

TD N° 1 : Introduction et rappels

EXERCICE 1.

- 1) Montrer que pour toutes matrices $A \in \mathbb{R}^{n_1 \times n_1}$, $B \in \mathbb{R}^{n_1 \times n_2}$, $C \in \mathbb{R}^{n_2 \times n_1}$, $D \in \mathbb{R}^{n_2 \times n_2}$ telle que les matrices $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$, A et D sont inversibles on a la relation suivante :

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix} .$$

Aide : on pourra utiliser un pivot de Gauss pour prouver le résultat précédent.

Rem: On appelle complément de Schur du bloc D de la matrice M , la matrice de dimension $n_1 \times n_2$ suivante : $A - BD^{-1}C$, qu'on note parfois M/D .

- 2) Procéder de même pour montrer que :

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} + D^{-1} \end{bmatrix}$$

- 3) Montrer que les déterminants suivants sont égaux.

$$\left| \begin{bmatrix} A & B \\ C & D \end{bmatrix} \right| = |A||D - CA^{-1}B| = |D||A - BD^{-1}C| .$$

De plus si $AC = CA$ montrer qu'alors $\left| \begin{bmatrix} A & B \\ C & D \end{bmatrix} \right| = |AD - CB|$.

- 4) Montrer l'identité de Woodbury sous les mêmes hypothèses :


$$(A + BDC)^{-1} = A^{-1} - A^{-1}B(D^{-1} + CA^{-1}B)^{-1}CA^{-1} . \quad (1)$$

- 5) Appliquer la formule précédente pour prouver la relation suivante pour des scalaires $x \neq 0$ et $\delta \neq 0$:

$$\frac{1}{x + \delta} = \frac{1}{x} - \left(\frac{1}{x}\right)^2 \cdot \left(\frac{1}{\delta} + \frac{1}{x}\right)^{-1} . \quad (2)$$

- 6) Montrer la relation suivante pour des vecteurs $\mathbf{u} \in \mathbb{R}^{n_1}$ et $\mathbf{v} \in \mathbb{R}^{n_1}$:

$$(A + \mathbf{u}\mathbf{v}^\top)^{-1} = A^{-1} - \frac{A^{-1}\mathbf{u}\mathbf{v}^\top A^{-1}}{1 + \mathbf{v}^\top A^{-1}\mathbf{u}} . \quad (3)$$

EXERCICE 2. (Moindres carrés en ligne ( : online)) Le but de cet exercice est de montrer que l'estimateur des moindres carrés $\hat{\beta}_n$ associé avec la matrice de co-variables $X_{(n)} \in \mathbb{R}^{n \times p}$ et comme vecteur observé $\mathbf{y}_{(n)} \in \mathbb{R}^n$ peut facilement être mis à jour quand un couple $(\mathbf{x}_{n+1}, y_{n+1}) \in \mathbb{R}^p \times \mathbb{R}$ est obtenu. On appliquera ce résultat à la validation croisée pour finir.

Pour clarifier les notations dans cette exercice :

$$X_{(n+1)} = \begin{pmatrix} X_{(n)} \\ \mathbf{x}_{n+1}^\top \end{pmatrix} \in \mathbb{R}^{(n+1) \times p}, \quad \text{et} \quad \mathbf{y}_{(n+1)} = \begin{pmatrix} \mathbf{y}_{(n)} \\ y_{n+1} \end{pmatrix} \in \mathbb{R}^{n+1} .$$

On suppose dorénavant que $X_{(n)}$ et $X_{(n+1)}$ sont de plein rang colonne (*i.e.*, pour chaque matrice, leur colonnes sont linéairement indépendantes).

1) En définissant $\mathbf{z}_{n+1} = (X_{(n)}^\top X_{(n)})^{-1} \mathbf{x}_{n+1}$ et $a_{n+1} = \mathbf{x}_{n+1}^\top (X_{(n)}^\top X_{(n)})^{-1} \mathbf{x}_{n+1}$ montrer que

$$(X_{(n+1)}^\top X_{(n+1)})^{-1} = (X_{(n)}^\top X_{(n)})^{-1} - \frac{\mathbf{z}_{n+1} \mathbf{z}_{n+1}^\top}{1 + a_{n+1}} \quad (4)$$

2) Exprimer $X_{(n+1)}^\top \mathbf{y}_{(n+1)}$ en fonction de $X_{(n)}^\top \mathbf{y}_{(n)}$ et $y_{n+1} \mathbf{x}_{n+1}$.

