K-means clustering : 거리기반, discrete label
Mixture of Gaussian : 확률기반, probabilistic
⇒ responsibility
공통점: iterative method

Inclass 12: Mixture of Gaussian Clustering

[SCS4049] Machine Learning and Data Science

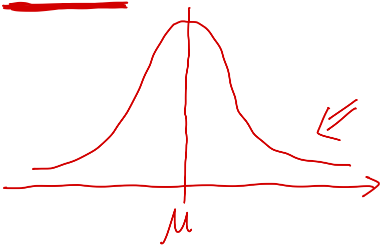
Seongsik Park (s.park@dgu.edu)

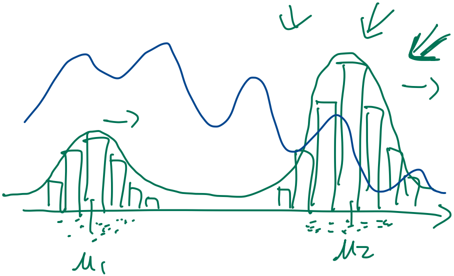
School of AI Convergence & Department of Artificial Intelligence, Dongguk University

1. mixture model, distribution
2. mixture of Gaussian → responsibility
3. iterative method

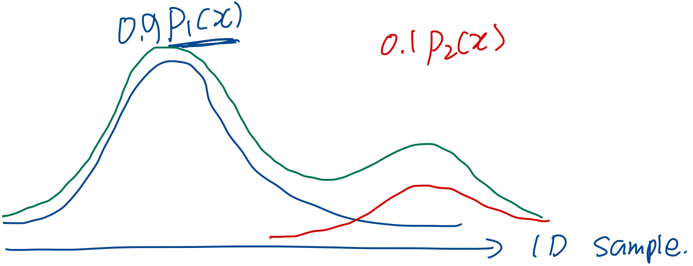
① **mixture** model : 기존확률분포 x 여러개 → 새확률분포 한다.

$$\sum_{k=1}^{K} \pi_k \, p_k(x) = p(x)$$





1D sample



$$p(x) = 0.5\,p_1(x) + 0.5\,p_2(x)$$

$$= \sum_{k=1}^{K} \pi_k \, p_k(x)$$

$$\sum \pi_k = 1$$

mixing coefficient

② mixture of Gaussian

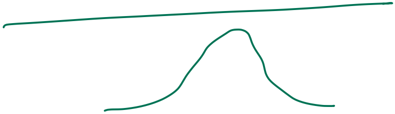
$$p(x) = \pi_1 p_1(x) + \pi_2 p_2(x) + \cdots$$

$$= \pi_1 N(\mu_1, \Sigma_1) + \pi_2 N(\mu_2, \Sigma_2) + \cdots$$

parameter

$$\{(\pi_1, \mu_1, \Sigma_1), (\pi_2, \mu_2, \Sigma_2), (\pi_3, \mu_3, \Sigma_3) \cdots \}$$

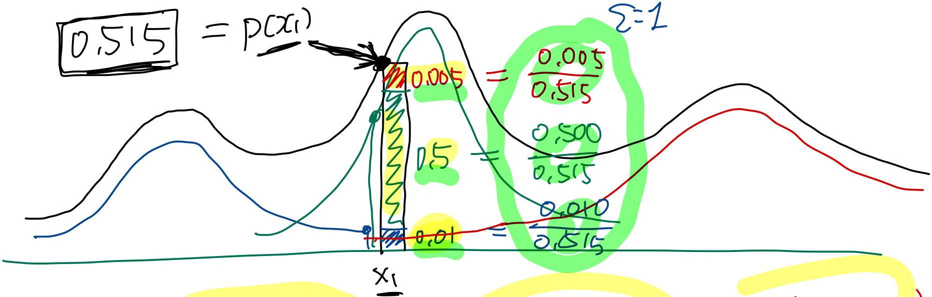
$$N(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



Responsibility ⇐ · probability로 나타낸 cluster label

γ_{nk}

· n번째 sample x_n이 k번째 cluster (π_k, μ_k, Σ_k)에 속할 확률



$$p(x_1) = \underbrace{\pi_1 N(\mu_1, \Sigma_1)}_{0.01} + \underbrace{\pi_2 N(\mu_2, \Sigma_2)}_{0.50} + \underbrace{\pi_3 N(\mu_3, \Sigma_3)}_{0.005}$$

