MVA-DM3 PROBABILISTIC GRAPHICAL MODELS

HUGO CISNEROS

1. Implementation - HMM

1.2 For a set of parameters θ in the Gaussian Mixture Hidden Markov model (GM-HMM), the complete likelihood can be written:

$$\begin{aligned} l_c(\theta) &= \log\left(p(q_0)\prod_{t=0}^{T-1} p(q_{t+1}|q_t)\prod_{t=0}^{T} p(\overline{y}_t|q_t)\right) \\ &= \sum_{i=1}^{K}\delta(q_0=i)\log(\pi_i) + \sum_{t=0}^{T-1}\sum_{i,j=1}^{K}\delta(q_{t+1}=i,q_t=j)\log(A_{i,j}) + \sum_{t=0}^{T}\sum_{i=1}^{K}\delta(q_t=i)\log\mathcal{N}(\overline{y}_t|\mu_t,\Sigma_t) \end{aligned}$$

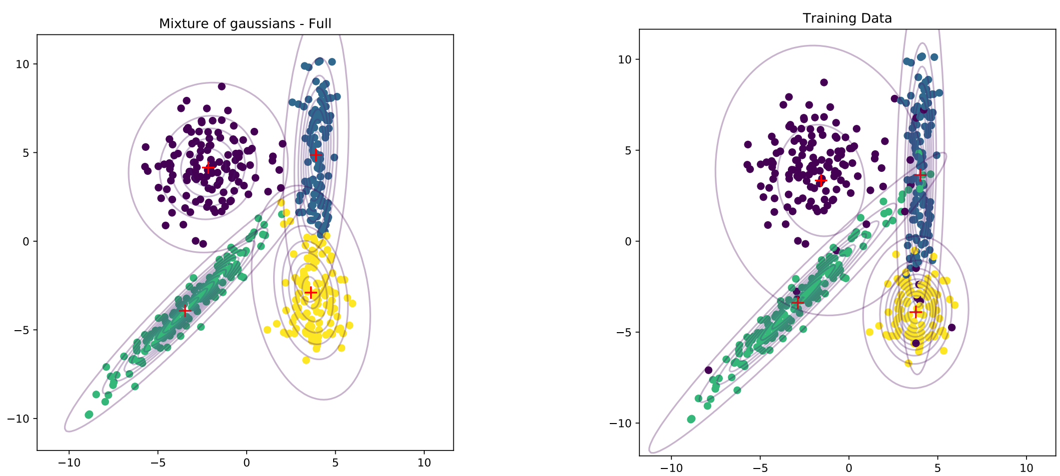
For the EM algorithm, we wish to compute the parameters of the HMM with the observations $(\overline{y}_t)$. A lower bound of the likelihood is found by using Jensen's inequality with the log function, which gives $\log p(\overline{y}_0,...,\overline{y}_T) \geq \mathbb{E}_q[l_c(\theta)]$.

We use the notations $\gamma(q_t)_i = p(q_t = i|\overline{y},\theta)$ and $\xi(q_{t+1},q_t)_{ij} = p(q_{t+1} = i, q_t = j|\overline{y},\theta)$. The expression $\mathbb{E}_q[l_c(\theta)]$ becomes

$$\mathbb{E}_q[l_c(\theta)] = \sum_{i=1}^{K}\gamma(q_0)_i\log(\pi_i) + \sum_{t=0}^{T-1}\sum_{i,j=1}^{K}\xi(q_{t+1},q_t)_{ij}\log(A_{i,j}) + \sum_{t=0}^{T}\sum_{i=1}^{K}\gamma(q_t)_i\log\mathcal{N}(\overline{y}_t|\mu_t,\Sigma_t)$$

By maximizing w.r.t π and A we get $\boxed{\pi_i = \frac{\gamma(q_0)_i}{\sum_{j=1}^{K}\gamma(q_0)_j}}$ and $\boxed{A_{i,j} = \frac{\sum_{t=0}^{T-1}\xi(q_{t+1},q_t)_{ij}}{\sum_{i=1}^{K}\sum_{t=0}^{T-1}\xi(q_{t+1},q_t)_{ij}}}$. The other derivations are similar to the ones from the Gaussian Mixture model in Homework 2.

We have : $\boxed{\mu_i = \frac{\sum_{t=0}^{T}\gamma(q_t)_i\overline{y}_t}{\sum_{t=0}^{T}\gamma(q_t)_i}}$ and $\boxed{\Sigma_i = \frac{\sum_{t=0}^{T}\gamma(q_t)_i(\overline{y}_t-\mu_i)(\overline{y}_t-\mu_i)^T}{\sum_{t=0}^{T}\gamma(q_t)_i}}$ weighted sums of the empirical means and covariances.



(A) Gaussian mixture

(B) GM-Hidden markov model

Figure 1. Comparison of mixture of gaussian profile and cluster assignment (most probable sequence of hidden state for HMM) for the Gaussian mixture model and GM-HMM.

1.5 The GM-HMM learns a slightly different mixture configuration than the GM model. The assumption that data is temporally distributed is strong and yields some artifacts such as outliers assigned to a very unlikely hidden state. However, the likelihood on both training data and test data is much higher than for the Gaussian mixtrure: -2101.4 and -2201 for GM-HMM against -2340.2 for the training set and -2431 for the test set for the GM model.