Machine Learning Fall 2015 Homework 4-Saket Vishwasrao

Written Problems

1. (a)

$$L(X, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \Theta) = \sum_{i=1}^{n} \log \left(\sum_{k=1}^{K} \theta_k \, \mathcal{N}(x_i | \boldsymbol{\mu}_k, \Sigma_k) \right) \tag{1}$$

using latent variables

$$L(X, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \Theta) = \sum_{i=1}^{n} \log \left(\sum_{k=1}^{K} \frac{q(z_i = k)}{q(z_i = k)} \theta_k \, \mathcal{N}(x_i | \boldsymbol{\mu}_k, \Sigma_k) \right) \tag{2}$$

(b)

$$L(X, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \Theta) = \sum_{i=1}^{n} \log \left(\sum_{k=1}^{K} q(z_i = k) \frac{\theta_k \, \mathcal{N}(x_i | \boldsymbol{\mu}_k, \Sigma_k)}{q(z_i = k)} \right) \tag{3}$$

Using Jenson's equality,

$$\sum_{i=1}^{n} \log \left(\sum_{k=1}^{K} q(z_i = k) \frac{\theta_k \, \mathcal{N}(x_i | \boldsymbol{\mu}_k, \Sigma_k)}{q(z_i = k)} \right) \geq \sum_{i=1}^{n} \sum_{k=1}^{K} q(z_i = k) \log \left(\frac{\theta_k \, \mathcal{N}(x_i | \boldsymbol{\mu}_k, \Sigma_k)}{q(z_i = k)} \right) \tag{4}$$

$$\implies \sum_{i=1}^{n} \log \left(\sum_{k=1}^{K} q(z_i = k) \frac{\theta_k \, \mathcal{N}(x_i | \boldsymbol{\mu}_k, \Sigma_k)}{q(z_i = k)} \right) \geq \sum_{i=1}^{n} \sum_{k=1}^{K} q(z_i = k) \log \left(\theta_k \, \mathcal{N}(x_i | \boldsymbol{\mu}_k, \Sigma_k) \right) - q(z_i = k) \log(q(z_i = k)) \tag{5}$$

(c) For finding the parameters,we have

$$\arg\max_{\Theta} \sum_{i=1}^{n} \sum_{k=1}^{K} q(z_i = k) \log \left(\theta_k \, \mathcal{N}(x_i | \boldsymbol{\mu}_k, \Sigma_k) \right) - q(z_i = k) \log(q(z_i = k)) + \lambda \left(1 - \sum_{k=1}^{K} \theta_k \right) \tag{6}$$

where λ is the langrange multiplier.
Diffrentiating w.r.t θ_k,

$$\frac{\partial}{\partial \theta_k} \sum_{i=1}^{n} \sum_{k=1}^{K} q(z_i = k) \log \left(\theta_k \, \mathcal{N}(x_i | \boldsymbol{\mu}_k, \Sigma_k) \right) - q(z_i = k) \log(q(z_i = k)) + \lambda \left(1 - \sum_{k=1}^{K} \theta_k \right) = 0 \tag{7}$$

$$\implies \sum_{i=1}^{n} \frac{q(z_i = k)}{\theta_k} - \lambda = 0 \tag{8}$$

since θ_k is independent of i

$$\implies \sum_{i=1}^{n} q(z_i = k) = \lambda \theta_k \tag{9}$$

Summing both sides over all k

$$\implies \sum_{i=1}^{n} \sum_{k=1}^{K} q(z_i = k) = \sum_{k=1}^{K} \lambda \theta_k \tag{10}$$

Now,

$$\sum_{k=1}^{K} q(z_i = k) = \sum_{k=1}^{K} \theta_k = 1 \tag{11}$$

$$\implies \lambda = N \tag{12}$$

Therefore from equation(9) and (12), we have

$$\theta_k = \frac{1}{N} \sum_{i=1}^{n} q(z_i = k) \tag{13}$$

(d) We want to maximise w.r.t to q_k for a particular i.The objective becomes for every i, Let q_k represent $q(z_i = k)$

$$\arg\max_{q_k} \mathbb{E}\left[\log\left(\frac{\theta_k \mathcal{N}(x_i|\boldsymbol{\mu}_k, \Sigma_k)}{q(z_i = k)}\right)\right] \tag{14}$$

$$\implies \arg\max_{q_k} \mathbb{E}\left[\log\theta_k + \log\left(\frac{\mathcal{N}(x_i|\boldsymbol{\mu}_k, \Sigma_k)}{q(z_i = k)}\right)\right] \tag{15}$$

$$\implies \arg\max_{q_k} \mathbb{E}\left[\log\theta_k\right] + \mathbb{E}\left[\log\left(\frac{\mathcal{N}(x_i|\boldsymbol{\mu}_k, \Sigma_k)}{q(z_i = k)}\right)\right] \tag{16}$$

$$\implies \arg\max_{q_k} \mathbb{E}\left[\log\theta_k\right] - \mathbb{E}\left[\log\left(\frac{q(z_i = k)}{\mathcal{N}(x_i|\boldsymbol{\mu}_k, \Sigma_k)}\right)\right] \tag{17}$$

Since $\log\theta_k$ is independent of q_i,

$$\implies \arg\min_{q_k} \mathbb{E}\left[\log\left(\frac{q(z_i = k)}{\mathcal{N}(x_i|\boldsymbol{\mu}_k, \Sigma_k)}\right)\right] \tag{18}$$

$$\frac{\partial}{\partial q_k} \sum_{k=1}^{K} q_k \log\left(\frac{q_k}{\mathcal{N}(x_i|\boldsymbol{\mu}_k, \Sigma_k)}\right) + \beta\left(1 - \sum_{k=1}^{K} q_k\right) = 0 \tag{19}$$

where β is the langrange multiplier.

$$\implies 1 + \log\left(\frac{q_k}{\mathcal{N}(x_i|\boldsymbol{\mu}_k, \Sigma_k)}\right) + \beta = 0 \tag{20}$$

$$\implies q_k = \mathcal{N}(x_i|\boldsymbol{\mu}_k, \Sigma_k) e^{-\beta-1} \tag{21}$$

Summing both sides over all k,

$$\implies \sum_{k=1}^{K} q_k = e^{-\beta-1} \sum_{k=1}^{K} \mathcal{N}(x_i|\boldsymbol{\mu}_k, \Sigma_k) \tag{22}$$

$\mathcal{N}(x_i|\boldsymbol{\mu}_k, \Sigma_k)$ represents the probability that x_i belongs to class k.Summing over all classes,the summation will be 1

$$\implies e^{-\beta-1} = 1 \tag{23}$$

$$\implies \beta = -1 \tag{24}$$

Using equation(20),

$$\implies 1 + \log\left(\frac{q_k}{\mathcal{N}(x_i|\boldsymbol{\mu}_k, \Sigma_k)}\right) - 1 = 0 \tag{25}$$

$$\implies \log\left(\frac{q_k}{\mathcal{N}(x_i|\boldsymbol{\mu}_k, \Sigma_k)}\right) = 0 \tag{26}$$

$$\implies q_k = \mathcal{N}(x_i|\boldsymbol{\mu}_k, \Sigma_k) \tag{27}$$

2. Project with Krati,Saurabh,Sina on Network anomaly detection. Proposal emailed seperately.

Programming Problems

I discussed gmmLL.m with my project team members.The results are in ”html” folder plotted using matlab publish mode.