1 Implémentation - HMM

Malgré nos efforts pour raccourcir la rédaction des calculs/explications, nous n'avons pas réussi à tout faire tenir sur une page. Une version courte est disponible dans l'archive du code mais est lacunaire.

Question 1.2

Notons :

$$\boxed{\begin{aligned}&(\pi^0)_i = p(q_1 = i)\ (i \in [1,K])\\&A_{ij} = p(q_{t+1} = j | q_t = i),\ (i,j \in [1,K])\\&u_t | q_t = i \sim \mathcal{N}(\mu_i, \Sigma_i)\\&\Theta = \{\pi^0, A, (\mu_i)_{i\in K}, (\Sigma_i)_{i\in K}\}\end{aligned}}$$

La log-vraisemblance complète du modèle peut se décomposer comme suit :

$$\begin{aligned}l(q,\Theta) =& \log\left[p(q_1)\prod_{t=1}^{T-1} p(q_{t+1}|q_t)\prod_{t=1}^{T} p(u_t|q_t)\right]\\ =&\sum_{i=1}^{K}\delta(q_1 = i)\log(\pi^0)_i\\ &+\sum_{t=1}^{T-1}\sum_{i=1}^{K}\sum_{j=1}^{K}\delta(q_{t+1}=j, q_t = i)\log A_{ij}\\ &+\sum_{t=1}^{T}\sum_{i=1}^{K}\delta(q_t = i)\log\mathcal{N}(u_t|\mu_i,\Sigma_i)\end{aligned}$$

A l'itération $n+1$ de l'algorithme EM, on maximise la quantité $E_{q\sim p(.|u,\Theta^n)}[l(Z,\Theta)]$.

Posons : $\forall t \in [1,T],\ \ \forall i \in [1,K]$:

$$\boxed{p^n_{t,i} = E_{q\sim p(.|u,\Theta^n)}[\delta(q_t = i)] = p(q_t = i|u,\theta^n)}$$

et $\forall t \in [1,T-1]\ \ \forall i,j \in [1,K]$:

$$\boxed{\begin{aligned}p^n_{t,i,j} =& E_{q\sim p(.|u,\Theta^n)}[\delta(q_{t+1}=j, q_t = i)]\\ =& p(q_{t+1}=j, q_t = i|u,\theta^n)\end{aligned}}$$

On cherche donc à maximiser : $E_{q\sim p(.|u,\Theta^n)}[l(q,\Theta)] = l_1(\pi^0) + l_2(A) + l_3(\mu,\mathbf{\Sigma})$ où :

$$\begin{aligned}l_1(\pi^0) =& \sum_{i=1}^{K} p^n_{1,i}\log(\pi^0)_i\\ l_2(A) =& \sum_{t=1}^{T-1}\sum_{i=1}^{K}\sum_{j=1}^{K} p^n_{t,i,j}\log A_{ij}\\ l_3(\mu,\mathbf{\Sigma}) =& \sum_{t=1}^{T}\sum_{i=1}^{K} p^n_{t,i}\log\mathcal{N}(u_t|\mu_i,\Sigma_i)\end{aligned}$$

Pour écrire les équations de mise à jour de l'algorithme EM pour le modèle GMM, nous avions déjà maximisé par rapport à x une expression du type $\sum_i a_i \log x_i$ sous la contrainte $\sum_i x_i = 1$, ainsi que par rapport aux (μ_i, Σ_i) une expression de la forme $\sum_t\sum_i b_{it}\log\mathcal{N}(u_t|\mu_i,\Sigma_i)$. En reprenant les résultats du DM2, nous obtenons pour maximiseurs de l_1 et l_3 :

$$\boxed{\begin{aligned}(\pi^0)_i^{n+1} &= \frac{p^n_{1,i}}{\sum_j p^n_{1,j}} = p^n_{1,i}\\ \mu_i^{n+1} &= \frac{\sum_{t=1}^{T} p^n_{t,i} u_i}{\sum_{t=1}^{T} p^n_{t,i}}\\ {\Sigma_i}^{n+1} &= \frac{\sum_{t=1}^{T} p^n_{t,i}(u_i - \mu_i^{n+1})(u_i - \mu_i^{n+1})^T}{\sum_{t=1}^{T} p^n_{t,i}}\end{aligned}}$$

Maximisons enfin l_2 sous les contraintes : $\forall i,\ \ \sum_{j=1}^{K} A_{ij} = 1$. l_2 étant concave et les contraintes étant affines, un point annulateur du Lagrangien est un maximiseur. Le Lagrangien s'écrit :

$$\begin{aligned}\mathcal{L}(A,\lambda) =& \sum_{t=1}^{T-1}\sum_{i=1}^{K}\sum_{j=1}^{K} p^n_{t,i,j}\log A_{ij}\\ &+\sum_{i=1}^{K}\lambda_i\left(\sum_{j=1}^{K} A_{ij} - 1\right)\end{aligned}$$

$$\frac{\partial\mathcal{L}}{\partial A_{ij}}(A^{n+1},\lambda) = 0 \implies \frac{\sum_{t=1}^{T-1} p^n_{t,i,j}}{A_{ij}^{n+1}} - \lambda_j = 0$$

