

Rendu DM3 Modèles probabilistes graphiques

Yoann Pradat

December 4, 2018

Exercise 1 1.1 The HMM model we are considering has chain $(q_t)_{t=1,\dots,T}$ and observations $(u_t)_{t=1,\dots,T}$. Each node of the chain can take $K = 4$ different states. We will note $\pi = (\pi_1, \dots, \pi_K)$ the distribution of the first node of the chain q_1 . In the class the formulas for the α and β recursions are given as functions of the potentials ψ . Let's define them in our model.

$$\frac{1}{Z} \psi_{q_1}(q_1) = \pi_{q_1}, \quad \psi_{q_{t-1}, q_t}(q_{t-1}, q_t) = a_{q_{t-1}, q_t}, \quad \psi_{q_t, u_t}(q_t, u_t) = \frac{(\det \Sigma_{q_t})^{-\frac{1}{2}}}{2\pi} e^{-\frac{1}{2}(u_t - \mu_{q_t})^T \Sigma_{q_t}^{-1} (u_t - \mu_{q_t})}$$

1.2 Let n be the iteration we are at in the EM algorithm. The complete log-likelihood for the data $(u_t)_{t=1,\dots,T}$ with states $(q_t)_{t=1,\dots,T}$ for the parameters $\theta_{n-1} = (\pi, a, \mu, \Sigma)$ is

$$l_c(\theta) = \sum_{k=1}^K \delta_{q_1=k} \log(\pi_k) + \sum_{t=2}^T \sum_{k,l=1}^K \delta_{q_{t-1}=k, q_t=l} \log(a_{k,l}) + \sum_{t=1}^T \sum_{k=1}^K \delta_{q_t=k} \log \mathcal{N}(u_t | \mu_k, \Sigma_k)$$

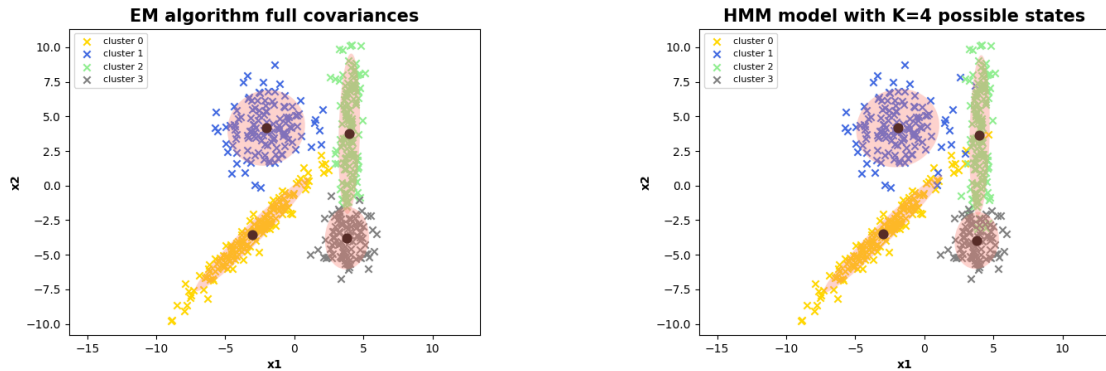
Realizing the E step is done by replace indicators δ with conditional distributions hence the following quantity to be maximized with respect to θ in the M step.

$$\mathcal{L}_\theta = \sum_{k=1}^K p_{\theta_{n-1}}(q_1=k | u_{1\dots T}) \log(\pi_k) + \sum_{t=2}^T \sum_{k,l=1}^K p(q_{t-1}=k, q_t=l | u_{1\dots T}) \log(a_{k,l}) + \sum_{t=1}^T \sum_{k=1}^K p_{\theta_{n-1}}(q_t=k | u_{1\dots T}) \log \mathcal{N}(u_t | \mu_k, \Sigma_k)$$

We deduce the following updates of the parameters θ in the M step at n^{th} iteration

$$\widehat{\pi}_k = \frac{p_{n-1}(1, k)}{\sum_{l=1}^K p_{n-1}(1, l)} \quad \widehat{a}_{k,l} = \frac{\sum_{t=2}^T p_{n-1}(t, k, l)}{\sum_{t=2}^T p_{n-1}(t, k)} \quad \widehat{\mu}_k = \frac{\sum_{t=1}^T p_{n-1}(t, k) u_t}{\sum_{t=1}^T p_{n-1}(t, k)} \quad \widehat{\Sigma}_k = \frac{\sum_{t=1}^T p_{n-1}(t, k) (u_t - \mu_k)(u_t - \mu_k)^T}{\sum_{t=1}^T p_{n-1}(t, k)}$$

where $p_{n-1}(t, k) = p_{\theta_{n-1}}(q_t=k | u_{1\dots T})$, $p_{n-1}(t, k, l) = p_{\theta_{n-1}}(q_{t-1}=k, q_t=l | u_{1\dots T})$. The calculation of conditional distributions in the formulas have been implemented in question 1.1.



1.5 We see that contrary to GMM model, HMM model may assign a point to a cluster whose center is not the closest (whatever the distance on \mathbb{R}^2). However the overall fit with HMM is better as the complete log-likelihood is lower both on the train set (-1,555.92 vs -2,345.97) and the test set (-1,636.79 vs -2,425.99). We also see that the parameters μ and Σ learned by GMM and HMM models are very similar.