GMDL, HW #2

Oren Freifeld and Ron Shapira Weber
The Department of Computer Science, Ben-Gurion University

Release Date: 12/5/2023
Submission Deadline: 8/6/2023, 23:59

Abstract

This assignment focuses on mixtures model.

Contents

1 Mixture Models **1**
1.1 EM GMM 1
1.2 Intensity-based Clustering 3

Version Log

- 1.00, 2/5/2023. Initial release.

1 Mixture Models

1.1 EM GMM

In the context of EM for GMM fitting, with GMM parameters $\theta = (\theta_1, \dots, \theta_K, \pi_1, \dots, \pi_K)$, *i.i.d.* data points $(\boldsymbol{x}_i)_{i=1}^N$ drawn from the GMM, and missing labels $(z_i)_{i=1}^N$, recall our "Q function":

$$\begin{aligned} Q(\theta, \theta^t) &= E\left(\sum\nolimits_{i=1}^N \log p(\boldsymbol{x}_i, z_i; \theta) \middle| (\boldsymbol{x}_i)_{i=1}^N; \theta^t\right) \\ &= \left(\sum\nolimits_{i=1}^N \sum\nolimits_{k=1}^K r_{i,k} \log \pi_k\right) + \left(\sum\nolimits_{i=1}^N \sum\nolimits_{k=1}^K r_{i,k} \log p(\boldsymbol{x}_i; \theta_k)\right) \end{aligned}$$

where $r_{i,k} \triangleq p(z_i = k | \boldsymbol{x}_i; \theta^t)$.

It was stated in the lecture that the E and M steps in EM GMM are as follows.

- E step:
  Given θ^t, the current estimate of $\theta = (\theta_1, \dots, \theta_K, \pi_1, \dots, \pi_K)$, compute, $\forall i \in \{1, \dots, N\}$ and $\forall k \in \{1, \dots, K\}$,

$$r_{i,k} \overset{\text{mixture}}{=} \frac{\pi_k p(\boldsymbol{x}_i | z_i = k; \theta_k^t)}{\sum_{k'=1}^K \pi_{k'} p(\boldsymbol{x}_i | z_i = k'; \theta_{k'}^t)} \overset{\text{GMM}}{=} \frac{\pi_k \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{k'=1}^K \pi_{k'} \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{\mu}_{k'}, \boldsymbol{\Sigma}_{k'})}$$

- M step:
  We optimize Q w.r.t. π (subject to the constraint $\sum_k \pi_k = 1$) and the θ_k's to obtain the following updates, $\forall k \in \{1, \dots, K\}$:

$$\pi_k = \frac{1}{N} \sum_i r_{i,k} = \frac{N_k}{N} \qquad N_k \triangleq \sum_i r_{i,k} \tag{1}$$

$$\boldsymbol{\mu}_k = \frac{\sum_{i=1}^N r_{i,k} \boldsymbol{x}_i}{N_k} \tag{2}$$

$$\boldsymbol{\Sigma}_k = \frac{\sum_{i=1}^N r_{i,k} \boldsymbol{x}_i \boldsymbol{x}_i^T}{N_k} - \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T \tag{3}$$

  Note: $\sum_{k=1}^K r_{i,k} = 1 \, \forall i \in \{1, \dots, N\}$ and $\sum_{k=1}^K \sum_{i=1}^N r_{i,k} = N$.

Problem 1 *Complete the missing details in the M step (**some parts are optional; see below**) omitted in class. In other words, show that indeed*

$$\operatorname*{arg\,max}_{(\theta_1, \dots, \theta_K, \pi_1, \dots, \pi_K)} Q(\theta, \theta^t) \tag{4}$$

subject to the constraint

$$\sum_{k=1}^K \pi_k = 1 \,, \tag{5}$$

is given by Equation 1 (this part is mandatory), Equation 2 (this part is optional), and Equation 3 (this part is optional). Reminder: this constraint should be handled using a Lagrange multiplier (and then you will see that the second constraint, that the values are nonnegative, will be satisfied as well without having to treat it explicitly). ◇

Remark 1 *In previous years, it has been pointed out by some students that they had never seen Lagrange multipliers before. It has been further suggested by them that they believed that this made Problem 1 illegitimate. While we respect their belief, we do not share it.* ◇

Problem 2 *Let $\boldsymbol{\pi}$, a K-dimensional categorical distribution, be drawn from a Dirichlet distribution prior*

$$\boldsymbol{\pi} \sim \mathrm{Dir}(\boldsymbol{\pi}; \alpha_1, \dots, \alpha_K) \,.$$

where

$$\mathrm{Dir}(\boldsymbol{\pi}; \alpha_1, \dots, \alpha_K) \propto \prod\nolimits_{k=1}^K \pi_k^{\alpha_k - 1} \tag{6}$$

and $\boldsymbol{\pi} = \begin{bmatrix} \pi_1 & \pi_2 & \dots & \pi_K \end{bmatrix}$.

Part (i) *What values of $(\alpha_k)_{k=1}^K$ will yield a uniform distribution over the space of all K-dimensional categorical distributions?*

Part (ii) *Explain the difference between a uniform $\boldsymbol{\pi}$ on the one hand, and $\boldsymbol{\pi}$ sampled from a uniform distribution.*

Part (iii) *In the context of mixture models and estimating $\boldsymbol{\pi}$, where $\boldsymbol{\pi}$ stands for the mixture weights, suggest some motivation for using a prior (as opposed to, say, using a pure likelihood-based approach) over $\boldsymbol{\pi}$.* ◇

1.2 Intensity-based Clustering

This section focuses on intensity-based clustering, using a Bayesian Gaussian Mixture model and Gibbs sampling inference, in order to segment the "landscape" image. Note that you are not asked to implement the model/inference yourself (even though it is not that hard to do); rather, you can use the provided code and merely play with some parameters.

Computer Exercise 1 *Segment the "landscape" image using a Bayesian GMM model – with a Dirichlet prior over the weights of the components and a Normal-Inverse Wishart prior over the Gaussians' parameters – applied to the RGB values of that image and display the resulting segmentations. For that aim, use the provided code and merely adapt the values of* K *and the hyper parameters in* `run_gmm.py`. *Particularly, experiment with 3 different values of* K *(*$K = 2$, $K = 3$ *and* $K = 4$*) and for each value of* K, *explore 4 different configurations of the hyper parameters. Try to get both good and bad segmentations (where bad/-good is determined by visual inspection). Speculate how the different choices affected the results.* ◊