K-Means Clustering and Gaussian Mixture Model

Il-Chul Moon
Dept. of Industrial and Systems Engineering
KAIST
icmoon@kaist.ac.kr

Weekly Objectives

- Understand the clustering task and the K-means algorithm
  - Know what the unsupervised learning is
  - Understand the K-means iterative process
  - Know the limitation of the K-means algorithm
- Understand the Gaussian mixture model
  - Know the multinomial distribution and the multivariate Gaussian distribution
  - Know why mixture models are useful
  - Understand how the parameter updates are derived from the Gaussian mixture model
- Understand the EM algorithm
  - Know the fundamentals of the EM algorithm
  - Know how to derive the EM updates of a model

Gaussian Mixture Model

- Let's assume that the data points are drawn from a mixture distribution of multiple multivariate Gaussian distributions
  - $P(x) = \sum_{k=1}^{K} P(z_k)P(x|z) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k)$
  - How to model such mixture?
    - Mixing coefficient, or Selection variable: z_k
      - The selection is stochastic which follows the multinomial distribution
      - $z_k \in \{0,1\}, \sum_k z_k = 1, P(z_k = 1) = \pi_k, \sum_{k=1}^{K} \pi_k = 1, 0 \le \pi_k \le 1$
      - $P(Z) = \prod_{k=1}^{K} {\pi_k}^{z_k}$
    - Mixture component
      - $P(X|z_k = 1) = N(x|\mu_k, \Sigma_k) \rightarrow P(X|Z) = \prod_{k=1}^{K} N(x|\mu_k, \Sigma_k)^{z_k}$
  - This is the marginalized probability. How about conditional?
    - $\gamma(z_{nk}) \equiv p(z_k = 1|x_n) = \frac{P(z_k=1)P(x|z_k = 1)}{\sum_{j=1}^{K} P(z_j=1)P(x|z_j = 1)}$

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

- Log likelihood of the entire dataset is
  - $\ln P(X|\pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln\{\sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k)\}$