoires i.i.d. de loi gaussienne de moyenne (inconnue) μ et de variance (connue) σ^2, *i.e.*,$\epsilon_i \overset{i.i.d}{\sim} \mathcal{N}(\mu, \sigma^2)$. On observe $(y_i, x_i)_{i=1,\dots,n}$ tel que, pour chaque $i = 1, \dots, n$, $y_i = x_i^T\beta + \epsilon_i$ et x_i déterministe.
   (a) La variable y_1 est-elle indépendante de y_2 ? La variable y_1 a-t-elle même loi que y_2 ? Donner la distribution de y_i.
   (b) Soit $\hat{\beta}$ l'estimateur OLS du modèle $y_i = x_i^T\beta + \epsilon_i$. A l'aide de résidus bootstrap $(\epsilon_i^*)_{i=1,\dots,n} \overset{i.i.d}{\sim} \mathcal{N}(\mu, \sigma^2)$, calculer $(y_i^*)_{i=1,\dots,n}$ et donner leur loi. Calculer l'estimateur bootstrap $\hat{\beta}^*$ et donner la loi de $\sqrt{n}(\hat{\beta}^* - \hat{\beta})$.

2) Proposer une procédure bootstrap sur les résidus pour estimer l'écart quadratique moyen de la méthode des moindres carrées dans le cas d'une régression linéaire.