

kmeans
MoG

6^o iterative⁹

discrete, label $\rightarrow$ one-hot
probabilistic

Inclass 10: Mixture of Gaussian Clustering

7/12/21.
kmeans
MoG.
unsupervised

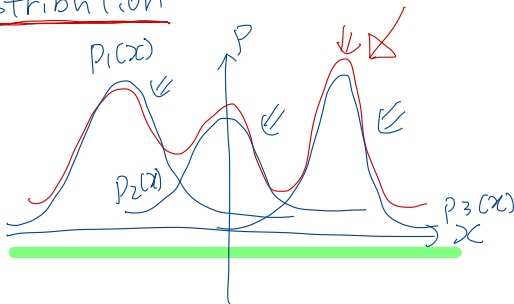
[SCS4049] Machine Learning and Data Science

Seongsik Park (s.park@dgu.ac.kr)

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

- ① mixture, mixture of Gaussian.
 - ② responsibility 
 - ③ param. update.
- 

mixture distribution



$$p(x) = 0.1 p_1(x) + 0.7 p_2(x) + 0.2 p_3(x)$$

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

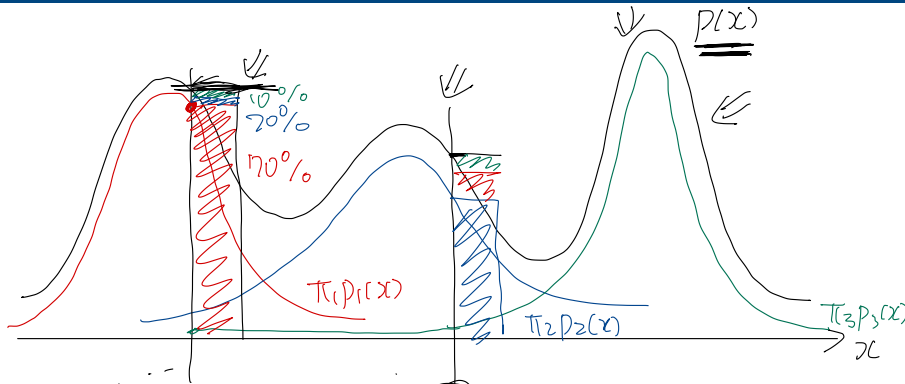
$$\sum \pi_k = 1$$

→ mixing coefficient, 0/1

mixture of Gaussian. (MoG, GMM)

$$p(x) = \sum_k \pi_k N(x; \mu_k, \sigma_k^2)$$

$$= \pi_1 N(x; \mu_1, \sigma_1^2) + \pi_2 N(x; \mu_2, \sigma_2^2) + \pi_3 N(x; \mu_3, \sigma_3^2)$$



$$\downarrow x_1$$

$$0.5$$

$$p(x_1)$$

$$= 0.35 \pi_1 p_1(x_1) + 0.10 \pi_2 p_2(x_1) + 0.05 \pi_3 p_3(x_1)$$

$$\gamma_1 = \frac{0.35}{0.5} = 0.7$$

$$\gamma_2 = \frac{0.10}{0.5} = 0.2$$

$$\gamma_3 = \frac{0.05}{0.5} = 0.1$$

γ_{nk} $=$

$$\frac{\pi_k N(x_n; \mu_k, \sigma_k^2)}{\sum_j \pi_j N(x_n; \mu_j, \sigma_j^2)}$$

x_n : n th sample

k th distribution

$$r_{nk} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \end{matrix}$$

5x3

$$\gamma_{nk} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0.3 & 0.6 & 0.1 \\ 0.1 & 0.2 & 0.7 \\ 0.9 & 0.1 & 0 \end{bmatrix} \end{matrix}$$

5x3

$$\textcircled{2} \quad \mu_1 = \frac{1}{2.3} (0.8x_1 + 0.2x_2 + 0.3x_3 + 0.1x_4 + 0.9x_5)$$

$$N_k = \sum_n \gamma_{nk}$$

$$\mu_2 = \frac{1}{1.7} (0.1x_1 + 0.7x_2 + 0.6x_3 + 0.2x_4 + 0.1x_5)$$

MoG: summary

1. **Initialize** the means μ_k , covariances Σ_k and mixing coefficients π_k .
2. **E-step** Evaluate the responsibilities using the current parameter values

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

Handwritten notes: *3. step sample x_n param*

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{x}_n | \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(\mathbf{x}_n | \mu_j, \Sigma_j)} \quad (1)$$

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) \mathbf{x}_n \quad (2)$$

$$\Sigma_k = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) (\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^T \quad (3)$$

$$\pi_k = \frac{N_k}{N} \quad (4)$$

Handwritten notes: $N_k = \sum_n \gamma_{nk}$

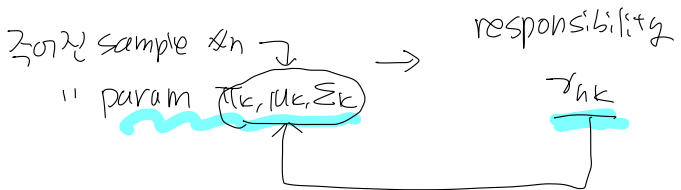
4. Evaluate the log likelihood

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

kmeans



MoG



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 $\mathbf{x}$.

