
LISTE DE QUESTIONS (À ÉVOLUER) : Modèles linéaires

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) Quel est un modèle naturel pour "un lancer de dé" (non-nécessairement 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 $\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_n)^2$ (\bar{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 $\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_n)^2$ de σ^2 (\bar{y}_n est la moyenne empirique) ?
- 7) Quelle est la projection orthogonal du vecteur $\mathbf{y} \in \mathbb{R}^n$ sur $\text{Vect}(1_n)$, avec $1_n = (1, \dots, 1)^\top \in \mathbb{R}^n$?
- 8) Quels sont les vecteurs $\mathbf{y} \in \mathbb{R}^n$ tels que $\text{var}_n(\mathbf{y}) = 0$ (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 $(\theta_0, \theta_1) \rightarrow \frac{1}{2} \sum_{i=1}^n (y_i - \theta_0 - \theta_1 x_i)^2$ 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 $x_i, y_i, \bar{x}_n, \bar{y}_n$

3 Moindres carrés : $\mathbf{y} = (y_1, \dots, y_n)^\top \in \mathbb{R}^n$ et $X \in \mathbb{R}^{n \times p}$

- 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 $\text{Ker}(X^\top X)$?
- 3) Pour une matrice $X \in \mathbb{R}^{n \times (p+1)}$, $n > 1$ et $p \geq 1$, 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

$$\sum_{i=1}^n (\tilde{X}_i - \hat{\mu}_n)(\tilde{X}_i - \hat{\mu}_n)^\top \text{ non-inversible,}$$

$$\text{où } \hat{\mu}_n = \sum_{i=1}^n \tilde{X}_i.$$

- 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 carrés.
- 6) Décrire les 2 cas possible quant à la définition de l'estimateur des moindres carrés (existence et unicité).
- 7) 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 $\sigma^2 \text{Id}_n$).
- 8) 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 $r = \text{rg}(X)$ et $s_1 \geq \dots \geq s_r > 0$.
- 9) Donner un estimateur sans biais du niveau du bruit σ^2 (dans le cas où le X est déterministe).
- 10) Donner une formule explicite du problème $\arg \min_{\theta} \frac{1}{2} (\mathbf{y} - X\theta)^\top \Omega (\mathbf{y} - X\theta)$ pour une matrice $\Omega = \text{diag}(w_1, \dots, w_n)$ définie positive, dans le cas où X est de plein rang.
- 11) Dans le cas du modèle de régression avec design aléatoire, décrire l'asymptotique de l'estimateur des moindres carrés. On donnera la loi asymptotique de $\sqrt{n}(\hat{\beta} - \beta^*)$.
- 12) 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 moindres carrés $\hat{\beta}$.
- 13) 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 $X = \text{Id}_n$.
- 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 $\mathbf{y} - X\theta^*$ est centré et de variance $\sigma^2 \text{Id}_n$.
- 4) Donner en fonction de X , y , $D \in \mathbb{R}^{p \times p}$ et λ , une formule explicite de

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

5 Lasso :

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

- 5) On suppose que l'on dispose d'un solveur $\text{Lasso}(X, y, \lambda)$ qui résout le problème du Lasso $\hat{\theta}_\lambda = \arg \min_{\theta \in \mathbb{R}^p} \frac{1}{2} \|y - X\theta\|_2^2 + \lambda \|\theta\|_1$. En utilisant ce solveur comment résoudre le problème suivant : $\hat{\theta}_\lambda = \arg \min_{\theta \in \mathbb{R}^p} \frac{1}{2} \|y - X\theta\|_2^2 + \lambda \sum_{j=1}^p w_j |\theta_j|$, pour des $w_j \geq 0$?

6 ACP/SVD :

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

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 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.*, $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Écrire un pseudo code de bootstrap pour le test sur la moyenne $\mu = 1$.
- 2) Proposer une procédure bootstrap sur les résidus pour estimer l'écart quadratique moyen de la méthode des moindres carrés dans le cas d'une régression linéaire.