PGM - HW3 - Mehdi Boubnan & Amine Sadeq

Exercice 1.2: Compute the estimation equations of EM

The expectation of the complete loglikelihood is concave wrt π, μ, A and Σ^{-1}. We'll compute the derivatives :

$$l(\pi,\mu,A,\Sigma^{-1}) = \sum_{i=1}^{K}\tau_0^i log(\pi_i) + \sum_{t=0}^{T-1}\sum_{i=1}^{K}\sum_{j=1}^{K}\tau_t^{ij} log(A_{i,j}) + \sum_{t=0}^{T}\sum_{i=1}^{K}\tau_t^i[log(\frac{1}{(2\pi)^{\frac{d}{2}}}) + \frac{1}{2}log(det(\Sigma_i^{-1})) - \frac{1}{2}(u_t-\mu_i)^T\Sigma_i^{-1}(u_t-\mu_i)]$$

with $\forall i \in [1,K]$ $\forall j \in [1,K]$ $\forall t \in [0,T]$ $\tau_t^i = p(q_t = i|\bar{y}_0,...,\bar{y}_T)$ and $\tau_t^{ij} = p(q_{t+1} = i, q_t = j|\bar{y}_0,...,\bar{y}_T)$

`Estimate of` π:
Since the constraint $\sum_{j=1}^{K}\pi_j = 1$ and the E-step are $\mathcal{C}^1$, we have strong duality, we'll use the lagrangian:

$$L(\pi,\lambda) = -\sum_{i=1}^{K}\tau_0^i log(\pi_i) + \lambda(\sum_{i=1}^{K}\pi_i - 1)$$

$$\nabla_{\pi_i}L(\pi,\lambda) = -\frac{\tau_0^i}{\hat{\pi}_i} + \lambda = 0 \iff \hat{\pi}_i = \frac{\tau_0^i}{\lambda}$$

$$\sum_{i=1}^{K}\hat{\pi}_i = 1 \iff \sum_{i=1}^{K}\frac{\tau_0^i}{\lambda} = 1 \iff \lambda = \sum_{i=1}^{K}\tau_0^i = 1$$

$$\boxed{\hat{\pi}_i = \tau_0^i}$$

`Estimate of` A:
Since the contraints : $\forall j \in [1,K]$ $\sum_{i=1}^{K}A_{i,j} = 1$ and the E-step are $\mathcal{C}^1$, we have strong duality, we'll use the lagrangian :

$$L(A,\lambda) = -\sum_{t=0}^{T-1}\sum_{i=1}^{K}\sum_{j=1}^{K}\tau_t^{ij} log(A_{i,j}) + \sum_{j=1}^{K}(\lambda_j(\sum_{i=1}^{K}A_{i,j} - 1))$$

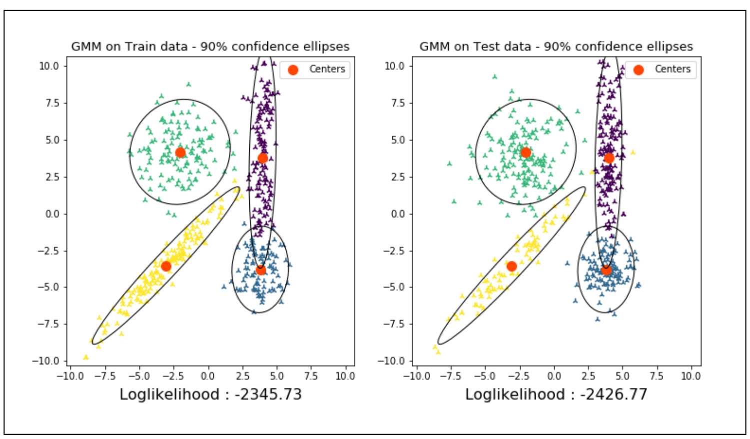
$$\nabla_{A_{i,j}}L(A,\lambda) = -\sum_{t=0}^{T-1}\frac{\tau_t^{ij}}{\hat{A}_{i,j}} + \lambda_j = 0 \iff \hat{A}_{ij} = \frac{\sum_{t=0}^{T-1}\tau_t^{ij}}{\lambda_j}$$

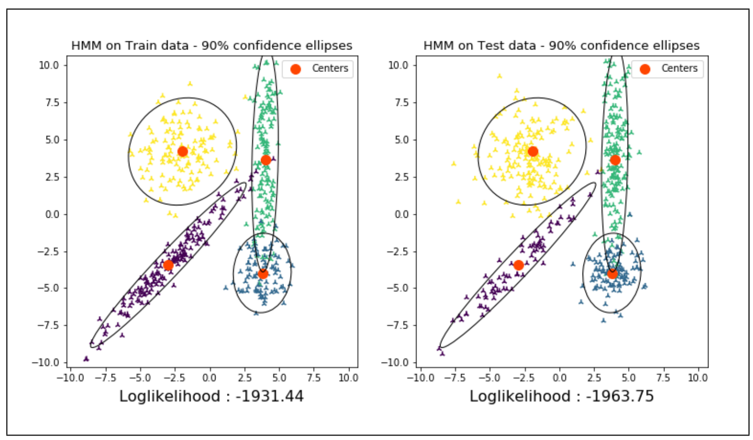
$$\sum_{i=1}^{K}\hat{A}_{i,j} = 1 \iff \sum_{i=1}^{K}\frac{\sum_{t=0}^{T-1}\tau_t^{ij}}{\lambda_j} = 1 \iff \lambda_j = \sum_{i=1}^{K}\sum_{t=0}^{T-1}\tau_t^{ij}$$

$$\boxed{\hat{A}_{i,j} = \frac{\sum_{t=0}^{T-1}\tau_t^{ij}}{\sum_{i=1}^{K}\sum_{t=0}^{T-1}\tau_t^{ij}} = \frac{\sum_{t=0}^{T-1}\tau_t^{ij}}{\sum_{t=0}^{T-1}\tau_t^{j}}}$$

`Estimate of` μ_j and Σ_j :
For μ_j and Σ_j, we get the same relations from the second homework :

$$\boxed{\hat{\mu}_j = \frac{\sum_{t=0}^{T}\tau_t^j u_t}{\sum_{t=0}^{T}\tau_t^j}} \qquad \boxed{\hat{\Sigma}_j = \frac{\sum_{t=0}^{T}\tau_t^j(u_t-\mu_j)(u_t-\mu_j)^T}{\sum_{t=0}^{T}\tau_t^j}}$$





Exercise 1.5 : Comments

We compare the loglikelihoods for each model shown in the table on the right. We notice that the train set loglikelihoods are greater than the test set ones which is expected because the estimators were learned on the training data. Moreover, there is a time dependence between our observations, the HMM algorithm performs therefore much better than the GMM (higher lokglikelihoods). Indeed, HMM is more general capturing also the pairwise dependencies between the latent variables (a shuffled dataset will give different results for the HMM).

Loglikelihoods

	GMM	HMM
Train set	-2345.75	-1931.44
Test set	-2426.77	-1963.75