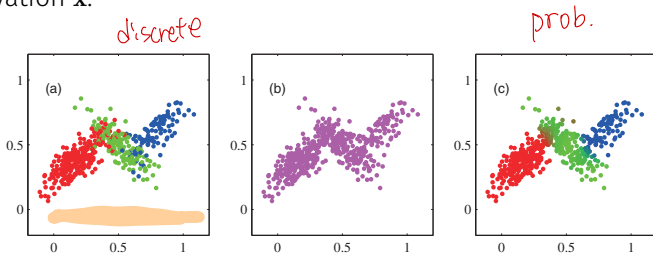


Figure 9.5 Example of 500 points drawn from the mixture of 3 Gaussians shown in Figure 2.23. (a) Samples from the joint distribution $p(\mathbf{z})p(\mathbf{x}|\mathbf{z})$ in which the three states of $\mathbf{z}$, corresponding to the three components of the mixture, are depicted in red, green, and 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

MoG: python example

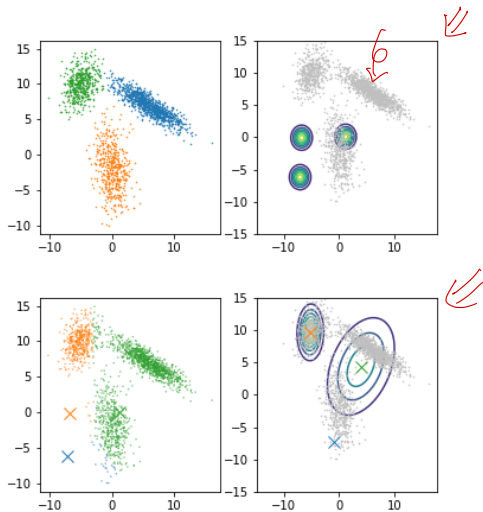


Figure 1: MoG python example: dataset and initialization (top) and the 1st iteration (bottom).

MoG: python example

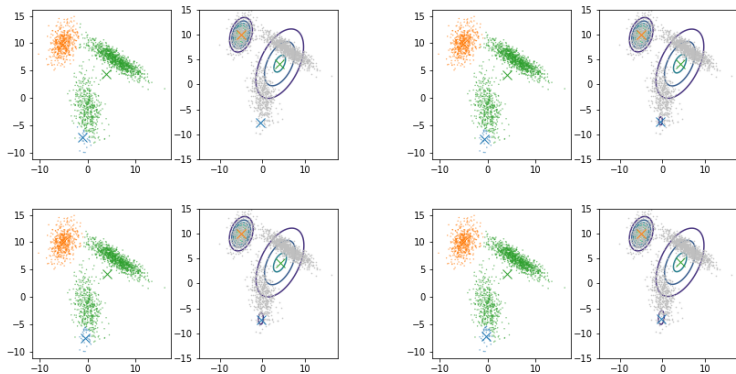


Figure 2: From the 2nd to 5th iterations.

MoG: python example

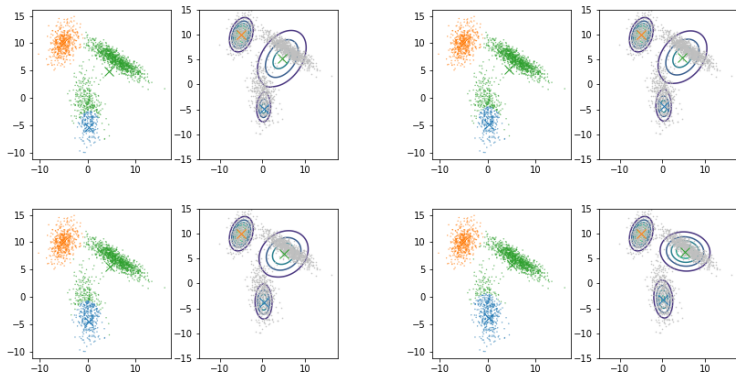


Figure 3: From the 12th to 15th iterations.

MoG: python example

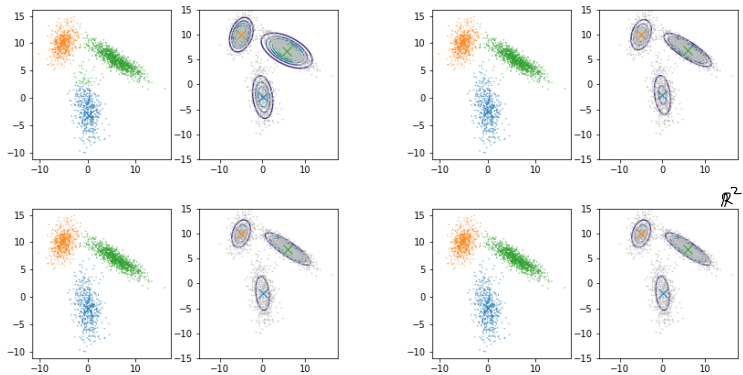


Figure 4: From the 16th to 19th iterations.

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