Examen Master SICOM, Cours Identification
Luc Pronzato, Décembre 2004

QUELQUES INDICATIONS

Les parties 2 et 3 sont totalement indépendantes. Dans la partie 3, les questions a, b, c et d sont largement indépendantes.

On prendra soin de **souligner les vecteurs** et de noter les matrices par des lettres capitales. Un vecteur $\mathbf{v}$ est toujours supposé être du type colonne, on notera $\mathbf{v}^\top$ un vecteur ligne.

ÉNONCÉ DU PROBLÈME

On considère une suite d'observations $y_k = y(x_k)$, $k = 1,2,\ldots$ dans un modèle de régression linéaire pour lequel

$$y(x_k) = \mathbf{r}^\top(x_k)\bar{\theta} + \varepsilon_k$$

avec $\bar{\theta} \in I\!R^p$ la vraie valeur (inconnue) des paramètres du modèle et (ε_k) une suite de variables aléatoires indépendantes et identiquement distribuées, de moyenne zéro et de variance σ^2. On notera $\mathbf{r}_i = \mathbf{r}(x_i)$ pour tout i et

$$\mathbf{X}_k = \begin{pmatrix} \mathbf{r}_1^\top \\ \vdots \\ \mathbf{r}_k^\top \end{pmatrix}, \ \mathbf{y}_1^k = (y_1,\ldots,y_k)^\top \text{ et } \mathbf{M}_k = \mathbf{X}_k^\top \mathbf{X}_k \text{ pour tout } k\,.$$

1) ESTIMATEUR DES MOINDRES CARRÉS

a) Donner l'expression de l'estimateur des moindres carrés $\hat{\theta}^k$ après observation de $\mathbf{y}_1^k$ en fonction de $\mathbf{X}_k$ et $\mathbf{y}_1^k$.

Quelle est sa moyenne $\mathtt{E}\{\hat{\theta}^k\}$?

Que vaut sa covariance $\mathrm{Var}\{\hat{\theta}^k\}$? (On la notera $\mathbf{C}_k$.)

b) On effectue une observation supplémentaire y_{k+1}. Exprimer $\hat{\theta}^{k+1}$ en fonction de $\mathbf{M}_{k+1}$, $\mathbf{X}_k$, $\mathbf{y}_1^k$ et $y_{k+1}\mathbf{r}_{k+1}$. En exprimant $\mathbf{M}_k$ en fonction de $\mathbf{M}_{k+1}$ et $\mathbf{r}_{k+1}$, en déduire la forme récursive de l'estimateur

$$\hat{\theta}^{k+1} = \hat{\theta}^k + \delta_k$$

où l'on exprimera le terme correctif δ_k en fonction de $\mathbf{M}_{k+1}$, $\mathbf{r}_{k+1}$ et de l'erreur de prédiction $e_{k+1}(\hat{\theta}^k) = y_{k+1} - \mathbf{r}_{k+1}^\top\hat{\theta}^k$.

c) Enfin, à partir du lemme d'inversion matricielle $(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$ (où l'on suppose inversibles les matrices à inverser...) on exprimera δ_k en fonction de $\mathbf{M}_k$, $\mathbf{r}_{k+1}$ et $e_{k+1}(\hat{\theta}^k)$.

En déduire la forme récursive de l'estimateur des moindres carrés

$$\begin{aligned} \mathbf{C}_{k+1} &= \mathbf{C}_k - \ldots \\ \hat{\theta}^{k+1} &= \hat{\theta}^k + \ldots \end{aligned}$$

où les membres de droite ne font intervenir que $\mathbf{C}_k$, $\mathbf{r}_{k+1}$, $e_{k+1}(\hat{\theta}^k)$ et σ^2.

2) ESTIMATEUR LMS (*Least Mean Squares*)

a) Considérons à présent l'algorithme de calcul suivant :

$$\tilde{\theta}^{k+1} = \tilde{\theta}^k + \lambda_{k+1}\mathbf{r}_{k+1}e_{k+1}(\tilde{\theta}^k)\,,$$

avec $\tilde{\theta}^k$ l'estimée après observation de $\mathbf{y}_1^k$ (et *a priori* $\tilde{\theta}^k \neq \hat{\theta}^k$) et λ_{k+1} un pas scalaire. Pourquoi est-il plus simple à mettre en œuvre que l'algorithme des moindres carrés récursifs ?

Supposons que $\mathtt{E}\{\tilde{\theta}^k\} = \bar{\theta}$. Que vaut $\mathtt{E}\{\tilde{\theta}^{k+1}\}$? (On rappelle que ε_{k+1} est indépendant de $\mathbf{y}_1^k$.)

Supposons que $\mathrm{Var}\{\tilde{\theta}^k\} = \mathbf{C}_k$. Exprimer $\mathbf{V}_{k+1} = \mathrm{Var}\{\tilde{\theta}^{k+1}\}$ en fonction de $\mathbf{C}_k$, $\mathbf{r}_{k+1}$, λ_{k+1} et σ^2.

b) On suppose que $\|\mathbf{r}_{k+1}\| = 1$. Calculer λ_{k+1}^*, la valeur de λ_{k+1} qui minimise $\mathrm{trace}(\mathbf{V}_{k+1})$, en fonction de $\rho_{k+1} = \sigma^2/(\mathbf{r}_{k+1}^\top\mathbf{C}_k\mathbf{r}_{k+1})$.

