Graphical models - HWK3: Implementation - HMM

Abdessamad Ed-dahmouni

December 2018

Exercise 1.2

With $\theta = (\pi_1, A, \mu_{1:K}, \Sigma_{1:K})$, the complete log-likelihood is:

$$\begin{aligned}
\ell_c(\theta) &= \log\Big(p(q_1)\prod_{t=1}^{T-1} p(q_{t+1}|q_t)\prod_{t=1}^{T} p(u_t|q_t)\Big) \\
&= \log(p(q_1)) + \sum_{t=1}^{T-1}\log p(q_{t+1}|q_t) + \sum_{t=1}^{T}\log p(u_t|q_t) \\
&= \sum_{i=1}^{K}\delta(q_1 = i)\log((\pi_1)_i) + \sum_{t=1}^{T-1}\sum_{i,j=1}^{K}\delta(q_{t+1} = i, q_t = j)\log(A_{i,j}) + \sum_{t=1}^{T}\sum_{i=1}^{K}\delta(q_t = i)\log(\mathcal{N}(u_t|\mu_i, \Sigma_i))
\end{aligned}$$

By maximizing $\mathbb{E}_{\hat{p}}(\ell_c(\theta))$ under the constraints $1^T\pi_1 = 1$ and $1^T A = 1^T$ where $\hat{p}(q_{1:T}) = \mathbb{P}(q_{1:T}|u_{1:T}, \theta)$, we update the following estimates:

$$(\hat{\pi}_1)_j = \hat{p}(q_1 = j), \qquad \hat{A}_{i,j} = \frac{\sum_{t=1}^{T-1}\hat{p}(q_{t+1} = i, q_t = j)}{\sum_{i'=1}^{K}\sum_{t=1}^{T-1}\hat{p}(q_{t+1} = i', q_t = j)}$$

$$\hat{\mu}_j = \frac{\sum_{t=1}^{T}\hat{p}(q_t = j)u_t}{\sum_{t=1}^{T}\hat{p}(q_t = j)}, \quad \hat{\Sigma}_j = \frac{\sum_{t=1}^{T}\hat{p}(q_t = j)(u_t - \mu_j)(u_t - \mu_j)^T}{\sum_{t=1}^{T}\hat{p}(q_t = j)}$$

Exercise 1.4

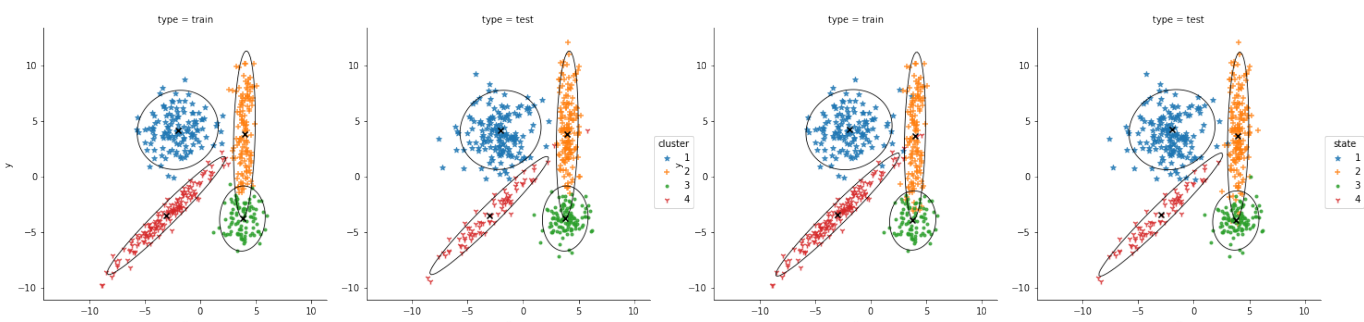

Figure 1: Most likely states for EM-GMM (left) and EM-HMM (right), with regions of 90% density.

Exercise 1.5

Method	Train	Test
GMM	-2327.72	-2408.98
HMM	-1896.97	-1954.46

Method	Train	Test
GMM	-4.6554	-4.8180
HMM	-3.7939	-3.9089

Table 1: Unnormalized and normalized log-likelihoods for GMM and HMM on training and test sets

HMM reaches a higher log-likelihood for both the training and the test set, with no sign of overfitting. While it doesn't usually make sense to compare the log-likelihood of two different models, here GMM can be viewed as an HMM where q_t are i.i.d. If this GMM assumption is true, then we shouldn't expect a net improvement in the log-likelihood, and the columns of the estimated matrix A should be very similar ($p(q_{t+1}|q_t = k) \approx p(q_{t+1}|q_t = k')$). The estimated A shows a clear dependency between q_{t+1} and q_t (for each state, there is a different state to transition to with probability ≥ 0.87) and this is confirmed by the higher likelihood. To further show the importance of this hidden Markov structure to our model, I shuffled the data and trained GMM and HMM on it, this was almost enough to break the Markov structure and render q_t independent, which was confirmed by the very close values of log-likelihoods for the two models.