3) Montrer que l'estimateur des moindres carrés $\hat{\boldsymbol{\beta}}_{n+1}$ avec la matrice $X_{(n+1)}$ et le signal $\mathbf{y}_{(n+1)}$ s'obtient de la façon suivante :

$$\hat{\boldsymbol{\beta}}_{n+1} = \hat{\boldsymbol{\beta}}_n + \frac{r_{n+1}}{1 + a_{n+1}} \mathbf{z}_{n+1} \quad , \quad (5)$$

avec $r_{n+1} = y_{n+1} - \mathbf{x}_{n+1}^\top \hat{\boldsymbol{\beta}}_n$ (résidu pour l'observation $n + 1$).


4) Supposons que la matrice $(X_{(n)}^\top X_{(n)})^{-1}$ et le vecteur $\hat{\boldsymbol{\beta}}_n$ ont été stockés, comment mettre à jour $\hat{\boldsymbol{\beta}}_{n+1}$ en utilisant le moins possible d'opérations de type : multiplication matrice-vecteur (pour des matrices de taille $p \times p$ et des vecteurs de taille p). Combien d'opérations sont nécessaires ?

5) En utilisant (4), montrer que $1 + a_{n+1} = \frac{1}{1 - h_{n+1}}$ où $h_{n+1} = \mathbf{x}_{n+1}^\top (X_{(n+1)}^\top X_{(n+1)})^{-1} \mathbf{x}_{n+1}$.

6) La prédiction de y_{n+1} donnée par le modèle est $\hat{y}_{n+1} := \mathbf{x}_{n+1}^\top \hat{\boldsymbol{\beta}}_{n+1}$. Avec la formule suivante

$$\hat{y}_{n+1} = \mathbf{x}_{n+1}^\top \hat{\boldsymbol{\beta}}_n + \frac{r_{n+1} a_{n+1}}{1 + a_{n+1}} \quad . \quad (6)$$

montrer que $y_{n+1} - \hat{y}_{n+1} = r_{n+1}(1 - h_{n+1})$.

7) Partant de données (\mathbf{y}, X) , la validation croisée "laissant un de côté" ( : *leave-one-out*) consiste à calculer le risque (ici quadratique) suivant

$$R_{cv} = \frac{1}{n} \sum_{i=1}^n (y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{(-i)})^2 \quad ,$$

où $\hat{\boldsymbol{\beta}}_{(-i)}$ est l'estimateur des moindres carrés basé sur $(\mathbf{y}_{(-i)}, X_{(-i)})$, *i.e.*, sur les données (\mathbf{y}, X) sans la i^e observation. En appliquant ce qui a été fait précédemment, montrer que

$$R_{cv} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 / (1 - \hat{h}_i)^2,$$

avec $\hat{h}_i = \mathbf{x}_i^\top (X^\top X)^{-1} \mathbf{x}_i$ et $\hat{y}_i = \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_n$, $\hat{\boldsymbol{\beta}}_n$ étant l'estimateur des moindres carrés de (\mathbf{y}, X) .

EXERCICE 3. (Modèle gaussien et loi(s) conditionnelle(s)) Soit $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \in \mathbb{R}^n$ (avec $n = n_1 + n_2$)

un vecteur gaussien suivant la loi $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, avec

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \text{et} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \in \mathcal{S}_n^+ \quad ,$$

où \mathcal{S}_n^+ est l'ensemble des matrices symétriques semi-définies positives¹ On notera donc que $\Sigma_{21} = \Sigma_{12}^\top$, et $\Sigma_{22}^\top = \Sigma_{22}$.

1) Montrer que \mathbf{x}_1 (resp. \mathbf{x}_2) est gaussien, d'espérance μ_1 (resp. μ_2) et de covariance Σ_{11} (resp. Σ_{22}).

2) Montrer que la distribution conditionnelle de \mathbf{x}_i sachant \mathbf{x}_j est normale, d'espérance $\mu_{i|j} := \mu_i + \Sigma_{ij} \Sigma_{jj}^{-1} (\mathbf{x}_j - \mu_j)$ et de matrice de covariance $\Sigma_{i|j} := \Sigma_{ii} - \Sigma_{ij}^\top \Sigma_{jj}^{-1} \Sigma_{ij}$. Aide : on pourra commencer par donner la densité jointe du vecteur \mathbf{x} .

3) Appliquer le résultat précédent quand $n_1 = n_2 = 1$. En notant ρ la corrélation entre les variables \mathbf{x}_1 et \mathbf{x}_2 , on pourra écrire la matrice de covariance sous la forme :

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix} \quad .$$

1. $\Sigma \in \mathcal{S}_n^+ \iff \forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x}^\top \Sigma \mathbf{x} \geq 0$