Université de Rennes 1
Master recherche STI
Examen du cours UE 14
“Filtre de Kalman et modèles de Markov cachés”
Jeudi 25 janvier 2007, 10:30 à 12:30

Problème

L'objectif de ce problème est d'étudier un algorithme récemment apparu en apprentissage automatique sous le nom d'*estimation bayésienne variationnelle*, pour estimer récursivement un paramètre inconnu dans un modèle à variables cachées. On se place dans le cadre de l'estimation bayésienne, c'est–à–dire que le paramètre inconnu est supposé aléatoire, et l'incertitude sur la valeur de ce paramètre est décrite par une distribution de probabilité *a priori*, supposée connue.

Soit X et Y deux vecteurs aléatoires, de dimension m et d respectivement. Comme d'habitude, la composante X est cachée et on observe seulement la composante Y. On suppose connue, à un paramètre inconnu A près, la densité de probabilité conjointe du couple (X, Y), notée $p_{X,Y|A=a}(x, y)$. Le paramètre inconnu est considéré comme une variable aléatoire de dimension s, et on se donne sa densité de probabilité $p_A(a)$.

(i) **Donner l'expression de la densité de probabilité conjointe du triplet (X, A, Y), et montrer que la densité de probabilité marginale du vecteur aléatoire Y s'écrit**

$$p_Y(y) = \int_{\mathbb{R}^m} \int_{\mathbb{R}^s} dx\, da\; p_{X,Y|A=a}(x, y)\; p_A(a) \ .$$

On veut calculer la densité de probabilité *a posteriori* du paramètre, c'est–à–dire la densité de probabilité conditionnelle, notée $p_{A|Y=y}(a)$, du vecteur aléatoire A sachant l'observation Y. A cet effet, il suffit de calculer la densité de probabilité jointe *a posteriori* de l'état caché et du paramètre, c'est–à–dire la densité de probabilité conditionnelle jointe, notée $p_{X,A|Y=y}(x, a)$, du vecteur aléatoire (X, A) sachant l'observation Y.

(ii) **Soit $q(x, a)$ une densité de probabilité arbitraire sur $\mathbb{R}^m \times \mathbb{R}^s$. Montrer que**

$$\log p_Y(y) \geq \int_{\mathbb{R}^m} \int_{\mathbb{R}^s} dx\, da\; q(x, a)\; \log \frac{p_{X,Y|A=a}(x, y)\; p_A(a)}{q(x, a)} \ .$$

Indication : on utilisera l'inégalité de Jensen suivante

$$\mathbb{E}[\Phi(Z)] \leq \Phi(\mathbb{E}[Z]) \ ,$$

valable pour toute variable aléatoire Z et pour toute fonction concave Φ (par exemple $z \mapsto \log z$).

(iii) **Montrer le maximum de la borne inférieure**

$$\int_{\mathbb{R}^m} \int_{\mathbb{R}^s} dx\, da\ q(x,a)\ \log \frac{p_{X,Y|A=a}(x,y)\ p_A(a)}{q(x,a)} \ ,$$

obtenue à la question (ii), par rapport à la densité de probabilité $q(x,a)$, est atteint précisément pour $q(x,a) = p_{X,A|Y=y}(x,a)$.

A partir de maintenant, l'objectif se résume à trouver un algorithme permettant de trouver le maximum de la borne inférieure par rapport à la densité de probabilité $q(x,a)$. L'idée de l'*estimation bayésienne variationnelle* est de rechercher $q(x,a)$ sous la forme factorisée $q(x,a) = q_X(x)\ q_A(a)$ et de maximiser itérativement par rapport à $q_X(x)$ puis par rapport à $q_A(a)$, et ainsi de suite. En d'autres termes, les solutions $q_X^{(p)}(x)$ et $q_A^{(p)}(a)$ à l'itération p, sont définies à partir des solutions $q_X^{(p-1)}(x)$ et $q_A^{(p-1)}(a)$ à l'itération $(p-1)$, à l'aide des formules de mise–à–jour suivantes

$$q_X^{(p)}(x) = \operatorname*{argmax}_{q_X(\cdot)} \int_{\mathbb{R}^m} \int_{\mathbb{R}^s} dx\, da\ q_X(x)\ q_A^{(p-1)}(a)\ \log \frac{p_{X,Y|A=a}(x,y)\ p_A(a)}{q_X(x)\ q_A^{(p-1)}(a)} \ ,$$

et

$$q_A^{(p)}(a) = \operatorname*{argmax}_{q_A(\cdot)} \int_{\mathbb{R}^m} \int_{\mathbb{R}^s} dx\, da\ q_X^{(p)}(x)\ q_A(a)\ \log \frac{p_{X,Y|A=a}(x,y)\ p_A(a)}{q_X^{(p)}(x)\ q_A(a)} \ .$$

On espère ainsi que la densité de probabilité $q_A^{(p)}(a)$ converge, quand le nombre p d'itérations tend vers l'infini, vers la densité de probabilité *a posteriori* du paramètre, c'est–à–dire vers la densité de probabilité $p_{A|Y=y}(a)$ du vecteur aléatoire A sachant l'observation Y.