Calculer $\mathrm{trace}(\mathbf{C}_{k+1})$, avec $\mathbf{C}_{k+1}$ la matrice de covariance de l'estimateur des moindres carrés $\hat{\theta}^{k+1}$.

Utiliser l'inégalité de Cauchy-Schwarz pour conclure quant aux performances des deux estimateurs (pour le critère trace(Var)).

3) PLANIFICATION D'EXPÉRIENCES

On utilise de nouveau l'estimateur des moindres carrés.

a) Lorsque $\mathbf{C}_k$ est fixée et que la seule contrainte sur $\mathbf{r}_{k+1}$ est $\|\mathbf{r}_{k+1}\| \leq r$, comment choisir $\mathbf{r}_{k+1}$ qui minimise $\det(\mathbf{C}_{k+1})$? (On rappelle que $\det(A+\mathbf{v}\mathbf{v}^\top) = (1+\mathbf{v}^\top A^{-1}\mathbf{v})\det(A)$.) On notera $\mathbf{r}^*_{k+1}$ ce choix, que l'on exprimera en fonction des vecteurs propres de $\mathbf{C}_k$.

b) On s'intéresse à la prédiction de la réponse du modèle en un point x. Après observation de $\mathbf{y}_1^k$, on prédit cette réponse par $\hat{y}^k(x) = \mathbf{r}^\top(x)\hat{\theta}^k$. Calculer la moyenne de cette prédiction en un x donné, $\mathtt{E}\{\hat{y}^k(x)|x\}$, ainsi que sa variance var$\{\hat{y}^k(x)|x\}$. Justifier le choix $\mathbf{r}^*_{k+1}$ de la question a).

c) On suppose que $\mathbf{r}(x) = \begin{pmatrix} 1 \\ x \end{pmatrix}$. On souhaite choisir les vecteurs $\mathbf{r}(x_i)$ qui forment $\mathbf{X}_k$, avec x_i appartenant à l'intervalle $[-1,1]$, $i = 1,\ldots,k$, de façon à minimiser

$$d(x^*,\mathbf{X}_k) = \max_{x\in[-1,1]} \text{var}\{\hat{y}^k(x)|x\} \,.$$

On pose

$$\mathbf{C}_k = \begin{pmatrix} c^2 & b \\ b & a^2 \end{pmatrix} .$$

Justifier l'écriture des termes diagonaux sous forme de carrés et montrer que a,b et c satisfont une inégalité élémentaire. Montrer que $d(x,\mathbf{X}_k) = \text{var}\{\hat{y}^k(x)|x\}$ est nécessairement maximum en -1 ou en +1. En déduire l'expression de $d(x^*,\mathbf{X}_k)$ en fonction de a,b et c. Exprimer a,b et c en fonction de

$$\mu_1 = \sum_{i=1}^k x_i \,,\, \mu_2 = \sum_{i=1}^k x_i^2 \,,\, \sigma^2 \text{ et } k \,.$$

En déduire les valeurs de μ_1 et μ_2 qui minimisent $d(x^*,\mathbf{X}_k)$. Comment réaliser cette expérience quand k est un nombre pair ?

d) On s'intéresse à présent à la moyenne par rapport à x de l'erreur quadratique moyenne de la prédiction en x, lorsque x a une densité de probabilité $\varphi(x)$. Montrer que le critère s'écrit

$$\Phi(\mathbf{X}_k) = \int \text{var}\{\hat{y}^k(x)|x\}\,\varphi(x)\,dx$$

puis exprimer $\Phi(\mathbf{X}_k)$ en fonction de $\mathbf{C}_k$ et de

$$\mathbf{V}_x = \int \mathbf{r}(x)\mathbf{r}^\top(x)\,\varphi(x)\,dx \,.$$

- On suppose que φ est de moyenne 0 et de variance s^2.
  Exprimer $\Phi(\mathbf{X}_k)$ en fonction de μ_1, μ_2, σ^2, k et s^2. On suppose de nouveau que k est pair. Comparer aux résultats de la question c).
- On suppose que φ est de moyenne 1 et de variance s^2. Exprimer de nouveau $\Phi(\mathbf{X}_k)$ en fonction de μ_1, μ_2, σ^2, k et s^2. On fait tendre k vers l'infini et on suppose qu'une proportion α des observations sont faites en $x_i = -1$ et une proportion $1-\alpha$ en $x_i = 1$. Exprimer $\psi(\alpha) = \lim_{k\to\infty} k\Phi(\mathbf{X}_k)$ en fonction de α, σ^2 et s^2 et déterminer la valeur optimale $\alpha^*(s)$ qui minimise $\psi(\alpha)$ (justifier qu'il s'agit bien d'un minimum).
  Que se passe-t-il quand s tend vers 0 ? Justifier la valeur limite de α^*.
  Que se passe-t-il quand $s \to \infty$?