University of Illinois Spring 2018

CS 446: Machine Learning
Homework 9

Due on Tuesday, April 10, 2018, 11:59 a.m. Central Time

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

   Your answer:
$$\log p(x;\theta) = \log \prod_{i=1}^{N} p(x_i;\theta) = \sum_{i=1}^{N} \log p(x_i;\theta)$$
$$= \sum_{i=1}^{N} \log \left(\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i|\mu_k, \sigma_k) \right)$$
   where,
$$\mathcal{N}(x_i|\mu_k, \sigma_k) = \left(\frac{1}{2\pi\sigma_k^2} \right)^{1/2} \exp\{-\frac{(x_i-\mu_k)^2}{2\sigma_k^2}\}$$

   (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: According bayes theorem:
$$p(A_i|B) = \frac{p(B|A_i)p(A_i)}{\sum_j p(B|A_j)p(A_j)}$$
   Here our prior is $p(A_i) : p(z_i = k) = \pi_k$ and
   conditional is $p(B|A_i) : p(x_i|\theta^{(t)}) = \mathcal{N}(x_i|\mu_k^{(t)}, \sigma_k^{(t)})$.
   Hence, our posterior will be
$$p(z_i = k|x_i; \theta^{(t)}) = \frac{\pi_k^{(t)} \mathcal{N}(x_i|\mu_k^{(t)}, \sigma_k^{(t)})}{\sum_{k=1}^{K} \pi_k^{(t)} \mathcal{N}(x_i|\mu_k^{(t)}, \sigma_k^{(t)})}$$

   (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:

$$
\begin{aligned}
\mathbb{E}[\log p(x_i, z_i; \theta)] &= \mathbb{E}\left[\log\left(\prod_{k=1}^{K}(\pi_k p(x_i;\theta_k))^{\mathbb{1}(z_i=k)}\right)\right] \\
&= \sum_{k=1}^{K}\mathbb{E}[\mathbb{1}(z_i=k)]\log(\pi_k p(x_i;\theta_k)) \\
&= \sum_{k=1}^{K} p(z_i=k; x_i, \theta^{(t)})\log[\pi_k p(x_i;\theta_k)] \\
&= \sum_{k=1}^{K} z_{ik}\log[\pi_k] + \sum_{k=1}^{K} z_{ik}\log[p(x_i;\theta_k)]
\end{aligned}
$$

(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: Proof given below,

$$\pi_k^{(t+1)} = \hat{\pi}_k = \frac{1}{N}\sum_{i=1}^{N} z_{ik}$$

$$\mu_k^{(t+1)} = \hat{\mu}_k = \frac{\sum_{i=1}^{N} z_{ik}x_i}{\sum_{i=1}^{N} z_{ik}}$$

$$\sigma_k^{2(t+1)} = \hat{\sigma}_k^2 = \frac{\sum_{i=1}^{N} z_{ik}(x_i-\mu_k)^2}{\sum_{i=1}^{N} z_{ik}}$$

Proof: Auxiliary function:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}) \triangleq \mathbb{E}\left[\sum_{i=1}^{N} \log p(x_i, z_i; \boldsymbol{\theta})\right] = \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik} \log[\pi_k] + \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik} \log[p(\mathbf{x}_i; \boldsymbol{\theta}_k)$$

Such that:

$$\sum_{k=1}^{K} \pi_k = 1$$

Hence,

$$L(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}) : \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik} \log[\pi_k] + \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik} \log[p(\mathbf{x}_i; \boldsymbol{\theta}_k) + \lambda \left(\sum_{k=1}^{K} \pi_k - 1\right)$$

Now taking derivative of $L(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)})$ w.r.t π_k and equating it to 0,

$$\frac{\partial}{\partial \pi_k} L(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}) = 0$$

$$\frac{\partial}{\partial \pi_k} \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik} \log[\pi_k] + \frac{\partial}{\partial \pi_k} \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik} \log[p(\mathbf{x}_i; \boldsymbol{\theta}_k) + \frac{\partial}{\partial \pi_k} \lambda \left(\sum_{k=1}^{K} \pi_k - 1\right) = 0$$

$$\sum_{i=1}^{N} \frac{z_{ik}}{\pi_k} + 0 + \lambda = 0$$

$$\hat{\pi}_k = \frac{1}{-\lambda} \sum_{i=1}^{N} z_i k$$

we also know $\sum_{i=1}^{N}(\lambda \sum_{k=1}^{K} \pi_k - 1) = 0$ which gives on simplification $\lambda = -N$
Now taking derivative of $L(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)})$ w.r.t $\boldsymbol{\theta}_k$ and equating it to 0
(considering only middle part of the L i.e neglecting parts that does not have $\boldsymbol{\theta} : \mu, \sigma$)

$$\frac{\partial}{\partial \mu_k} L(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}) = 0$$

$$\frac{\partial}{\partial \mu_k} \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik} \log\left(\left(\frac{1}{2\pi\sigma_k^2}\right)^{1/2} \exp\{-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}\}\right) = 0$$

$$\sum_{i=1}^{N} z_{ik} \frac{(x_i - \mu_k)}{\sigma_k^2} = 0$$

$$\hat{\mu}_k = \frac{\sum_{i=1}^{N} z_{ik} x_i}{\sum_{i=1}^{N} z_{ik}}$$

Proof (continued):
Similarly,

$$\frac{\partial}{\partial \sigma_k} L(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}) = 0$$

$$\frac{\partial}{\partial \sigma_k} \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \log \left(\left(\frac{1}{2\pi\sigma_k^2} \right)^{1/2} \exp\{-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}\} \right) = 0$$

$$\sum_{i=1}^{N} z_{ik} \sigma_k^2 = \sum_{i=1}^{N} z_{ik} (x_i - \mu_k)^2$$

$$\hat{\sigma}_k^2 = \frac{\sum_{i=1}^{N} z_{ik} (x_i - \mu_k)^2}{\sum_{i=1}^{N} z_{ik}}$$

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

Your answer: K means can be obtained from E-M on GMM via :

- Uniform Mixture Weights i.e 1/K
- Diagonal Covariances i.e $\sigma_k^2 = cI$
- c $\downarrow$ 0