Feuille de TD 4 : Modèle linéaire

Exercice 1 (Théorème de Cochran)

On se place dans l'espace vectoriel euclidien $\mathbb{R}^n$ muni du produit scalaire usuel, défini par $\langle x, y\rangle = x_1y_1+\cdots+x_ny_n$. Soit F un sous-espace vectoriel de dimension p de $\mathbb{R}^n$, et $F^\perp$ l'orthogonal de F. On note u_F (resp. $u_{F^\perp}$) l'opérateur de projection orthogonale de $\mathbb{R}^n$ dans F (resp. $F^\perp$) et A la matrice de u_F dans la base canonique.

1. Montrer que $A^2 = A$. À quelle application linéaire la matrice $I - A$ correspond-elle ? Montrer que $(I - A)^\top A = 0$ puis que $A^\top = A$.
   Indication : $\forall u \in \mathbb{R}^n, [u = 0] \Leftrightarrow [\forall y \in \mathbb{R}^n, y^\top u = 0]$.
2. Montrer que u_F et $u_{F^\perp}$ sont simultanément diagonalisables dans une base orthonormée de $\mathbb{R}^n$. Déterminer les matrices de u_F et $u_{F^\perp}$ dans cette base. Montrer que la matrice de passage P vérifie $P^\top = P^{-1}$.
3. Soit $X = (X_1 \dots X_n)^\top$ un vecteur gaussien de loi $\mathcal{N}(0, I)$.
   (a) Calculer $\mathrm{Cov}((I - A)X, AX)$. Que peut-on dire des variables $(I - A)X$ et AX ?
   (b) Quelles sont les lois des vecteurs $P^\top AX$ et $P^\top(I - A)X$?
   (c) Montrer que l'on a $\|AX\|^2 = \|P^\top AX\|^2$. En déduire la loi de $\|AX\|^2$. Faire de même pour $\|(I - A)X\|^2$.

   Les résultats des questions (a) et (c) constituent le théorème de Cochran : si on projette un vecteur aléatoire gaussien standard sur deux sous-espaces vectoriels orthogonaux, alors les projections sont indépendantes. De plus le carré de la norme de chacune des projections suit une loi du χ^2 dont le nombre de degrés de liberté est égal à la dimension du sous-espace correspondant.
4. Application. Soit $X = (X_1, \dots, X_n)^\top \sim \mathcal{N}(0, I)$. On pose $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$ et $s_n^2 = n^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$. Montrer que les variables aléatoires $\overline{X}_n$ et s_n^2 sont indépendantes et déterminer leurs lois.

Exercice 2 (Conditionnement gaussien)

Soit $(X_1, \dots, X_n)^\top$ un vecteur gaussien. On se donne une matrice A de taille $m \times n$. Donner l'espérance et loi conditionnelle de $(X_1, \dots, X_n)$ sachant $A(X_1, \dots, X_n)^\top$.

Exercice 3 (Modèle linéaire)

Le principe de la régression linéaire est de modéliser une variable y à partir de variables explicatives

$$x = (x_1, \dots, x_p)^\top,$$

c'est-à-dire de considérer

$$y \simeq \beta_1 x_1 + \cdots + \beta_p x_p,$$

où $\beta = (\beta_1, \dots, \beta_p)^\top$ est inconnu.

En pratique, on dispose d'un échantillon $(x_1, y_1), \dots, (x_n, y_n)$, mais on n'obtient jamais réellement une droite (erreurs de mesures...). On va donc considérer le modèle linéaire

$$y = \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2).$$

On parle alors de *modèle linéaire gaussien.*

On suppose maintenant que les données suivent le modèle :

$$Y_i = \beta_1 x_{i,1} + \cdots + \beta_p x_{i,p} + \varepsilon_i,$$

avec :

— Y_i est une variable aléatoire et on observe les réalisations y_i,
— les $x_i = (x_{i,1}, \ldots, x_{i,p})^\top$ sont déterministes,
— le paramètre $\beta = (\beta_1, \ldots, \beta_p)^\top$ est inconnu et déterministe,
— les ε_i sont i.i.d. et $\varepsilon_1 \sim \mathcal{N}(0, \sigma^2)$.

1. Écrire le modèle de manière matricielle.
2. Donner la loi du vecteur Y. Quelle est la loi de Y_i ?
3. On considère à partir de maintenant que $\text{rang}(X) = p$, et on note $D = \text{Im}(X)$. On s'intéresse à l'estimateur des moindres carrés défini par

   $$\hat{\beta} = \arg\min_{h \in \mathbb{R}^p} \|Y - Xh\|^2.$$

   Montrer que la matrice $X^\top X$ est inversible et en déduire $\hat{\beta}$.
4. Montrer que

   $$P_D = X(X^\top X)^{-1} X^\top$$

   est le projecteur orthogonal sur D parallèlement à $D^\perp$.
5. Que pouvez-vous en déduire sur $X\hat{\beta}$?
6. Donner la loi de $X\hat{\beta}$ et en déduire celle de $\hat{\beta}$.
7. On suppose σ^2 connu. Soit $x_0 \in \mathbb{R}^p \setminus \{0\}$, donner un intervalle de confiance de niveau au moins $1 - \alpha$ de $x_0^\top \beta$.
8. On suppose maintenant que σ^2 est inconnu et on considère l'estimateur

