Mathias Marciano

Exercice 1 : Discrete distribution

1) Pour générer une réalisation de la variable aléatoire X, il suffit d'appliquer la méthode d'inversion en calculant l'inverse généralisé de X donné par :

$F^{\leftarrow}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}$ avec F la fdr de X et $u \in [0;1]$

X étant d'une loi discrète, l'inverse généralisé est donné par :

$$F^{\leftarrow}(u) = \inf\{x \in \mathbb{R} : \sum_{i=1}^{n} p_i \mathbb{1}_{\{x_i \leq x\}} \geq u\}$$

$$= \{x_k : S_{k-1} < u \leq S_k\}$$

avec $S_0 = 0$ et $S_k = (\sum_{i=1}^{k} p_i)$

Et on a alors $F^{\leftarrow}(U) \sim X$ pour $U \sim Unif([0;1])$

Exercice 2 : Gaussian mixture model

1) On identifie dans ce modèle :

$$\boxed{\theta = (\alpha_1, \dots \alpha_m, \mu_1, \dots \mu_m, \Sigma_1, \dots \Sigma_m)}$$

De plus, on a :

$$\mathcal{L}(x_1, \dots x_n ; \theta) = \prod_{i=1}^{n} f_\theta(x_i)$$

$$= \prod_{i=1}^{n} \mathbb{P}_\theta (X_i = x_i)$$

formule des probabilités totales

$$= \prod_{i=1}^{n} \sum_{j=1}^{m} \mathbb{P}(Z_i = j) \mathbb{P}(X_i = x_i | Z_i = j)$$

$$\boxed{\mathcal{L}(x_1, \dots x_n ; \theta) = \prod_{i=1}^{n} \sum_{j=1}^{m} \alpha_j \phi(x_i, \mu_j, \Sigma_j)} \quad \square$$

où $\phi(\cdot, \mu_j, \Sigma_j)$ est la densité de la loi normale multivariée $\mathcal{N}(\mu_j, \Sigma_j)$

3) Calcul de α

Mathias Marciano

3) <u>Etape (E)</u> :

$Q(\theta,\theta_t) = \mathbb{E}[\log P_\theta(z,x) \mid x,\theta_t]$

$Q(\theta,\theta_t) = \sum_{i=1}^{m} \mathbb{E}[\log P_\theta(z_i,x_i) \mid x_i,\theta_t]$

<u>De plus</u> : $\log P_\theta(z_i,x_i) = \sum_{j=1}^{m} \log(\mathbb{P}(z_i=j)) + \log(f_{\theta_j}(x_i))$

où f est la densité de X comme dans l'énoncé

$\Rightarrow \log P_\theta(z_i,x_i) = \sum_{j=1}^{m} \log(\alpha_j) + \log(f_{\theta_j}(x_i))$

$\Rightarrow Q(\theta,\theta_t) = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathbb{P}(z_i=j \mid x_i,\theta_t)[\log(\alpha_j) + \log(f_{\theta_j}(x_i))]$

$Q(\theta,\theta_t) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}(\theta_t)[\log(\alpha_j) + \log(f_{\theta_j}(x_i))]$

avec $p_{ij}(\theta_t) = \mathbb{P}(z_i=j \mid x_i,\theta_t)$

En appliquant la formule de Bayes on a :

$$p_{ij}(\theta_t) = \frac{\mathbb{P}(z_i=j)\,\phi(x_i \mid \theta_t, z_i=j)}{\sum_{k=1}^{m} \mathbb{P}(z_i=k)\,\phi(x_i \mid \theta_t, z_i=k)}$$

$$\boxed{p_{ij}(\theta_t) = \frac{\alpha_j\,\phi(x_i;\mu_j,\Sigma_j)}{\sum_{k=1}^{m} \alpha_k\,\phi(x_i;\mu_k,\Sigma_k)}}$$

avec $\phi(\cdot;\mu,\Sigma)$ densité de la loi normale multidimensionnelle

Etape (M) :

