Gaussian mixture models and Expectation maximization

Clustering methods

K Means Clustering

Hierarchical Clustering

Gaussian Mixture Models

Gaussian Distribution

- In one dimension probability density function of a Gaussian Distribution

$$G(X|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- where μ and σ are respectively the mean and variance of the distribution

Multivariate Gaussian distribution

$$G(X|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(X-\mu)^T \boldsymbol{\Sigma}^{-1}(X-\mu)\right)$$

Here μ is a d dimensional vector denoting the mean of the distribution and Σ is the d X d covariance matrix.

Gaussian Mixture Model

- Suppose there are K clusters
- the probability density is defined as a linear function of densities of all these K distributions

$$p(X) = \sum_{k=1}^{K} \pi_k G(X|\mu_k, \Sigma_k)$$

Gaussian Mixture Model

- $\prod_k$ is the mixing coefficient for kth distribution
- For estimating the parameters by the maximum log-likelihood method

$$\ln p(X|\mu, \Sigma, \pi) = \sum_{i=1}^{N} p(X_i)$$
$$= \sum_{i=1}^{N} \ln \sum_{k=1}^{K} \pi_k G(X_i|\mu_k, \Sigma_k)$$

Bayes theorem

$$\gamma_k(X) = \frac{p(X|k)p(k)}{\sum_{k=1}^{K} p(k)p(X|k)}$$

$$= \frac{p(X|k)\pi_k}{\sum_{k=1}^{K} \pi_k p(X|k)}$$

Gaussian Mixture Model

- *So equating the derivative of p(X|\mu,\Sigma,\pi)with respect to \mu to zero and rearranging the terms,*

$$\mu_k = \frac{\sum_{n=1}^{N} \gamma_k(x_n) x_n}{\sum_{n=1}^{N} \gamma_k(x_n)}$$

Gaussian Mixture Model

$$\Sigma_k = \frac{\sum_{n=1}^{N} \gamma_k(x_n)(x_n - \mu_k)(x_n - \mu_k)^T}{\sum n=1^N \gamma_k(x_n)}$$

$$\pi_k = \frac{1}{N} \sum_{n=1}^{N} \gamma_k(x_n)$$

denotes the total number of sample points in the kth cluster.