1 Implementation - HMM

Exercice 1.2

Tout d'abord, notons π la distribution initiale de q_0, $A = (p(q_t = i, q_{t+1} = j))_{i,j}$ et $\theta = \left((\pi_j)_j, (\mu_j)_j, (\Sigma_j)_{j=1}^K, A\right)$ l'ensemble des paramètres du modèle. Commençons par écrire le "complete log-likelihood" :

$$\begin{aligned} l_c = \ln(p_\theta(x,z)) &= \ln\left[\prod_{t=1}^{T} p_\theta(u_t, q_t)\right] \\ &= \sum_{j=1}^{K} \mathbb{I}(q_t = j)\ln \pi_j + \sum_{t=0}^{T-1}\sum_{i,j=1}^{K} \mathbb{I}(q_{t+1} = j, q_t = i)\ln(A_{i,j}) + \sum_{t=0}^{T}\sum_{i=1}^{K} \mathbb{I}(q_t = i)\ln\left(\mathcal{N}(u_t|\mu_j, \Sigma_j)\right) \end{aligned}$$

Afin de déterminer l'E-step, nous passons à l'espérance conditionnelle dans l'expression précédente :

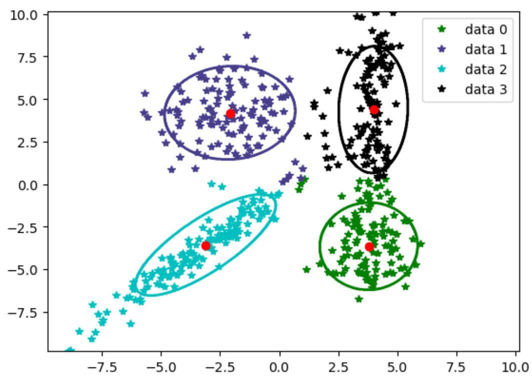
$$\mathbb{E}_{Q|U,\theta}\left[\log(p_\theta(u,q))\right] = \sum_{j=1}^{K} p(q_0 = i|u,\theta)\ln \pi_j + \sum_{t=0}^{T-1}\sum_{i,j=1}^{K} p(q_{t+1} = j, q_t = i|u,\theta)\ln(A_{i,j}) + \sum_{t=0}^{T}\sum_{j=1}^{K} p(q_t = j|u,\theta)\ln\left(\mathcal{N}(u_i|\mu_j, \Sigma_j)\right)$$

Désormais, attaquons-nous à l'M-step. Elle consiste à maximiser la quantité précédente en $\theta = ((\pi_j)_j, (\mu_j)_j, (\Sigma_j)_j, A)$. Comme nous souhaitons maximiser des quantités indépendantes, on peut majorer chacun des termes indépendamment. En reprenant la démarche du TP précédent, ainsi que la convention $0/0 = 0$, on trouve :

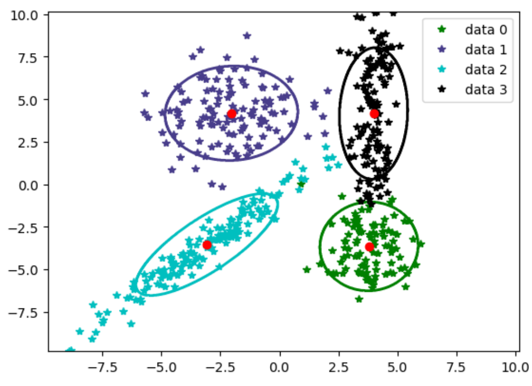
$$\pi'_j = \frac{p(q_0 = j|u,\theta)}{\sum_{i=1}^{K} p(q_0 = i|u,\theta)}; \quad \mu'_j = \frac{\sum_{t=0}^{T} p(q_t = j|u,\theta)u_t}{\sum_{t=0}^{T} p(q_t = j|u,\theta)}; \quad \Sigma'_j = \frac{\sum_{t=0}^{T} p(q_t = j|u,\theta)\|u_t - \mu_{j,t}\|^2}{\sum_{t=0}^{T} p(q_t = j|u,\theta)}$$

Intéressons-nous à la matrice de transition A. Comme la somme des éléments de chaque ligne est égale à 1, nous étudions le lagrangien $\mathcal{L}(A,\lambda) := \sum_{i,j=1}^{K} p(q_{t+1} = j, q_t = i|u,\theta)\ln(A_{i,j}) - \lambda^T(A\mathbb{1} - \mathbb{1})$. Comme $A \longmapsto \mathcal{L}(A,\lambda)$ est concave, il suffit de déterminer un point d'annulation du gradient. Cela se produit lorsque $\lambda_i = \sum_t \sum_{j=1}^{K} p(q_{t+1} = j, q_t = i|u,\theta)$. En utilisant la contrainte $A\mathbb{1} = \mathbb{1}$, on obtient la valeur de λ_i. En définitif, $A'_{i,j} = \frac{\sum_t p(q_{t+1}=j, q_t=i|u,\theta)}{\sum_t \sum_{j'=1}^{K} p(q_{t+1}=j', q_t=i|u,\theta)}$.

Exercice 1.4



(a) EM - GMM



(b) EM - HMM

Exercice 1.5

Log-likelihood	EM - GMM	EM - HMM
Unormalized Train	-3022	-2155
Normalized Train	-6.04	-4.31
Unormalized Test	-3077	-2202
Normalized Test	-6.15	-4.41

Nous constatons que l'entropie du modèle GMM est plus petite que celle du modèle HMM. En effet, contrairement au modèle GMM, le modèle HMM introduit une dépendance markovienne entre les variables cachées q_t. Cela lui permet de mieux s'adapter aux observations, donc de produire un modèle plus adapté aux u_t. Par conséquent, il est cohérent que le likehood de HMM soit plus grand que celui de GMM.