PGM - HW3 - Ayman CHAOUKI & Otmane SAKHI

1 Implementation - HMM

Exercise 1.2

We want to estimate the parameters $\Theta = \{(\pi_i)_{1\leq i\leq K}\,,\ \ A = (a_{ij})_{1\leq i,j\leq K}\,,\ \ (\mu_i, \Sigma_i)_{1\leq i\leq K}\}$, the E-step consists on computing:

$$\mathcal{L}(\Theta) = \mathbb{E}\left[l_c(\theta)\right] = \sum_{i=1}^{K} p(q_1 = i|u)\, log(\pi_i) + \sum_{t=1}^{T-1}\sum_{i,j=1}^{K} p(q_{t+1} = i, q_t = j|u)\, log(a_{ij})$$
$$+ \sum_{t=1}^{T}\sum_{i=1}^{K}\left[p(q_t = i|u)\left(-\frac{1}{2}(u_t-\mu_i)^T \Sigma_i^{-1}(u_t-\mu_i) - \frac{1}{2}log|\Sigma_i| - log(2\pi)\right)\right]$$

Maximizing w.r.t: $(\pi_i)_{1\leq i\leq K}$ under the constraint $\sum_{i=1}^{K}\pi_i = 1$

Lagrangian $L(\pi,\lambda) = -\sum_{i=1}^{K} p(q_1 = 1|u)\, log(\pi_i) + \lambda\left(\sum_{i=1}^{K}\pi_i - 1\right)$

Setting $\nabla_\pi L = 0 \Rightarrow \forall i \in \{1,\ldots,K\}$, $\hat{\pi}_i = \frac{p(q_1=i|u)}{\sum_{j=1}^{K} p(q_1=j|u)} = p(q_1 = i|u)$

Maximizing w.r.t: $A = (a_{ij})_{1\leq i,j\leq K}$ under the constraint $\forall j \in \{1,\ldots,K\}$, $\sum_{i=1}^{K} a_{ij} = 1$

Lagrangian $L(\pi,\lambda) = -\sum_{t=1}^{T-1}\sum_{i,j=1}^{K} p(q_{t+1} = i, q_t = j|u)\, log(a_{ij}) + \lambda_1\left(\sum_{i=1}^{K} a_{i1} - 1\right) + \ldots \lambda_K\left(\sum_{i=1}^{K} a_{iK} - 1\right)$

Setting $\nabla_{A_{.j}} L = 0 \Rightarrow \forall i,j \in \{1,\ldots,K\}$, $\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} p(q_{t+1}=i,q_t=j|u)}{\sum_{t=1}^{T-1}\sum_{i=1}^{K} p(q_{t+1}=i,q_t=j|u)}$

Maximizing w.r.t: $\mu = (\mu_i)_{1\leq i\leq K}$

$\nabla\mathcal{L}_{\mu_i} = 0 \Rightarrow \forall i \in \{1,\ldots K\} : \hat{\mu}_i = \frac{\sum_{t=1}^{T} p(q_t=i|u)u_t}{\sum_{t=1}^{T} p(q_t=i|u)}$

Maximizing w.r.t: $\Sigma = (\Sigma_i)_{1\leq i\leq K}$, denote $\Lambda = (\Lambda_i)_{1\leq i\leq K} = \left(\Sigma_i^{-1}\right)_{1\leq i\leq K}$

$\nabla_{\Lambda_i}\mathcal{L} = 0 \Rightarrow \forall i \in \{1,\ldots,K\} : \hat{\Sigma}_i = \tilde{\Sigma}_i = \frac{\sum_{t=1}^{T} p\left(p(q_t=i|u)(u_t-\mu_i)(u_T-\mu_i)^T\right)}{\sum_{t=1}^{T} p(q_t=i|u)}$

Exercise 1.4: Graphs

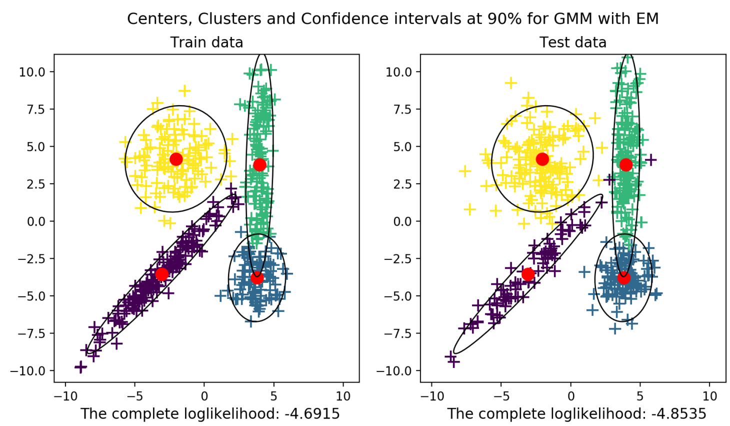

Figure 1: GMM results

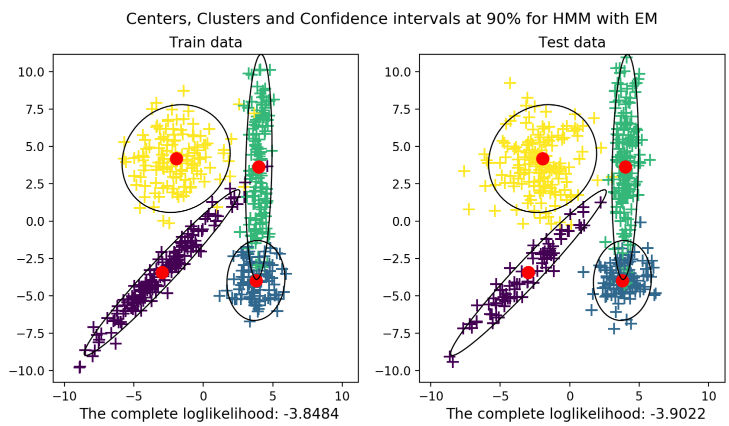

Figure 2: HMM results

Exercise 1.5 : Results comparaison

To compare the models, we used the Normalized loglikelihood as in the second homework to keep the same benchmarks. We tried at first to check if we can showcase temporal dependencies on the data to justify the use of the HMM that can be seen as a GMM that takes into account the pairwise potentials between the latent variables. So we compared an HMM that we fit on the true training data, HMM trained on a shuffled version of the data, and a GMM model. We can see that the HMM trained on a shuffled version of the dataset performs worse that a GMM and an HMM trained on the true data outperforms all the other models on both the training and the test data. It means that the use of an HMM is justified and respects the distribution assumptions of the data that is some sort of a time series.

Normalized Loglikelihoods

Model	Train	Test
GMM	-4.6915	-4.8535
HMM on shuffled	-4.6872	-4.9823
HMM	**-3.8484**	**-3.9022**