CS4269/6362 Machine Learning, Spring 2016: Homework 4

Sixie Yu

Question 1:

(a) Suppose we have a dataset D, from first order condition we can have:

$$\overline{x} = \underset{\hat{x}}{argmin} \sum_{x \in D} \|x - \hat{x}\|^2 \tag{1}$$

Where $\overline{x} = \frac{1}{|D|} \sum_{x \in D} x$. We denote μ_n^m as the mean vector of the n_{th} cluster in all m clusters out of D (i.e. μ_1^1 denotes global mean). Then with (1) we have:

$$\underset{S=\{D\}}{min} \sum_{k=1}^{1} \sum_{x_i \in D} \|x_i - \mu_k\|_2^2 = \underset{S=\{D\}}{min} \sum_{x_i \in S_1} \left\|x_i - \mu_1^1\right\|_2^2 + \underset{S=\{D\}}{min} \sum_{x_i \in S_2} \left\|x_i - \mu_1^1\right\|_2^2 \tag{2}$$

$$\geq \underset{S=\{S_1,S_2\}}{min} \sum_{x_i \in S_1} \left\|x_i - \mu_1^2\right\|^2 + \underset{S=\{S_1,S_2\}}{min} \sum_{x_i \in S_2} \left\|x_i - \mu_2^2\right\|^2 \tag{3}$$

$$= \underset{S=\{S_1,S_2\}}{min} \sum_{k=1}^{2} \sum_{x_i \in S_k} \|x_i - \mu_k\|_2^2 \tag{4}$$

where μ_1^2, μ_2^2 are mean vectors of cluster S_1 and cluster S_2 respectively. From (2), (3), (4) we know when K turns from 1 to 2, γ_K is non-increasing in K. Next we assume when K turns from $n-1$ to n, γ_K is non-increasing in K ($n < |D|$). That is:

$$\underset{S=\{S_1,S_2...S_{n-1}\}}{min} \sum_{k=1}^{n-1} \sum_{x_i \in S_k} \|x_i - \mu_k\|_2^2 \geq \underset{S=\{S_1,S_2...S_n\}}{min} \sum_{k=1}^{n} \sum_{x_i \in S_k} \|x_i - \mu_k\|_2^2 \tag{5}$$

In order to complete proof of induction, we need to show when K turns from n to $n+1$, γ_K is non-increasing in K. We can substitute $n = n+1$ into (5) and prove that γ_K is non-increasing in K.

(b) The Euclidean distance from $\phi(x_i)$ to α_k is:

$$\|\phi(x_i) - \alpha_k)\|^2 = \phi(x_i)\phi(x_i) - \frac{2}{|S_k|} \sum_{x_j \in S_k} \phi(x_j)\phi(x_i) + \frac{1}{|S_k|^2} \sum_{x_j \in S_k, x_h \in S_k} \phi(x_j)\phi(x_h) \tag{6}$$

$$= k(x_i, x_i) - \frac{2}{|S_k|} \sum_{x_j \in S_k} k(x_i, x_j) + \frac{1}{|S_k|^2} \sum_{x_j \in S_k, x_h \in S_k} k(x_j, x_h) \tag{7}$$

The first step: for each data instance x_i, find its nearest cluster $C^*(x_i)$ using (7):

$$C^*(x_i) = \underset{k}{argmin} \|\phi(x_i) - \alpha_k\|^2 \tag{8}$$

The second step: update the mean vector of each cluster:

$$\alpha_k = \frac{\sum_{x_i \in S_k} \phi(x_i)}{|S_k|} \tag{9}$$

Question 2:

(a) Given cluster assignments, we want to update cluster center $\hat{x}$ with following formula:

$$\hat{x}^* = \underset{\hat{x}}{argmin} \sum_{x \in S_k} \|x - \hat{x}\|_1 \tag{10}$$

$$= \underset{\hat{x}}{argmin} \sum_{x \in S_k} \sum_{i=1}^{m} |x_i - \hat{x}_i| \tag{11}$$

$$= \underset{\hat{x}}{argmin} \sum_{i=1}^{m} \sum_{x \in S_k} |x_i - \hat{x}_i| \tag{12}$$

where m is feature dimension. From useful fact mentioned in the question and (12) we can know $\hat{x}_i^*$ is computed as median of ith dimensional features of all data in S_k.

(b) The ith dimension of a cluster representative is computed as median of ith dimensional features of all data in the cluster.

Question 3:

(a) E-step:

$$\gamma(z_k) = \frac{\pi_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \tag{13}$$

$\gamma(z_k)$ is expected value of the indicator variable z_k calculated from $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\pi}$ in ith iteration.
M-step:

$$\mathbb{E}_{\boldsymbol{Z}}[\ln p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})] = \sum_{k=1}^{K} \gamma(z_k)\{\ln \pi_k + \ln \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\} \tag{14}$$

Plug into the expected value $\gamma(z_k)$ from (13), and maximize complete data log-likelihood (14) with respect to $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\pi}$.

(b) $O(Km^2)$, where K is the dimension of $\boldsymbol{\pi}$, and m is the dimension of data.