Multipliant par A_{ij}^{n+1}, sommant les égalités sur j et utilisant la contrainte $\sum_{j=1}^{K} A_{ij}^{n+1} = 1$, on trouve : $\lambda_j = \sum_{j=1}^{K}\sum_{t=1}^{T-1} p^n_{t,i,j} = \sum_{t=1}^{T-1} p^n_{t,i}$ et donc :

$$\boxed{A_{ij}^{n+1} = \frac{\sum_{t=1}^{T-1} p^n_{t,i,j}}{\sum_{t=1}^{T-1} p^n_{t,i}}}$$

Question 1.4

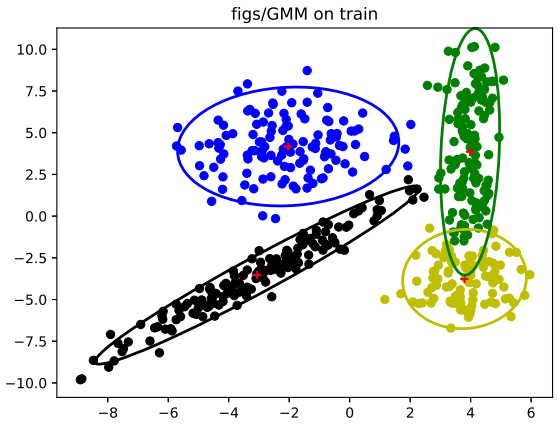

FIGURE 1 – Clusters prédits par le modèle GMM (général, ie. non isotropique) sur le jeu de Train. Initialisation avec KMeans. Centres des clusters en rouge. Chaque ellipse contient 90% de la masse de la distribution gaussienne correspondante.

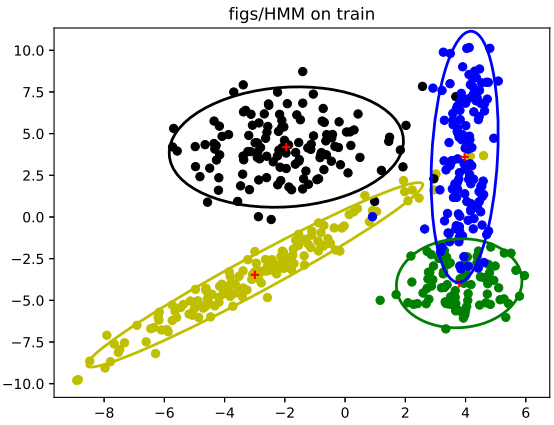

FIGURE 2 – Clusters prédits par le modèle HMM, sur le jeu de Train. Initialisation des moyennes et matrices de covariances des Gaussiennes avec GMM. Centres des clusters en rouge. Chaque ellipse contient 90% de la masse de la distribution gaussienne correspondante.

Question 1.5

Dans le tableau suivant, on donne, pour chacun des deux modèles et sur chacun des jeux, les log-vraisemblances finales "moyennes" (ie. divisée par le nombre de données du jeu, ce afin de pouvoir comparer les résultats sur jeux de Train et de Test).

Modèle | Jeu de données	Train	Test
GMM (général)	-4.66	-4.82
HMM	-3.78	-3.92

La log-vraisemblance pour le modèle GMM a été calculée comme $\sum_{t=1}^{T} \log\left(\sum_{i=1}^{K} \pi_i \mathcal{N}(u_t|\mu_i, \Sigma_i)\right)$ et celle pour le modèle HMM comme $\log p(u_1, .., u_T|\Theta) = \log\left(\sum_{i=1}^{K} p(z_T = i, u_1, .., u_T|\Theta)\right) = \log\left(\sum_{i=1}^{K} \alpha_T(z_T = i)\right)$ (et utilisant les log des α_T-messages pour calculer plus précisément la somme, cf. code).

1) Comme attendu, on observe que, pour chacun des deux modèles, la log-vraisemblance obtenue en Test est plus petite que celle en Train.

2) On observe aussi que les log-vraisemblances en Train et en Test obtenues avec le modèle HMM sont plus grandes que celles obtenues avec le modèle GMM.

Si l'on regarde les labels des données de Train dans l'ordre de leur génération (prédits par le modèle GMM ou HMM), on peut en effet observer une relation temporelle entre deux points générés consécutivement : un point dans un des deux clusters du bas sera souvent suivi par un point dans ce meme cluster, quand un point dans un des deux clusters du haut sera souvent suivi par un point dans l'autre cluster du haut. Le fait que les points des clusters du bas soient générés "en paquets" et ceux des clusters du haut "en alternance" est "expliquée" par la matrice de transition entre états cachées A estimée par le modèle HMM et donnée ci-dessous (probabilités arrondies au centième) :

$$\begin{bmatrix} 0.91 & 0.00 & 0.02 & 0.07 \\ 0.06 & 0.02 & 0.05 & 0.87 \\ 0.03 & 0.04 & 0.88 & 0.05 \\ 0.03 & 0.93 & 0.01 & 0.02 \end{bmatrix}$$

(associations états/couleurs de la figure 2 : état 1 - jaune, état 2 - noir, état 3 - vert, état 4 - bleu).

Le modèle HMM, qui cherche ainsi à expliquer la structure temporelle de la séquence, aboutit à une assignation des données aux clusters qui diffère légèrement de celle de GMM (cf. graphiques), et conduit à une vraisemblance plus élevée que GMM en Train.

Sa vraisemblance est aussi plus élevée le jeu de Test, la séquence de Test ayant été générée selon une structure temporelle similaire.