COMPTE-RENDU DU DEVOIR 1

Vincent Matthys

1 Leaning in discrete graphical models

Etant donné que z et x sont des variables à valeurs discrètes prenant respectivement M et K valeurs, on peut procéder au *one hot encoding*, notant respectivement $\boldsymbol{Z}$ et $\boldsymbol{X}$ leur encodage sur, respectivement $\mathbb{R}^M$ et $\mathbb{R}^K$. Ainsi, on peut écrire :

$$p(z = m) = P(Z_m = 1) = \pi_m$$
$$p(x = k|z = m) = p(X_k = 1|Z_m = 1) = \theta_{mk}$$

En supposant que l'on ait un échantillon de n observations de (x, z), on peut exprimer la probabilité jointe d'une obersvation i :

$$\begin{aligned} p(x^{(i)}, z^{(i)}; \boldsymbol{\pi}, \boldsymbol{\theta}) &= p(z^{(i)}; \boldsymbol{\pi}) p(x^{(i)}|z^{(i)}; \boldsymbol{\theta}) \\ p(x^{(i)}, z^{(i)}; \boldsymbol{\pi}, \boldsymbol{\theta}) &= \prod_{m=1}^{M} \prod_{k=1}^{K} \theta_{mk}^{X_k^{(i)} Z_m^{(i)}} \prod_{l=1}^{M} \pi_l^{Z_l^{(i)}} \end{aligned} \tag{1}$$

On peut alors écrire la log-vraisemblance de l'échantillon dans le modèle $(\boldsymbol{\pi}, \boldsymbol{\theta})$, composé d'observations i.i.d., en utilisant (1) :

$$\begin{aligned} \ell(\boldsymbol{\pi}, \boldsymbol{\theta}) &= \sum_{i=1}^{n} \ln\big(p(x^{(i)}, z^{(i)}; \boldsymbol{\pi}, \boldsymbol{\theta})\big) \\ &= \sum_{i=1}^{n} \left(\ln \left(\prod_{m=1}^{M} \prod_{k=1}^{K} \theta_{mk}^{X_k^{(i)} Z_m^{(i)}} \prod_{l=1}^{M} \pi_l^{Z_l^{(i)}} \right) \right) \\ &= \sum_{i=1}^{n} \left(\sum_{m=1}^{M} \sum_{k=1}^{K} \ln \left(\theta_{mk}^{X_k^{(i)} Z_m^{(i)}} \right) + \sum_{l=1}^{M} \ln \left(\pi_l^{Z_l^{(i)}} \right) \right) \\ &= \sum_{i=1}^{n} \left(\sum_{m=1}^{M} \sum_{k=1}^{K} X_k^{(i)} Z_m^{(i)} \ln(\theta_{mk}) + \sum_{l=1}^{M} Z_l^{(i)} \ln(\pi_l) \right) \\ &= \sum_{m=1}^{M} \sum_{k=1}^{K} \left(\sum_{i=1}^{n} X_k^{(i)} Z_m^{(i)} \right) \ln(\theta_{mk}) + \sum_{l=1}^{M} \left(\sum_{i=1}^{n} Z_l^{(i)} \right) \ln(\pi_l) \\ &= \sum_{m=1}^{M} \sum_{k=1}^{K} \alpha_{mk} \ln(\theta_{mk}) + \sum_{l=1}^{M} \beta_l \ln(\pi_l) \end{aligned} \tag{2}$$

avec

$$\alpha_{mk} = \sum_{i=1}^{n} X_k^{(i)} Z_m^{(i)}$$
$$\beta_l = \sum_{i=1}^{n} Z_l^{(i)}$$

D'après (2), la log-vraisemblance est donc strictement concave en chaque composantes de $\boldsymbol{\pi}$ et $\boldsymbol{\theta}$, par combinaison linéaire de logarithmes. D'autre part, on a les contraintes linéaires suivantes :

$$\begin{aligned} \sum_{m=1}^{M} \pi_m - 1 = 0 \\ \sum_{k=1}^{K} \theta_{mk} - 1 = 0 \,, \forall m : 1..M \end{aligned} \tag{3}$$

On peut donc écrire le Lagrangien, en utilisant les multiplicateurs de Lagrange correspondants, associé au problème de maximisation de la log-vraisemblance (2) :

$$\mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\theta}, \lambda, \boldsymbol{\gamma}) = \sum_{m=1}^{M} \sum_{k=1}^{K} \alpha_{mk} \ln(\theta_{mk}) + \sum_{l=1}^{M} \beta_l \ln(\pi_l) - \lambda \left(\sum_{m=1}^{K} \pi_m - 1 \right) - \boldsymbol{\gamma}^{\mathsf{T}} \begin{pmatrix} \vdots \\ \sum_{k=1}^{K} \theta_{mk} - 1 \\ \vdots \end{pmatrix} \tag{4}$$

Le problème admet une solution unique que l'on trouve par dérivation de (4)

a Par rapport à π_m

$$\frac{\partial \mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\theta}, \lambda, \boldsymbol{\gamma})}{\partial \pi_m} = \frac{\beta_m}{\pi_m} - \lambda = 0 \implies \pi_m \propto \beta_m \implies \pi_m = \frac{\beta_m}{\sum_{m=1}^{M} \beta_m} = \frac{\sum_{i=1}^{n} Z_m^{(i)}}{\sum_{m=1}^{M} \sum_{i=1}^{n} Z_m^{(i)}}$$

D'où finalement :

$$\pi_m = \frac{1}{n} \sum_{i=1}^{n} Z_m^{(i)} \tag{5}$$

b Par rapport à θ_{mk}

$$\frac{\partial \mathcal{L}(\boldsymbol{\pi}, \boldsymbol{\theta}, \lambda, \boldsymbol{\gamma})}{\partial \theta_{mk}} = \frac{\alpha_{mk}}{\theta_{mk}} - \gamma_m = 0 \implies \theta_{mk} \propto \alpha_{mk} \implies \theta_{mk} = \frac{\alpha_{mk}}{\sum_{k=1}^{K} \alpha_{mk}} = \frac{\sum_{i=1}^{n} X_k^{(i)} Z_m^{(i)}}{\sum_{i=1}^{n} \left(\sum_{k=1}^{K} X_k^{(i)} \right) Z_m^{(i)}}$$

D'où finalement :

$$\theta_{mk} = \frac{\sum_{i=1}^{n} X_k^{(i)} Z_m^{(i)}}{\sum_{i=1}^{n} Z_m^{(i)}} \tag{6}$$