University of Illinois Spring 2018

CS 446: Machine Learning
Homework 9

Due on Tuesday, April 3, 2018, 11:59 a.m. Central Time

1. **[16 points]** Gaussian Mixture Models & EM
Consider a Gaussian mixture model with K components ($k \in \{1, \ldots, K\}$), each having mean μ_k, variance σ_k^2, and mixture weight π_k. All these are parameters to be learned, and we subsume them in the set θ. Further, we are given a dataset $X = \{x_i\}$, where $x_i \in \mathbb{R}$. We also use $Z = \{z_i\}$ to denote the latent variables, such that $z_i = k$ implies that x_i is generated from the k^{th} Gaussian.

   (a) What is the log-likelihood of the data $\log p(X; \theta)$ according to the Gaussian Mixture Model? (use μ_k, σ_k, π_k, K, x_i, and X). Don't use any abbreviations.

   Your answer:
$$\log p(X;\theta) = \sum_i log \sum_k \pi_k \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu_k)^2}{2\sigma_k^2}}$$

   (b) For learning θ using the EM algorithm, we need the conditional distribution of the latent variables Z given the current estimate of the parameters $\theta^{(t)}$ (we will use the superscript (t) for parameter estimates at step t). What is the posterior probability $p(z_i = k|x_i; \theta^{(t)})$? To simplify, wherever possible, use $\mathcal{N}(x_i|\mu_k, \sigma_k)$ to denote a Gaussian distribution over $x_i \in \mathbb{R}$ having mean μ_k and variance σ_k^2.

   Your answer:
$$p(z_i = k|x_i;\theta^{(t)}) = \frac{\pi_k^{(t)} \mathcal{N}(x_i|\mu_k^{(t)}, \sigma_k^{(t)})}{\sum_{c \in K} \pi_c \mathcal{N}(x_i|\mu_c^{(t)}, \sigma_c^{(t)})}$$

   (c) Find $\mathbb{E}_{z_i|x_i;\theta^{(t)}}[\log p(x_i, z_i; \theta)]$. Denote $p(z_i = k|x_i; \theta^{(t)})$ as z_{ik}, and use all previous notation simplifications.

   Your answer:
$$\mathbb{E}_{z_i|x_i;\theta^{(t)}}[\log p(x_i, z_i;\theta)] = \sum_k z_{ik} \log(\pi_k \mathcal{N}(\mathbf{x}_i|\theta_k))$$

   (d) $\theta^{(t+1)}$ is obtained as the maximizer of $\sum_{i=1}^N \mathbb{E}_{z_i|x_i;\theta^{(t)}}[\log p(x_i, z_i; \theta)]$. Find $\mu_k^{(t+1)}$, $\pi_k^{(t+1)}$, and $\sigma_k^{(t+1)}$, by using your answer to the previous question.

   Your answer:
$$\mu_k^{(t+1)} = \frac{\sum_i z_{ik} x_i}{\sum_i z_{ik}}$$
$$\pi_k^{(t+1)} = \frac{\sum_i z_{ik}}{N}$$
$$\sigma_k^{2(t+1)} = \frac{\sum_i z_{ik} ||x_i - \mu_k^{(t)}||^2}{\sum_i z_{ik}}$$

(e) How are kMeans and Gaussian Mixture Model related? (There are three conditions)

Your answer:

- kMeans assumes the data is IID, whereas GMM does not make this assumption
- kMeans uses $(x-\mu_k)^2$ as it's distance metric, whereas GMM uses $\frac{(x-\mu_k)^2}{\sigma^2}$ (euclidean distance) vs (euclidean distance divided by variance)
- kMeans assigns each point the label of the closest center, whereas GMM assigns each point the label of the most probable latent variable