MoG model

$$p(x_1) = 0{,}01 + 0{,}500 + 0{,}005$$

$$\gamma_{11} = \frac{10}{515} \simeq 2\% \simeq 0{,}02$$

$$\gamma_{12} = \frac{500}{515} \simeq 97\% \simeq 0{,}97$$

$$\gamma_{13} = \frac{5}{515} \simeq 1\% \simeq 0.01$$

sample $x_1, x_2, \cdots x_N$ $\longrightarrow$ responsibility

parameter (π, μ, Σ) $\longrightarrow$ γ_{nk}

K-means

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$\downarrow \quad \downarrow \quad \downarrow$

$\mu_1 \quad \mu_2 \quad \mu_3$

$$\mu_1 = \frac{1}{2}(x_1 + x_3)$$

MoG

$$\gamma = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0.2 & 0.7 & 0.1 \\ 0.8 & 0.1 & 0.1 \\ 0.0 & 0.2 & 0.8 \end{bmatrix}$$

π_k: $\frac{1.9}{4} \quad \frac{1.1}{4} \quad \frac{1.0}{4}$

$$\mu_1 = \frac{1}{1.9}(0.9x_1 + 0.2x_2 + 0.8x_3)$$

MoG: summary

EM method

1. **Initialize** the means $\boldsymbol{\mu}_k$, covariances $\boldsymbol{\Sigma}_k$ and mixing coefficients π_k.
2. **E-step** Evaluate the responsibilities using the current parameter values

expectation

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

3. **M-step** Re-estimate the parameters using the current responsibilities

maximization

sample respons.

$$\boldsymbol{\mu}_k = \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})\mathbf{x}_n \tag{2}$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})(\mathbf{x}_n-\boldsymbol{\mu}_k)(\mathbf{x}_n-\boldsymbol{\mu}_k)^T \tag{3}$$

$$\pi_k = \frac{N_k}{N} \tag{4}$$

4. Evaluate the log likelihood

$$\log p(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}) = \sum_{n=1}^{N}\log\left\{\sum_{k=1}^{K}\pi_k\mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right\} \tag{5}$$

MoG: model

We shall view π_k as the prior probability of $z_k = 1$, and the quantity $\gamma(z_k)$ as the corresponding posterior probability once we have observed $\mathbf{x}$. As we shall see later, $\gamma(z_k)$ can also be viewed as the *responsibility* that component k takes for 'explaining' the observation $\mathbf{x}$.

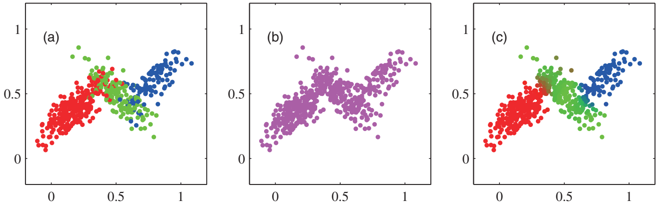


Figure 9.5 Example of 500 points drawn from the mixture of 3 Gaussians shown in Figure 2.23. (a) Samples from the joint distribution $p(\mathbf{z})p(\mathbf{x}|\mathbf{z})$ in which the three states of $\mathbf{z}$, corresponding to the three components of the mixture, are depicted in red, green, and blue, and (b) the corresponding samples from the marginal distribution $p(\mathbf{x})$, which is obtained by simply ignoring the values of $\mathbf{z}$ and just plotting the $\mathbf{x}$ values. The data set in (a) is said to be *complete*, whereas that in (b) is *incomplete*. (c) The same samples in which the colours represent the value of the responsibilities $\gamma(z_{nk})$ associated with data point $\mathbf{x}_n$, obtained by plotting the corresponding point using proportions of red, blue, and green ink given by $\gamma(z_{nk})$ for $k = 1, 2, 3$, respectively

