CS 559

Assignment

Chenxi Wang

Problem 2.

(1) We have $L = \sum_{j=1}^{k} \cdot \sum_{x_i \in S_j} \| x_i - \mu_j \|^2$

Since $\|V\|_2^2 = V^T \cdot V$ for vector V.

$$\therefore \frac{\partial L}{\partial \mu_1} = \frac{\partial \sum_{x_i \in S_1} (x_i - \mu_1)^T \cdot (x_i - \mu_1)}{\partial \mu_1}$$

We have $(u \cdot v)' = u' \cdot v + u \cdot v'$

$$= \sum (-1)(x_i - \mu_1) + (x_i - \mu_1)^T \cdot (-1)$$

$$= \sum_{x_i \in S_1} 2 \cdot (\mu_1 - x_i)$$

$\therefore$

$$\mu_1 \leftarrow \mu_1 - \epsilon \cdot \sum_{x_i \in S_1} 2 \cdot (\mu_1 - x_i)$$

(2)

We have the result from (1):

$$\frac{\partial L}{\partial \mu_1} = \sum_{x_i \in S_1} 2 \cdot (\mu_1 - x_i)$$

For the stochastic gradient desent =

when $x_i \in S_1$:

$$\mu_1 \leftarrow \mu_1 - 2\varepsilon \cdot (\mu_1 - x_i)$$

when $x_i \notin S_1$:

$$\mu_1 \leftarrow \mu_1$$

(3).

For the batch gradient desent:

$$\mu_1 \leftarrow \mu_1 - \varepsilon \cdot \sum_{x_i \in S_1} 2 \cdot (\mu_1 - x_i)$$

For the standard k-means algorithm:

$$\mu_1 \leftarrow \sum_{x_i \in S_1} \frac{1}{|S_1|} \cdot x_i$$

When two algorithms perform the same updata for μ_1:

$$\mu_1 - \varepsilon \sum_{x_i \in S_1} 2 \cdot (\mu_1 - x_i) = \sum_{x_i \in S_1} \frac{1}{|S_1|} \cdot x_i$$

$$2 \cdot \varepsilon \cdot \sum_{x_i \in S_1} (x_i - \mu_1) = \sum_{x_i \in S_1} \frac{1}{|S_1|} \cdot x_i - \sum_{x_i \in S_1} \frac{1}{|S_1|} \cdot \mu_1$$

$$\varepsilon = \frac{\frac{1}{|S_1|} \cdot \sum_{x_i \in S_1} (x_i - \mu_1)}{2 \cdot \sum_{x_i \in S_1} (x_i - \mu_1)} = \frac{1}{2 \cdot |S_1|}$$

Problem 3.

(1). The compact form $p(z)$ is:

$$p(z) = \prod_{k=1}^{K} \pi_k^{z_k}$$

The compact form $p(x|z)$ is:

$$p(x|z) = \prod_{k=1}^{K} N(x|\mu_k, \Sigma_k)^{z_k}$$

(2)

$$p(x) = \sum_{z} p(x, z) = \sum_{z} p(z) \cdot p(x|z)$$

We have:

$$p(z) = \prod_{k=1}^{K} \pi_k^{z_k} = p(z_k = 1) = \pi_k$$

$$p(x|z) = \prod_{k=1}^{K} N(x|\mu_k, \Sigma_k)^{z_k} = p(x|z_k=1) = N(x|\mu_k, \Sigma_k)$$

$\because$ $z_k \in \{0, 1\}$, $\sum_k z_k = 1$

$$p(x) = \sum_{z} p(z) \cdot p(x|z) = p(z_1) \cdot p(x|z_1) + \cdots\cdots p(z_k) \cdot p(x|z_k)$$

$$= p(z_1 = 1) \cdot p(x|z_1 = 1) + \cdots + p(z_k = 1) \cdot p(x|z_k = 1)$$

$$= \sum_{k=1}^{K} p(z_k = 1) \cdot p(x|z_k = 1) = \sum_{k=1}^{K} \pi_k \cdot N(x|\mu_k, \Sigma_k)$$

(3) For finding maximum likelihood solution for GMM, we should use the expectation-maximization algorithm (EM algorithm).

The parameters: π, μ, Σ

$$X = \{x_1, x_2, \cdots, x_N\} \in R^{N \times D}$$

$$\ln p(X \mid \pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \cdot N(x_n \mid \mu_k, \Sigma_k) \right\}$$

- E step: Evaluate the responsibilities using the current parameter values:

$$\gamma_k(x) = p(k \mid x) = \frac{\pi_k N(x \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(x \mid \mu_j, \Sigma_j)}$$

- M step: Re-estimate the parameters using the current responsibilities

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma_k(x_n)\, x_n \quad , \quad \Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma_k(x_n)(x_n - \mu_k) \cdot (x_n - \mu_k)^T$$

$$\pi_k = \frac{N_k}{N}$$

- Evaluate $\ln p(X \mid \pi, \mu, \Sigma)$ check for convergence.

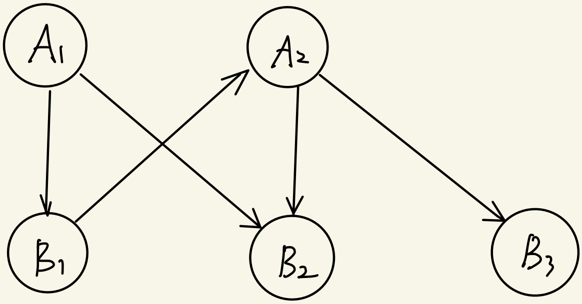
K means vs. GMM:

k-means is hard assignment, each point is associated uniquely with one cluster.

EM for GMM is soft assignment, based on the posterior probabilities.

Problem 4:

(1).



(2)

$$P(A_1, A_2, B_1, B_2, B_3) = \prod_{i=1}^{2} P(A_i \mid Pa_i) \cdot \prod_{j=1}^{3} P(B_j \mid Pa_j)$$

$$= P(A_1) \cdot P(A_2 \mid B_1) \cdot P(B_1 \mid A_1) \cdot P(B_2 \mid A_1, A_2) \cdot P(B_3 \mid A_2)$$

$$1 + 4 + 2 + 4 + 4 = 15$$

(3) There are 15 parameters are needed to fully specify the joint distribution.

(4) $P(A_1, A_2, B_1, B_2, B_3)$

$$= P(A_1) \cdot P(A_2 \mid A_1) \cdot P(B_1 \mid A_1, A_2) \cdot P(B_2 \mid A_1, A_2, B_1) \cdot P(B_3 \mid A_1 A_2 \cdot B_1 B_2)$$

$$1 \qquad 2 \qquad 4 \qquad 8 \qquad 16$$

$$= 31 = 2^5 - 1$$