Quiz: pool de questions

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”?
4. 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?
5. On suppose que l’on observe $y_1, \ldots, y_n$, 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)?
6. Quelle est la projection du vecteur $\mathbf{y} \in \mathbb{R}^n$ sur $\mathbf{Vect}(\mathbf{1}_n)$, avec $\mathbf{1}_n = (1, \ldots, 1)^\top \in \mathbb{R}^n$?
7. Quels sont les vecteurs $\mathbf{y} \in \mathbb{R}^n$ tels que $\mathbf{var}_n(\mathbf{y}) = 0$ ($\mathbf{var}_n$ est la variance empirique)?

Moindres carrés unidimensionnel: on observe $\mathbf{y} = (y_1, \ldots, y_n)^\top$ et $\mathbf{x} = (x_1, \ldots, x_n)^\top$

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 θ_0 correspond au coefficient des constantes et $\hat{\theta}_1$ correspond à l'influence de $\mathbf{x}$ sur $\mathbf{y}$. On les exprimeras en fonction des $x_i, y_i, \overline{x}_n, \overline{y}_n$

Moindre carrés: $\mathbf{y} = (y_1, \ldots, y_n)^\top$ 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 $\mathbf{Ker}(X^\top X)$?

3. Si la matrice $X \in \mathbb{R}^{n \times p}$ est de plein rang, donner une formule exacte de l'estimateur des moindres carrés.

4. 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 $\epsilon = \mathbf{y} - X\theta^*$ centré et de matrice de covariance $\sigma^2 \mathbf{Id}_n$).

5. Donner la formulation de la pseudo inverse si la SVD de X peut s'écrire $X = \sum_{i=1}^r s_i \mathbf{v}_i \mathbf{u}_i^\top$.

6. 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 $\Omega = \mathbf{diag}(w_1, \ldots, w_n)$ définie positive.

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 $\mathbf{y} - X\theta^*$ est centré et de variance $\sigma^2 \mathbf{Id}_n$.

4. Donner une formule explicite pour l'estimateur Ridge généralisé:
$\hat{\theta} = \arg\min_{\theta} \frac{1}{2}\|\mathbf{y} - X\theta\|_2^2 + \frac{\lambda}{2}\|D\theta\|_2^2,$
en fonction de $X, y, D \in \mathbb{R}^{p\times p}$ et λ.

Lasso:

1. Calculer $\eta_\lambda(Z) = \arg\min_{x\in\mathbb{R}} x \mapsto \frac{1}{2}(z-x)^2 + \lambda|x|$ 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*:
$\hat{\theta}_\lambda = \arg\min_{\theta\in\mathbb{R}^p} \left[\frac{1}{2}\|\mathbf{y} - X\theta\|_2^2 + \lambda\left(\alpha\|\theta\|_1 + (1-\alpha)\frac{\|\theta\|_2^2}{2}\right)\right].$
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 $\mathbf{Lasso}(X, y, \lambda)$ qui résout le problème du Lasso
$\hat{\theta}_\lambda = \arg\min_{\theta\in\mathbb{R}^p} \frac{1}{2}\|\mathbf{y} - X\theta\|_2^2 + \lambda\|\theta\|_1$. En utilisant ce solveur comment résoudre le problème suivant:
$\hat{\theta}_\lambda = \arg\min_{\theta\in\mathbb{R}^p} \frac{1}{2}\|\mathbf{y} - X\theta\|_2^2 + \lambda\sum_{j=1}^{p} w_j|\theta_j|$, pour des $w_j \geq 0$?

ACP/SVD

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

Test:

1. Pour des $X_1, \ldots, X_n$ 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 $X_1, \ldots, X_n$ des variables aléatoires i.i.d selon des lois gaussiennes de moyenne μ et de variance connue σ, *i.e.,* $X_i \sim \mathcal{N}(0, \sigma^2)$. Décrire une procédure de test de l'hypothèse $\mu = 1$ contre son contraire.

3. Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes et distribuées selon des lois gaussiennes de moyenne μ et de variances connues σ_i^2, *i.e.,* $X_i \sim \mathcal{N}(0, \sigma_i^2)$. Décrire une procédure de test de l'hypothèse $\mu = 1$ contre son contraire.

Bootstrap

1. Soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d selon des lois gaussiennes de moyenne μ et de variance connue σ, *i.e.,* $X_i \sim \mathcal{N}(0, \sigma^2)$. Écrire un pseudo code de bootstrap pour le test sur la moyenne $\mu = 1$.

2. Soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d et $w_1, \ldots, w_n$ une suite de variables i.i.d. de moyenne 1 et de variance 1. A l'aide des X_i et des w_i construire un intervalle de confiance à 99% pour la quantité $\mathbb{P}(X_1 \geq 10)$.

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