Devoir maison Algorithmes stochastiques

BOUACHA LAZHAR

8 Novembre 2020

Partie théorique

Question 1 :

La proportion d'étudiants qui répond 1 à la question m peut s'écrire

$$q_m = \frac{1}{N} \sum_{n=1}^{N} x_{n,m}.$$

Si cette proposition est supérieure à 0.5, il y a une majorité d'étudiants ayant répondu 1 à la question m. Sinon il y a une majorité d'étudiants ayant répondu 0 à la question m. On en déduit des estimateurs $\hat{t}_m$ des vraies réponses aux questions

$$t_m = \begin{cases} 1 & \text{si, } q_m > 0.5 \text{ une majorité d'élèves a répondu 1 à la question } m \\ 0 & \text{sinon.} \end{cases}$$

Question 2 :

On note $F(\mathbf{x}, \mathbf{t}, \theta) = \log p(\forall m, T_m = t_m, \text{ et } \forall n, m,\ X_{n,m} = x_{n,m} | \theta)$.

Les réponses aux différentes questions sont indépendantes, donc on peut écrire, $F(\mathbf{x}, \mathbf{t}, \theta) = \log \prod_{m=1}^{M} p(T_m = t_m, \text{ et } \forall n, X_{n,m} = x_{n,m} | \theta)$

En utilisant la formule de Bayes, on obtient ensuite :

$$\begin{aligned} F(\mathbf{x},\mathbf{t},\theta) &= \log \prod_{m=1}^{M} p(\forall n, X_{n,m} = x_{n,m} | T_m = t_m, \theta) p(T_m = t_m) && (1) \\ &= M \log 0.5 + \log \prod_{m=1}^{M} p(\forall n, X_{n,m} = x_{n,m} | T_m = t_m, \theta) && (2) \end{aligned}$$

Puis par indépendance des réponses des élèves entre elles, et en remarquant que
$p(X_{n,m} = x_{n,m} | T_m = t_m, \theta) = \theta_n^{\mathbf{1}_{x_{n,m}=t_m}} (1-\theta_n)^{\mathbf{1}_{x_{n,m}=1-t_m}}$,

$$\begin{aligned} F(\mathbf{x},\mathbf{t},\theta) &= M \log 0.5 + \log \prod_{m=1}^{M} \prod_{n=1}^{N} p(X_{n,m} = x_{n,m} | T_m = t_m, \theta) && (3) \\ &= M \log 0.5 + \sum_{n=1}^{N} \sum_{m=1}^{M} \log \left(\theta_n^{\mathbf{1}_{x_{n,m}=t_m}} (1-\theta_n)^{\mathbf{1}_{x_{n,m}=1-t_m}} \right) && (4) \end{aligned}$$

Question 3 :

a $= 1$: $\mathbf{1}_{a=1} = 1 + (1-1)(1-1) = 1$
a $= 0$: $\mathbf{1}_{a=0} = 0 + (1-0)(1-0) = 1$

Question 4 :

Lorsqu'on dérive $F(\mathbf{x},\mathbf{t},\theta)$ par rapport θ_n on obtient

$$\frac{\partial F}{\partial \theta_n} = \sum_{m=1}^{M} \left(\frac{1_{x_{n,m}=t_m}}{\theta_n} - \frac{1_{x_{n,m}=1-t_m}}{1-\theta_n} \right)$$

En regardant les points d'annulation de cette dérivée on en déduit l'estimateur suivant pour θ_n :

$\hat{\theta}_n = \frac{1}{M} \sum_{m=1}^{M} \mathbf{1}_{x_{n,m}=t_m} = \frac{1}{M} \sum_{m=1}^{M} (x_{n,m} t_m + (1-x_{n,m})(1-t_m))$

Question 5 :

On remarque que :

$$\mathbb{E}[T_m|\mathbf{x},\theta] = \mathbb{P}[T_m = 1|\mathbf{x},\theta] = \frac{\mathbb{P}[T_m = 1 \text{ et } \mathbf{x} = x|\theta]}{\mathbb{P}[\mathbf{x} = x|\theta]} \tag{5}$$

$$= \frac{\mathbb{P}[T_m = 1 \text{ et } \mathbf{x} = x|\theta]}{\mathbb{P}[T_m = 1 \text{ et } \mathbf{x} = x|\theta] + \mathbb{P}[T_m = 0 \text{ et } \mathbf{x} = x|\theta]} \tag{6}$$

Donc on a

$$\mathbb{P}[T_m = 1 \text{ et } \mathbf{x} = x|\theta] = \mathbb{P}[\mathbf{x} = x|T_m = 1, \theta]\mathbb{P}[T_m = 1] \tag{7}$$

$$= 0.5 \prod_{n=1}^{N} \theta_n^{x_{n,m}} (1 - \theta_n)^{1 - x_{n,m}} \tag{8}$$

$$\mathbb{P}[T_m = 0 \text{ et } \mathbf{x} = x|\theta] = \mathbb{P}[\mathbf{x} = x|T_m = 0, \theta]\mathbb{P}[T_m = 0] \tag{9}$$

$$= 0.5 \prod_{n=1}^{N} \theta_n^{1 - x_{n,m}} (1 - \theta_n)^{x_{n,m}} \tag{10}$$

Question 6 :

1. Initialiser les estimateurs de $\mathbf{t}$, $t_m = \begin{cases} 1 & \text{si } q_m > 0.5 \\ 0 & \text{sinon.} \end{cases}$

2. Pour i allant de 1 à N_{iter} (nombre d'itérations de l'algorithme)

- Étape M : mettre à jour le vecteur θ avec pour tout n allant de 1 à N

$$\theta_n = \frac{1}{M} \sum_{m=1}^{M} (x_{n,m} t_m + (1 - x_{n,m})(1 - t_m))$$

- Étape E : mettre à jour le vecteur $\mathbf{t}$, avec pour tout m allant de 1 à M

$$t_m = \frac{\prod_{n=1}^{N} \theta_n^{x_{n,m}} (1 - \theta_n)^{1 - x_{n,m}}}{\prod_{n=1}^{N} \theta_n^{x_{n,m}} (1 - \theta_n)^{1 - x_{n,m}} + \prod_{n=1}^{N} \theta_n^{1 - x_{n,m}} (1 - \theta_n)^{x_{n,m}}}$$

3. A la fin de l'algorithme, seuiller tous les coefficients du vecteur $\mathbf{t}$ (qui prend ses valeurs entre 0 et 1) à 0.5 pour obtenir un vecteur binaire.