2D sample

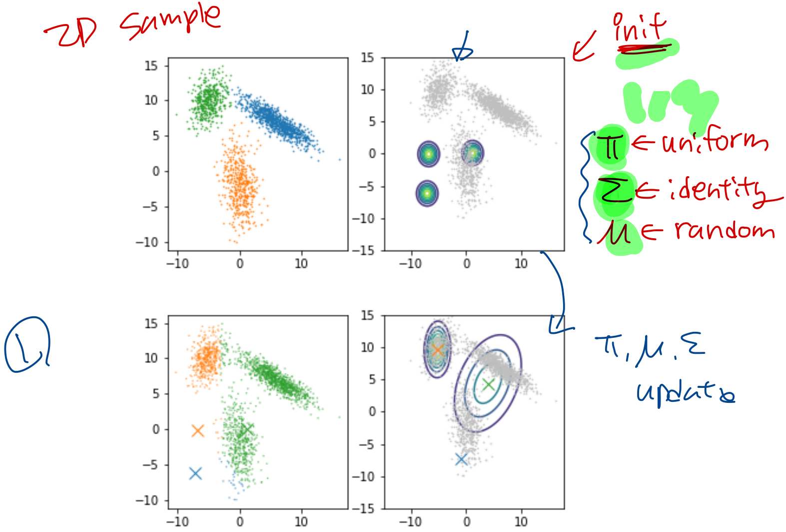


Figure 1: MoG python example: dataset and initialization (top) and the 1st iteration (bottom).

MoG: python example

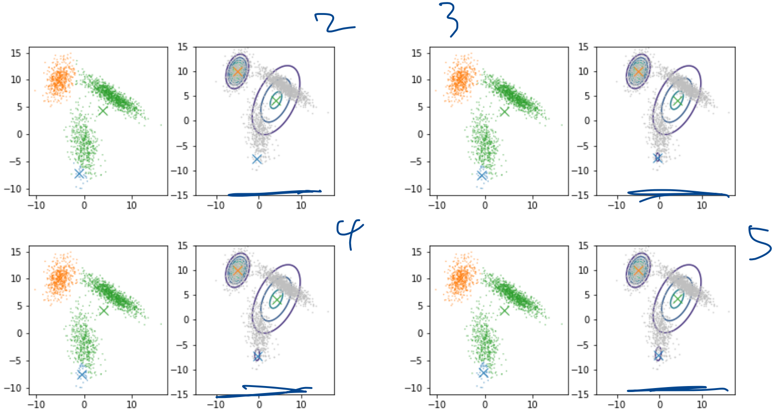


Figure 2: From the 2nd to 5th iterations.

MoG: python example

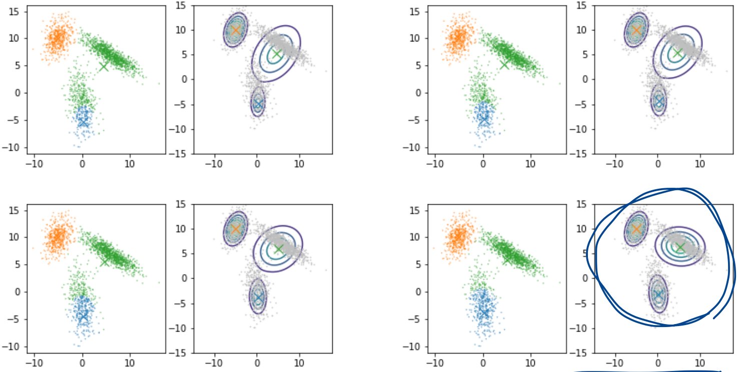
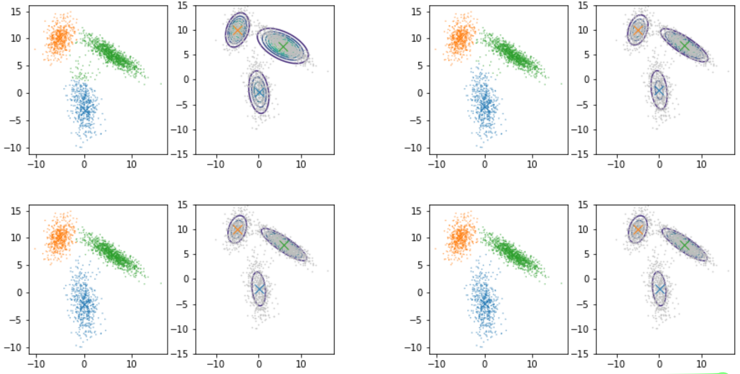


Figure 3: From the 12th to 15th iterations.

15

MoG: python example

16



19

Figure 4: From the 16th to 19th iterations.

$$p(x) = \pi_1 p_1(x) + \pi_2 p_2(x) + \pi_3 p_3(x)$$

cluster 개수 → k-means
→ MoG ?!

→ cost, performance measure

결정

model selection

Reference and further reading

- “Chap 9” of A. Geron, Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow
- “Chap 9” of C. Bishop, Pattern Recognition and Machine Learning
- Variational Bayesian mixtures of Gaussians