Gestion des incertitudes et analyse de risque
TD1 : Régression linéaire et inférence bayésienne

Lucie Le Briquer

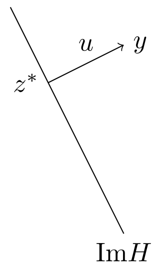
12 janvier 2018

Exercice 1 (régression linéaire)

1. ok
2. On a $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathrm{id})$, donc $y \sim \mathcal{N}(H\beta, \sigma^2 \mathrm{id})$.
3. $g \colon t \mapsto \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp(t)$ croissante en $t \Rightarrow$ minimum de $g \circ (-f)$ atteint en $\hat{\beta} = \arg\min f$.
   Géométriquement, correspond à une projection de y sur $\Im(H)$. On cherche donc $\min_{z \in \Im H} \|y - z\|^2$. Notons z^* l'arg min. Alors comme $z \in \Im H$, $\exists \hat{\beta}$ tel que $H\hat{\beta} = z^*$.

   Sinon de façon analytique, on peut considérer $J(\beta) = \|y - H\beta\|^2$. On a :
   $$\langle \nabla J(\beta), \delta\beta \rangle = 2\langle y - H\beta, -H\delta\beta \rangle$$
   $$\langle J(\beta_0) - J(\beta_1), \beta_0 - \beta_1 \rangle = 2\|H(\beta_0 - \beta_1)\| \geqslant 0$$
   $\forall \delta\beta,\ \langle \nabla J(\hat{\beta}), \delta\beta \rangle = 0.$
   $$H^T H \hat{\beta} = H^T y$$
   Indication. Vérifier que $\Im(H^T) = \Im(H^T H)$.
   Preuve : $(\Im M)^\perp = \mathrm{Ker}(M^T)$ et $\ker(M^T M) = \ker(M)$

   Soit $\hat{\beta}_0$ tel que $\hat{H} H \hat{\beta}_0 = \hat{H} y$:
   $$\hat{\beta} \in \hat{\beta}_0 + \ker(H^T H)$$
4. Pour $p = 0$:
   $$H = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \qquad H^T H = n \qquad H^T y = \sum_{i=1}^{n} y_i \qquad \hat{\beta}_0 = \frac{1}{n} \sum_{i=1}^{n} y_i$$

   Notons $u = y - H_p\hat{\beta}_p$ et $v = H_p\hat{\beta}_p - H_0\hat{\beta}_0$, il suffit de montrer que $\langle u, v \rangle = 0$. Ce qui est immédiat par la question 3 (cf schéma).

$$R^2 = \frac{\|H_p\hat{\beta}_p - H_0\hat{\beta}_0\|^2}{\|y - H_0\hat{\beta}_0\|^2} \geqslant 0 \qquad 1 = R^2 + \frac{\|H_p\hat{\beta}_p - y\|^2}{\|y - H_0\hat{\beta}_0\|^2}$$

Donc $R^2 \in [0,1]$.

5. On suppose que $\mathrm{rg}(H) = p+1$. alors $\dim\ker H = 0 \;\Rightarrow\; \ker(H) = \{0\} = \ker(H^TH)$. On a donc unicité de la solution. Et comme H^TH est désormais inversible, on peut écrire :

   $$\hat{\beta} = (H^TH)^{-1}H^Ty$$

   Pour se ramener à ce cas il suffit de supprimer des colonnes.

   Pour calculer $\hat{\beta}$ en pratique, on peut effectuer une décomposition de Cholesky de H^TH comme C^TC avec C triangulaire supérieure.

6. Soit $A \in \mathbb{R}^{n\times(p-1)}$ avec $(p+1) \leqslant n$. Σ doit être de la forme :

   $$\Sigma = \begin{pmatrix} \sigma_1 & \dots & & 0 \\ & \ddots & & \\ 0 & \dots & \sigma_r & 0 \\ \vdots & & & \vdots \\ 0 & \dots & \dots & 0 \end{pmatrix}$$

   En considérant A^TA qui est symétrique, on a donc une décomposition $A^TA = V^TDV$ avec D diagonale :

   $$D = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_{p+1} \end{pmatrix}$$

   avec $d_1 \geqslant d_2 \geqslant \ldots \geqslant d_r \geqslant 0$. On pose alors $\sigma_i = \sqrt{d_i}$. On cherche donc U telle que

   $$AV = U\Sigma \qquad \text{en notant } V = [V_1 \ldots V_{p+1}] \quad U = [U_1 \ldots U_p]$$

   on se ramène à $AV_i = \sigma_i U_i$. Posons donc $\forall i \in \{1,\ldots,n\}$: $U_i = \frac{1}{\sigma_i}AV_i$.
   On peut remarquer que les U_i sont orthonormaux puisque :

   $$U_i^TU_j = \frac{1}{\sigma_i\sigma_j}V_i^TA^TAV_j = \frac{1}{\sigma_i\sigma_j}V_i^T(\sigma_j^2V_j) = \frac{\sigma_j}{\sigma_i}V_i^TV_j = \delta_{i,j}$$

   On complète alors avec $U_{r+1}\ldots U_n$ tels que $(U_1,\ldots,U_n)$ base orthonormée de $\mathbb{R}^n$. On vérifie alors $U^TAV = \Sigma$.

