Remarques correction TD Exercice 24, Mapsi

Formalisation du probleme

Nous allons poser explicitement les variables cachées et observées du problème. On ne vous demande pas de savoir refaire cette formalisation. Le but de cette formalisation est de ramener l'exercice à une application directe de l'algorithme EM vu en cours.

On distingue trois cas : $i = 0$, A et B sont observés ; $i = 1$, seul B est observé et A reste caché ; et $i = 2$, seul A est observé et B reste caché.

Variables cachées

La variable cachée est un couple (x, i) dont les valeurs dépendent des cas :

— Si $i = 0$,
$$x^h = (x = 1, i = 0)$$

— Si $i = 1$,
$$x^h = (x \in \{a_1, a_2\}, i = 1)$$
donc la première variable (la variable A) n'est pas observée.

— Si $i = 2$,
$$x^h = (x \in \{b_1, b_2\}, i = 2)$$
donc la deuxième variable (la variable B) n'est pas observée.

Variables observées

Les variables observées sont $x^o = (y, i)$, selon les cas :

— Si $i = 0$,
$$x^o = (y \in \{a_1, a_2\} \times \{b_1, b_2\}, i = 0)$$
c'est-à-dire que y est un couple (les variables A et B sont toutes deux observées).

— Si $i = 1$,
$$x^o = (y \in \{b_1, b_2\}, i = 1)$$
donc seule la deuxième variable (B) est observée.

— Si $i = 2$,
$$x^o = (y \in \{a_1, a_2\}, i = 2)$$
donc seule la première variable (A) est observée.

Loi jointe

On se donne une loi jointe sur A, B qu'on note $Q_\theta(a,b)$ avec $a = a_1$ ou a_2 et $b = b_1$ ou b_2.

La loi jointe sur les variables cachées et observées est dénotée $P_\theta(x^h, x^o)$ et est définie selon les cas : considérons $a \in \{a_1, a_2\}$ et $b \in \{b_1, b_2\}$,

— Si $i = 0$,
$$P_\theta(x^h, x^o) = P_\theta\left((1, i=0), ((a,b), i=0)\right) \propto Q_\theta(a,b)$$
où $\propto$ signifie "proportionnel à".

— Si $i = 1$,
$$P_\theta(x^h, x^o) = P_\theta\left((a, i=1), (b, i=1)\right) \propto Q_\theta(a,b)$$

— Si $i = 2$,
$$P_\theta(x^h, x^o) = P_\theta\left((b, i=2), (a, i=2)\right) \propto Q_\theta(a,b)$$

La loi jointe sur x^h, x^o est entièrement déterminée pas $Q_\theta(a,b)$ où a peut prendre les valeurs a_1, a_2 et b les valeurs b_1, b_2.

Lois conditionnelles

En fonction des différents cas, la loi conditionnelle sur les variables observées x^o est :

— Si $i = 0$:
$$P_\theta(x^h \mid x^o) = P_\theta\left((1, i=0) \mid (a,b), i=0\right) = 1$$

— Si $i = 1$:
$$P_\theta(x^h \mid x^o) = P_\theta\left((a, i=1) \mid (b, i=1)\right) = \frac{P_\theta\left((a, i=1) \mid (b, i=1)\right)}{\sum_a P_\theta\left((a, i=1) \mid (b, i=1)\right)}$$
Mais comme
$$P_\theta\left((a, i=1), (b, i=1)\right) \propto Q_\theta(a,b)$$
donc,
$$\frac{P_\theta\left((a, i=1) \mid (b, i=1)\right)}{\sum_a P_\theta\left((a, i=1) \mid (b, i=1)\right)} = \frac{Q_\theta(a,b)}{\sum_a Q_\theta(a,b)}$$
Par conséquent,
$$P_\theta\left((a, i=1) \mid (b, i=1)\right) = \frac{Q_\theta(a,b)}{\sum_a Q_\theta(a,b)} = Q_\theta(A = a \mid B = b)$$

— Si $i = 2$: De manière similaire, avec le même argument que dans le cas $i = 1$, nous obtenons :
$$P_\theta(x^h \mid x^o) = P_\theta\left((b, i=2) \mid (a, i=2)\right) = Q_\theta(B = b \mid A = a)$$

Ce sont les quantités calculées dans cette dernière sous-section 'Lois conditionnelles' qui vont nous servir à déterminer $Q_i^{t+1}(x^h)$ car $Q_i^{t+1}(x^h) = P_{\theta^t}(x^h | x_i^o)$.

