University of Illinois Spring 2018

CS 446: Machine Learning
Homework 9

1. **[16 points]** Gaussian Mixture Models & EM
Consider a Gaussian mixture model with K components ($k \in \{1, \dots, K\}$), 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.

$$\text{log likelihood} = \sum_{x_i \in \mathcal{D}} \log \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}) \tag{1}$$

   (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.

By using Bayes theorem,

$$p(z_i = k|x_i;\theta^{(t)}) = \frac{p(z_i = k;\theta^{(t)})p(x_i|z_i = k;\theta^{(t)})}{p(x_i;\theta^{(t)})} \tag{2}$$

$$p(z_i = k|x_i;\theta^{(t)}) = \frac{\pi_k^t \mathcal{N}(x_i|\mu_k^t, \sigma_k^t)}{\sum_{\hat{k}} \pi_{\hat{k}}^t \mathcal{N}(x_i|\mu_{\hat{k}}^t, \sigma_{\hat{k}}^t)} \tag{3}$$

   (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.

$$\mathbb{E}_{z_i|x_i;\theta^{(t)}}[\log p(x_i, z_i;\theta)] = \sum_k p(z_i = k|x_i, \theta^t)\log\, p(x_i, z_i;\theta) \tag{4}$$

The expectation is over the distribution of z_i conditioned on x_i, computed with paramters θ set to θ^t. Bayes theorem gives us:

$$p(x_i, z_i;\theta) = p(z_i;\theta)p(x_i|z_i, \theta) \tag{5}$$

On substituting the values,

$$\mathbb{E}_{z_i|x_i;\theta^{(t)}}[\log p(x_i, z_i;\theta)] = \sum_k z_{ik}\log\, \pi_k^t \mathcal{N}(x_i|\mu_k^t, \sigma_k^t) \tag{6}$$

   (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.

Solve optimization from part C with the constraint $\sum_k \pi_k = 1$ using Lagrangian multipliers. This gives:

$$\mu_k^{t+1} = \frac{\sum_i z_{ik} x_i}{\sum z_{ik}} \tag{7}$$

$$(\sigma_k^{t+1})^2 = \frac{\sum_i z_{ik} (x_i - \mu_k^{t+1})^2}{\sum z_{ik}} \tag{8}$$

$$\mu_k^{t+1} = \frac{\sum_i z_{ik}}{N} \tag{9}$$

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

1) Same variance for all Gaussian mixtures.
2) Change to uniform distribution for the latent variable. $\pi_k = 1 \ \ \forall k$
3) Zero-temperature (hard-version) of the Gaussian Mixture Model.