On sait que la valeur de α_j qui maximise notre log-vraisemblance est sa probabilité d'appartenir à un cluster soit :

$$\boxed{\alpha_j = \frac{\sum_{i=1}^{n} p_{ij}}{\sum_{k=1}^{m} \sum_{i=1}^{n} p_{ij}}}$$

Cela vient de la résolution du problème de maximisation :

$$\begin{cases} \max_{\alpha} \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \left(\log(\alpha_j) + \log(\phi(x_i; \mu_j, \Sigma_j)) \right) \\ \text{tel que } \sum_{j=1}^{m} \alpha_j = 1 \end{cases}$$

Calcul de μ : on résout

$$\max_{\mu} \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \left(\log(\alpha_j) + \log\left(\frac{1}{(2\pi|\Sigma|)^{1/2}} \exp\left(-\frac{1}{2}(x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j)\right)\right)\right)$$

$$= \max_{\mu} \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \left(\log(\alpha_j) - \frac{1}{2}(x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) \right) + K$$

$\to$ constante par rapport à μ

C'est une fonction concave donc on résout la condition d'optimalité (gradient nulle) :

$$\sum_{i=1}^{n} p_{ij} \left(\Sigma_j^{-1} x_i - \Sigma_j^{-1} \mu_j \right) = 0$$

Mathias
Marciano

$$\Rightarrow \sum_{i=1}^{n} p_{iy} x_i = \sum_{i=1}^{n} p_{iy} \mu_y$$

$$\Rightarrow \boxed{\mu_y = \frac{\sum_{i=1}^{n} p_{iy} x_i}{\sum_{i=1}^{n} p_{iy}}}$$

Pour la matrice de covariance Σ_y, ma démarche n'a pas abouti. Je prendrai donc le résultat du document de Telecom Paris dont le lien est dans mon notebook :

$$\boxed{\Sigma_y = \frac{\sum_{i=1}^{n} p_{iy} (x_i - \mu_y)^T (x_i - \mu_y)}{\sum_{i=1}^{n} p_{iy}}}$$

Mathias Marciano

Exercice 3 : Importance sampling

(iii) : $\max\limits_{\theta} \sum_{i=1}^{n} \widehat{w}_i^{(0)} \log\left(\sum_{j=1}^{M} \alpha_j \varphi(X_i^{(0)}; \theta_j)\right)$

qui est une approximation de $\max\limits_{\theta} \int \log\left(\sum_{i=1}^{M} \alpha_i \phi(x, \mu_i, \Sigma_i)\right) f(x) dx$
pour $f(x)$ une certaine densité
qui remplace les pondérations $\widehat{w}_i^{(0)}$

Ce problème de maximisation ressemble à celui de l'exercice 2 où il ~~fallait~~ fallait maximiser la log-vraisemblance avec un modèle de mélange gaussien donc cela suggère qu'on peut résoudre ce problème grâce à l'algorithme EM en faisant quelques analogies et en prenant en compte les pondérations d'importance $\widehat{w}_i^{(0)}$:

$$p_{ic} = \frac{\alpha_c \, \phi(x_i | \mu_c, \Sigma_c)}{\sum_{k=1}^{M} \alpha_k \phi(x_i | \mu_k, \Sigma_k)}$$ ϕ densité d'une gaussienne multi-dimensionnelle

$$\alpha_c = \frac{\sum_{i=1}^{n} p_{ic} \widehat{w}_i}{\sum_{k=1}^{M} \sum_{i=1}^{n} p_{ik} \widehat{w}_i}$$

$$\mu_c = \frac{\sum_{i=1}^{n} p_{ic} \widehat{w}_i x_i}{\sum_{i=1}^{n} p_{ic} \widehat{w}_i}$$

$$\Sigma_c = \frac{\sum_{i=1}^{n} p_{ic} \widehat{w}_i (x_i - \mu_c)^T (x_i - \mu_c)}{\sum_{i=1}^{n} p_{ic} \widehat{w}_i}$$