(iv) **Montrer que les deux étapes définies ci–dessus admettent les solutions explicites suivantes**

$$q_X^{(p)}(x) = c_X\ \exp\Big\{ \int_{\mathbb{R}^s} da\ q_A^{(p-1)}(a)\ \log p_{X,Y|A=a}(x,y) \Big\} \ ,$$

et

$$q_A^{(p)}(a) = c_A\ p_A(a)\ \exp\Big\{ \int_{\mathbb{R}^m} dx\ q_X^{(p)}(x)\ \log p_{X,Y|A=a}(x,y) \Big\} \ ,$$

où les constantes c_X et c_A désignent des constantes de normalisation.

A titre d'exemple, on considère le cas particulier suivant

$$\begin{cases} X = M + W \ , \\ \\ Y = H\, X + V \ , \end{cases}$$

où le paramètre inconnu $A = (M, H)$ est supposé aléatoire. On fait les hypothèses suivantes

- le vecteur aléatoire W est gaussien centré, de matrice de covariance identité,
- le vecteur aléatoire V est gaussien centré, de matrice de covariance identité,
- la densité de probabilité *a priori* $p_A(m, h)$ du vecteur aléatoire $A = (M, H)$ est gaussienne, de moyenne $(\bar{M}, \bar{H})$ et de matrice de covariance identité,
- les vecteurs aléatoires W, V et A sont mutuellement indépendants.

(v) **Expliquer pourquoi le problème de l'estimation du vecteur aléatoire (X, A) sachant Y ne rentre pas dans le cadre de l'estimation MMSE pour les vecteurs aléatoires gaussiens.**

(vi) **Monter par récurrence sur l'indice p que les solutions successives $q_X^{(p)}(x)$ et $q_A^{(p)}(a)$ sont des densités de probabilité gaussiennes, dont on caractérisera la moyenne et la matrice de covariance.**

En revenant au cas général, on peut poser alternativement

$$q_X^{(p)}(x) = r_X^{(p)}(x)\; p_{X|Y=y,A=a_0}(x) = r_X^{(p)}(x)\; \frac{p_{X,Y|A=a_0}(x,y)}{p_{Y|A=a_0}(y)} \ ,$$

où a_0 est une valeur fixe du paramètre, et où l'inconnue est maintenant $r_X^{(p)}(x)$.

(vii) **Montrer que les solutions données à la question (iv) peuvent se ré–écrire**

$$r_X^{(p)}(x) = c_X\ \exp\Big\{ \int_{\mathbb{R}^s} da\ q_A^{(p-1)}(a)\ \log \frac{p_{X,Y|A=a}(x,y)}{p_{X,Y|A=a_0}(x,y)} \Big\} \ ,$$

et

$$\begin{aligned} q_A^{(p)}(a) &= c_A\, p_A(a)\ \exp\Big\{ \int_{\mathbb{R}^m} dx\ r_X^{(p)}(x)\ p_{X|Y=y,A=a_0}(x)\ \log \frac{p_{X,Y|A=a}(x,y)}{p_{X,Y|A=a_0}(x,y)} \Big\} \\ &= c_A\, p_A(a)\ \exp\Big\{ \mathbb{E}[r_X^{(p)}(X)\ \log \frac{p_{X,Y|A=a}(X,Y)}{p_{X,Y|A=a_0}(X,Y)} \mid Y=y, A=a_0] \Big\} \ , \end{aligned}$$

où les constantes c_X et c_A désignent d'autres constantes de normalisation.

On considère finalement le cas d'un système linéaire gaussien

$$\begin{cases} X_k = F\, X_{k-1} + W_k \ , \\ \\ Y_k = H\, X_k + V_k \ , \end{cases}$$

où le paramètre inconnu $A = (F, H)$ est supposé aléatoire. On fait les hypothèses suivantes :

- le vecteur aléatoire X_0 est gaussien, de moyenne $\bar{X}_0$ et de matrice de covariance Q_0^X,
- la suite aléatoire $\{W_k\}$ est un bruit blanc gaussien centré de matrice de covariance identité,
- la suite aléatoire $\{V_k\}$ est un bruit blanc gaussien centré de matrice de covariance identité,
- la densité de probabilité *a priori* $p_A(f, h)$ du vecteur aléatoire $A = (F, H)$ est gaussienne, de moyenne $(\bar{F}, \bar{H})$ et de matrice de covariance identité,
- les vecteurs / suites aléatoires X_0, $A = (F, H)$, $\{W_k\}$ et $\{V_k\}$ sont mutuellement indépendants.

(viii) **En utilisant les expressions obtenues à la question (vii), montrer que la mise en œuvre de l'algorithme précédent pour l'estimation du paramètre inconnu $A = (F, H)$ à partir des observations $Y_{0:n}$ se ramène à chaque itération au calcul de l'espérance conditionnelle d'une certaine fonction des états cachés $X_{0:n}$ sachant les observations $Y_{0:n}$ et pour une valeur particulière $A_0 = (F_0, H_0)$ du paramètre.**