   Remarques. r est le rang de A et $\mathrm{Vect}(V_{r+1},\ldots,V_n) = \ker(A^TA) = \ker A$

   Ainsi :

   $$A = \sum_{i=1}^{r} \sigma_i U_i V_i^T$$

7. On a déjà vu que l'ensemble des solutions est un espace affine de la forme :

$$\hat{\beta}_0 + \ker(H)$$

Montrons que H^+y est une solution particulière.

$$H^T H H^+ y = V\Sigma U^T U \Sigma V^T V \Sigma^+ U^T y = V\Sigma^T \Sigma\Sigma^+ U^T y$$

Soit $y \in \mathbb{R}^n$, décomposons le dans la base de U^T : $y = \sum_{i=1}^n \tilde{y}_i u_i$. Alors :

$$U^T y = \begin{pmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_n \end{pmatrix} \qquad \Sigma\Sigma^+ U^T y = \begin{pmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_r \\ 0 \\ \vdots \end{pmatrix} \qquad \Sigma^T \Sigma\Sigma^+ U^T y = \begin{pmatrix} \sigma_1 \tilde{y}_1 \\ \vdots \\ \sigma_r \tilde{y}_r \\ 0 \\ \vdots \end{pmatrix}$$

Ainsi $V\Sigma^T\Sigma\Sigma^+U^T y = \sum_{i=1}^r \sigma_i \tilde{y}_i v_i$. Et d'un autre côté on a bien :

$$H^T y = V\Sigma^T U^T y = \sum_{i=1}^r \sigma_i \tilde{y}_i v_i$$

Caractérisation géométrique de H^+y.

$$\hat{\beta} \in \left\{ \sum_{i=1}^r \frac{\tilde{y}_i}{\sigma_i} v_i + \sum_{i=r+1}^{p+1} \alpha v_i,\ \alpha_i \in \mathbb{R} \right\}$$

$\|\hat{\beta}\|^2 = \|H^+y\|^2 + \|\pi_{\ker(H)}(\hat{\beta})\|^2$. Donc H^+y est la solution de norme minimale.

8. On prend désormais comme estimateur naturel de β : $\hat{\beta} = H^+y$. On suppose qu'il existe β^* tel que $y = H\beta^* + \varepsilon$. On se demande maintenant si notre estimateur approche β^*. Comme $\mathbb{E}(\varepsilon) = 0$, $\mathbb{E}(\hat{\beta}) = \mathbb{E}(H^+(H\beta^* + \varepsilon)) = H^+H\beta^* = \pi_{(\ker H)^\perp}(\beta^*) = \beta^*$. Puis comme :

$$\mathbb{E}(\|\hat{\beta} - \beta^*\|) = \mathrm{Var}(\hat{\beta}) + \mathrm{biais}(\hat{\beta}, \beta^*) = \mathrm{Var}(\hat{\beta})$$

Regardons la matrice de covariance de $\hat{\beta}$.

$$\hat{\beta} = H^+H\beta^* + H^+\varepsilon$$

On sait que $\mathbb{E}(\hat{\beta}) = \beta^*$.

Remarque. Soit $X \in \mathbb{R}^n$ un vecteur aléatoire. Soit M une matrice de $\mathbb{R}^{n\times n}$ une matrice constante. Notons $\Gamma = \mathbb{E}[\bar{X}\bar{X}^T)$ la matrice de covariance de X (où $\bar{X} = X - \mathbb{E}(X)$). Alors $\mathbb{E}(M\bar{X}\bar{X}^T M) = M\Gamma M^T$.

Ainsi dans notre cas :

$$\mathrm{Cov}(\hat{\beta}) = \sigma^2 H^+(H^+)^T = \sigma^2 V\Sigma^+U^TU(\Sigma^+)^TV^T = \sigma^2 V\Sigma^+(\Sigma^+)^TV^T$$

Donc $\mathbb{E}(\|\hat{\beta} - \beta^*\|^2) = \mathrm{Tr}(\mathrm{Cov}(\hat{\beta})) = \sigma^2\mathrm{Tr}(V\Sigma^+(\Sigma^+)^TV^T) = \sigma^2\sum_{i=1}^r \frac{1}{\sigma_i^2}$. On a donc une très grande variance s'il y a des valeurs singulières très petites.