   $$\hat{\sigma}^2 = \frac{1}{n-p} \|Y - X\hat{\beta}\|^2.$$

   (a) Expliquer ce choix d'estimateur.
   (b) Exprimer $\hat{\sigma}^2$ à l'aide de projections.
   (c) Énoncer le théorème de Cochran dans ce cas.
   (d) En déduire un intervalle de confiance pour $x_0^\top \beta$.
   (e) En déduire un test de niveau α pour tester $H_0 : \beta_j = \beta_{j,0}$.
   (f) Construire un intervalle de confiance pour σ.
   (g) Construire un test de niveau α pour tester $H_0 : \beta = \beta_0$.
9. On considère maintenant une $(n+1)$-ème donnée x_{n+1} et on souhaite prédire Y_{n+1}.
   (a) Proposer un prédicteur $\hat{Y}_{n+1}$.
   (b) Donner la loi de l'erreur de prédiction $\hat{\varepsilon}_{n+1} = Y_{n+1} - \hat{Y}_{n+1}$.
   (c) En déduire un intervalle de prédiction.

Exercice 4 (Moindres carrés –encore)

On considère le modèle suivant

$$Y = X\beta + \varepsilon,$$

où X est une matrice de taille $n \times p$, $\beta \in \mathbb{R}^p$, ε est un vecteur alétoire centré de $\mathbb{R}^n$ et Y est un vecteur aléatoire de $\mathbb{R}^n$. On note $\hat{\beta}$ l'estimateur des moindres carrés donné par

$$\hat{\beta} \in \text{argmin}_\alpha \|Y - X\alpha\|_2^2. \tag{1}$$

On pose aussi $\hat{Y} = X\hat{\beta}$ et $\hat{\varepsilon} = Y - \hat{Y}$.

Expression de l'EMC.

1. Montrer que le sous-espace $\mathrm{Im}(X) = \{X\alpha, \alpha \in \mathbb{R}^p\}$ est de dimension p et que $X^\top X$ est inversible.
2. En calculant un gradient, montrer que

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y \,.$$

Point de vue géométrique. On veut retrouver la formule précédente en tuilisant des arguments géométriques.

3. Notons P_X la matrice de projection sur $\mathrm{Im}(X)$. En utilisant (1), trouver une relation entre P_X , X, Y et $\hat{\beta}$.
4. En utilisant le fait que Y peut se décomposer en un élément de $\mathrm{Im}(X)$ plus un élément de $\mathrm{Im}(X)^\perp$, retrouver l'expression de $\hat{\beta}$ et $P_{\mathrm{Im}(X)}$. Montrer que $\hat{\varepsilon} = P_{\mathrm{Im}(X)^\perp} Y$.
5. Montrer que le biais de $\hat{\beta}$ donné par $\mathbb{E}[\hat{\beta}] - \beta$ est nul.

Erreurs gaussiennes. On suppose maintenant que ε est un vecteur gaussien centré, avec une matrice de variance-covariance $\sigma^2 I_n$.

6. Donner la distribution de Y.
7. En utilisant le théorème de Cochran, trouver la distribution de $\hat{Y}$ et $\hat{\varepsilon}$ et montrer qu'elles sont indépendantes.
8. Finalement, expliciter la distribution du vecteur aléatoire

$$\begin{pmatrix} X\hat{\beta} \\ \hat{\beta} \\ \hat{\varepsilon} \end{pmatrix}.$$

Exercice 5 (Régression ridge)

On considère le modèle de régression linéaire

$$\underset{(n\times 1)}{Y} = \underset{(n\times k)}{X}\underset{(k\times 1)}{\theta} + \underset{(n\times 1)}{\xi} \,.$$

où X est une matrice déterministe, $\mathbb{E}[\xi] = 0$, $\mathbb{E}[\xi\xi^\top] = \sigma^2 I_n$,

1. Supposons que $k > n$. Que peut-on dire de l'estimateur des moindres carrés ?
2. On appelle *régression ridge* avec hyperparamètre de régularisation $\lambda > 0$ l'estimateur

$$\hat{\theta}_\lambda = \arg\min_{\theta \in \mathbb{R}!^k} \left\{ \|Y - X\theta\|^2 + \lambda \|\theta\|^2 \right\}.$$

   Donner une formule pour $\hat{\theta}_\lambda$ en fonction de X, Y et λ. Cet estimateur est-il bien défini pour $k > n$?
3. Calculer son espérance et sa matrice de variance-covariance. Est-il biaisé ?
4. Supposons que $k = 1$ (régression simple unidimensionnelle), montrer qu'il existe une valeur de λ pour laquelle le risque quadratique de l'estimateur ridge d'hyperparamètre λ est plus faible que le risque de l'estimateur des moindres carrés.