Graphical models homework 3 - Léo Alberge, Romain Petit

January 9, 2019

Exercise 1.2

Let us define :

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum\limits_{i'=1}^{K} \alpha_t(i')\beta_t(i')} = p(q_t = i|\bar{u}) \qquad \xi_t(i,j) = \frac{\alpha_t(i)a_{ij}\beta_{t+1}(j+1)f(\bar{u}_{t+1}, j+1)}{\sum\limits_{i',j'=1}^{K} \alpha_t(i)a_{ij}\beta_{t+1}(j+1)f(\bar{u}_{t+1}, j+1)} = p(q_t = i, q_{t+1} = j|\bar{u})$$

Then the estimation equations of EM write :

$$\boxed{\pi_i^k = \gamma_1(i) \qquad a_{ij}^k = \frac{\sum\limits_{t=1}^{T-1} \xi_t(i,j)}{\sum\limits_{t=1}^{T-1}\sum\limits_{j'=1}^{K} \xi_t(i,j')} \qquad \mu_i^k = \frac{\sum\limits_{t=1}^{T} \gamma_t(i)\bar{u}_t}{\sum\limits_{t=1}^{T} \gamma_t(i)} \qquad \Sigma_i^k = \frac{\sum\limits_{t=1}^{T} \gamma_t(i)(\bar{u}_t - \mu_i^k)(\bar{u}_t - \mu_i^k)^T}{\sum\limits_{t=1}^{T} \gamma_t(i)}}$$

Exercise 1.4

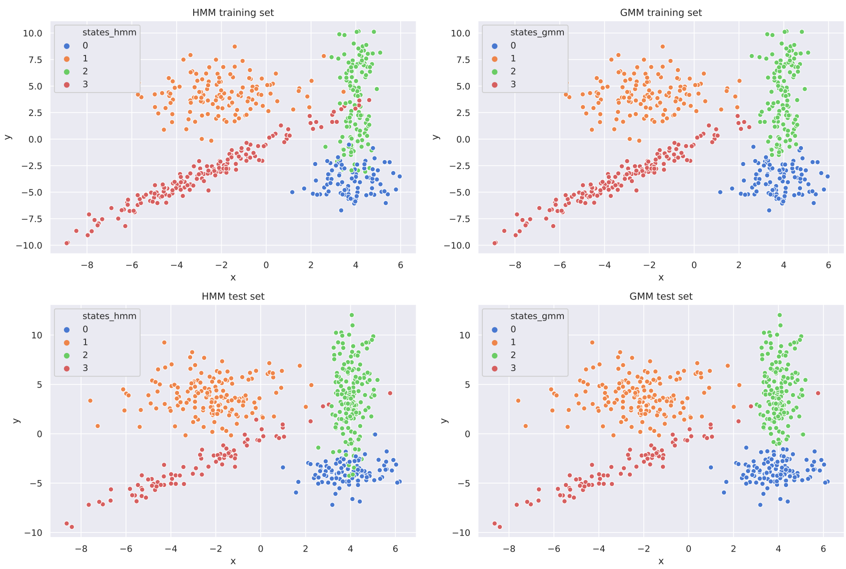

Figure 1: More likely hidden states for the hidden Markov model and the Gaussian mixture model on both datasets

Exercise 1.5

The following table shows the log likelihood of both models on the training and test sets :

Model	log likelihood (training set)	log likelihood (test set)
GMM	$-2.3\ 10^3$	$-2.4\ 10^3$
HMM	$-1.9\ 10^3$	$-2.0\ 10^3$

The following table shows the log likelihood of both models on the shuffled training and test sets :

Model	log likelihood (training set)	log likelihood (test set)
GMM	$-2.3\ 10^3$	$-2.4\ 10^3$
HMM	$-2.3\ 10^3$	$-2.6\ 10^3$

Our comments on these results can be found in the appendix.

Appendix

Exercise 1.2

The expectation step of EM consists in computing the α and β variables, that allow to compute $\phi(q_1, ..., q_T) = p(q_1, ..., q_T | \bar{u}_1, ...\bar{u}_T; \theta^{k-1})$, which maximizes $\phi \mapsto \mathbb{E}_\phi[l(\theta)]$.

The maximization step consists in maximizing the previously computed expression $\mathbb{E}_\phi[l(\theta)]$ with respect to $\theta = (\pi, a, \mu, \Sigma)$.

$$\begin{aligned}\mathbb{E}_\phi[l(\theta)] = &\sum_{i=1}^{K} p(q_1 = i|\bar{u}; \theta^{k-1}) log(\pi_i) + \sum_{t=1}^{T-1} \sum_{i,j=1}^{K} p(q_t = i, q_{t+1} = j|\bar{u}; \theta^{k-1}) log(a_{ij}) \\ &+ \sum_{t=1}^{T} \sum_{i=1}^{K} p(q_t = i|\bar{u}; \theta^{k-1}) log(f(\bar{u}_t, q_t = i))\end{aligned}$$

This shows that maximizing over θ boils down to solve the following independent problems :

$$\max_{\pi} \sum_{i=1}^{K} p(q_1 = i|\bar{u}; \theta^{k-1}) log(\pi_i) \qquad \max_{a_i} \sum_{t=1}^{T-1} \sum_{j=1}^{K} p(q_t = i, q_{t+1} = j|\bar{u}; \theta^{k-1}) log(a_{ij}) \quad i = 1, ..., K$$

$$max_{\mu_i, \Sigma_i} \sum_{t=1}^{T} p(q_t = i|\bar{u}; \theta^{k-1}) log(f(\bar{u}_t, q_t = i; \theta^{k-1})) \quad i = 1, ..., K$$

The first two optimization problems are exactly the same than the maximum likelihood estimation problem for the multinomial model. The third problem is exactly the same than the maximum likelihood estimation problem for the gaussian mixture model (solved in the previous homework).

If we define :

$$\gamma_t(i) = p(q_t = i|\bar{u}) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i'=1}^{K} \alpha_t(i')\beta_t(i')} \qquad \xi_t(i, j) = p(q_t = i, q_{t+1} = j|\bar{u}) = \frac{\alpha_t(i)a_{ij}\beta_{t+1}(j+1)f(\bar{u}_{t+1}, j+1)}{\sum_{i',j'=1}^{K} \alpha_t(i)a_{ij}\beta_{t+1}(j+1)f(\bar{u}_{t+1}, j+1)}$$

We therefore have :

$$\boxed{\pi_i^k = \gamma_1(i) \qquad a_{ij}^k = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \sum_{j'=1}^{K} \xi_t(i, j')} \qquad \mu_i^k = \frac{\sum_{t=1}^{T} \gamma_t(i)\bar{u}_t}{\sum_{t=1}^{T} \gamma_t(i)} \qquad \Sigma_i^k = \frac{\sum_{t=1}^{T} \gamma_t(i)(\bar{u}_t - \mu_i^k)(\bar{u}_t - \mu_i^k)^T}{\sum_{t=1}^{T} \gamma_t(i)}}$$

Exercise 1.5

We observe that **the hidden Markov model achieves a higher log likelihood than the Gaussian mixture model on both datasets**.

Since the hidden Markov model considers the input data as sequential, **we decided to shuffle the training and test data** to find out if the datasets had been generated to induce a special sequential structure. **In this setting, we observed that both models perform similarly** (the performance of the hidden Markov model is worse than in the original setting), the Gaussian mixture model even performing slightly better on the test set.

We were also concerned with the **influence of the hidden Markov model's parameters initialization**. In the end, we found out that **modifying the initial a and π has really little influence** since parameters after training end up being extremly close.