1 Réponses aux questions

Q24.1

Θ^t est l'ensemble des valeurs possibles que peut prendre Q_θ. Notons $Q_\theta(a_1,b_1) := \theta_{11}$, $Q_\theta(a_1,b_2) := \theta_{12}$, $Q_\theta(a_2,b_1) := \theta_{21}$, et $Q_\theta(a_2,b_2) := \theta_{22}$.

Q_θ est une distribution de probabilité sur $\{a_1,a_2\} \times \{b_1,b_2\}$, donc

$$\sum_{a,b} Q_\theta(a,b) = 1$$

Ainsi, $\theta_{22} = 1 - \theta_{11} - \theta_{12} - \theta_{21}$. Donc $\Theta^t = (\theta_{11},\theta_{12},\theta_{21},\theta_{22})$.

Appliquer votre cours : slides 17 cours 4 EM.

Q24.2

On pose $x_1^o := (a_1,b_1)$, $x_2^o := (a_2,b_1)$, $x_3^o := (?,b_2)$, $x_4^o := (a_2,?)$, $x_5^o := (a_2,b_2)$, $x_6^o := (a_1,b_2)$, $x_7^o := (?,b_1)$, $x_8^o := (a_1,?)$, $x_9^o := (a_1,b_2)$, $x_{10}^o := (a_2,b_1)$.

On a $Q_1^{t+1}(x^h) = P_\theta(x^h|x_1^o)$ donc $Q_1^{t+1}(1) = 1$.

De même $Q_2^{t+1}(x^h) = 1$

Pour : $Q_3^{t+1}(x^h)$ on a $Q_3^{t+1}(a_1) = Q_3^{t+1}(a_2) = 0,5$.
Pour : $Q_4^{t+1}(x^h)$ on a $Q_4^{t+1}(b_1) = Q_4^{t+1}(b_2) = 0,5$.

$Q_5^{t+1}(1) = 1$; $Q_6^{t+1}(1) = 1$; $Q_7^{t+1}(a_1) = Q_7^{t+1}(a_2) = 0.5$; $Q_8^{t+1}(b_1) = Q_8^{t+1}(b_2) = 0.5$; $Q_9^{t+1}(1) = 1$; $Q_{10}^{t+1}(1) = 1$.

Q24.3

$$\log L^{t+1}(\mathbf{x}^o,\Theta) = 2\log\theta_{11} + 3\log\theta_{12} + 3\log\theta_{21} + 2\log\theta_{22} - 4\log 0.5 \tag{1}$$

Q24.4

Dans l'expression précédente, l'équation 1, on remplace θ_{22} par $\theta_{22} = 1 - \theta_{11} - \theta_{12} - \theta_{21}$; il n'y a plus de contraintes sur θ_{11}, θ_{12}, et θ_{21} :

$$\log L^{t+1}(\mathbf{x}^o,\theta_{11},\theta_{12},\theta_{21}) = 2\log\theta_{11} + 3\log\theta_{12} + 3\log\theta_{21} + 2\log(1-\theta_{11}-\theta_{12}-\theta_{21}) - 4\log 0.5 \tag{2}$$

On annule le gradient de la fonction précédente :

$$\nabla \log L^{t+1} = \left(\frac{\partial \log L^{t+1}}{\partial\theta_{11}}, \frac{\partial \log L^{t+1}}{\partial\theta_{12}}, \frac{\partial \log L^{t+1}}{\partial\theta_{21}}\right)$$

Θ^1 est tel que les trois conditions $\frac{\partial \log L^{t+1}}{\partial \theta_{11}} = 0$, $\frac{\partial \log L^{t+1}}{\partial \theta_{21}} = 0$, et $\frac{\partial \log L^{t+1}}{\partial \theta_{12}} = 0$ sont satisfaites, ce qui nous dit que :

$$\Theta^1 = \{\theta_{11} = 0.2, \theta_{12} = 0.3, \theta_{21} = 0.3, \theta_{22} = 0.2\}$$

Q24.5

$Q_{\theta^t}(A|B = b_1) = (0.4, 0.6)$; $Q_{\theta^t}(A|B = b_2) = (0.6, 0.4)$.

$Q_{\theta^t}(B|A = a_1) = (0.4, 0.6)$; $Q_{\theta^t}(B|A = a_2) = (0.6, 0.4)$.

Q24.6

$\log L^{t+1}(\mathbf{x}^o, \Theta) = 1.8 \log \theta_{11} + 3.2 \log \theta_{12} + 3.2 \log \theta_{21} + 1.8 \log \theta_{22} - 2.4 \log 0.6 - 1.6 \log 0.4$