# Chapter 9 Mixture Models and EM

Yue Yu

## 9.1 K-means Clustering

### (9.5)

Recall the distortion function

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2.$$

Setting the derivative of $J$ with respect to $\boldsymbol{\mu}_k$ to $\mathbf{0}$, we have

$$\frac{\partial}{\partial \boldsymbol{\mu}_k} J = -2 \sum_{n=1}^{N} r_{nk} (\mathbf{x}_n - \boldsymbol{\mu}_k) = \mathbf{0},$$

which implies that

$$-\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} r_{nk} (\mathbf{x}_n - \boldsymbol{\mu}_k) = \mathbb{E}[-r_{nk}(\mathbf{x}_n - \boldsymbol{\mu}_k)] = \mathbf{0}.$$

Applying the Robbings-Monro algorithm while setting $a^{\text{old}} r_{nk} = \eta_n$, we obtain

$$\boldsymbol{\mu}_k^{\text{new}} = \boldsymbol{\mu}_k^{\text{old}} + \eta_n (\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{old}}).$$

## 9.2 Mixtures of Gaussians

### (9.12)

$$\begin{aligned}
p(\mathbf{x}) &= \sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{z}) \\
&= \sum_{\mathbf{z}} p(\mathbf{z}) p(\mathbf{x}|\mathbf{z}) \\
&= \sum_{k=1}^{K} p(z_k = 1) p(\mathbf{x}|z_k = 1) \\
&= \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k).
\end{aligned}$$

### (9.22)

As (9.21) indicates, by setting the derivative of the Lagrangian with respect to $\pi_k$ to 0, we have

$$\frac{\partial}{\partial \pi_k} L = \sum_{n=1}^{N} \frac{\mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_j \pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} + \lambda = 0.$$

Multiplying both sides by $\pi_k$ and sum over $k$, we have

$$
\begin{aligned}
0 &= \sum_{k=1}^{K} \sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_j \pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} + \sum_{k=1}^{K} \lambda \pi_k \\
&= \sum_{n=1}^{N} \frac{\sum_k \pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_j \pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} + \lambda \sum_{k=1}^{K} \pi_k \\
&= N + \lambda,
\end{aligned}
$$

which implies that

$$
\lambda = -N.
$$

Plugging it into

$$
\underbrace{\sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_j \pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}}_{N_k} + \lambda \pi_k = 0,
$$

we obtain

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