Concrete
Machine Learning

Deep User : 2020 Summer Program

Gaussian Mixture Model

Clustering algorithm

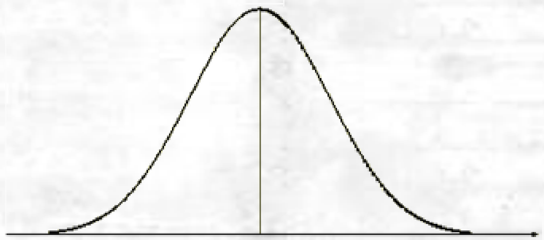
Unsupervised learning

Bayesian theory Base

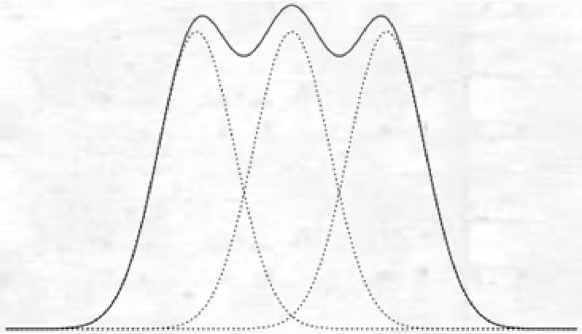
Labeling process

Get internal structure information

Knowledge discovery in data



Gaussian distribution



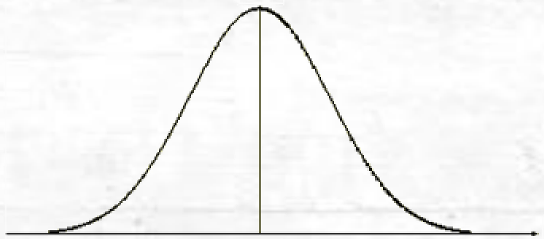
Mixture of Gaussian distributions

A | Gaussian Mixture Model

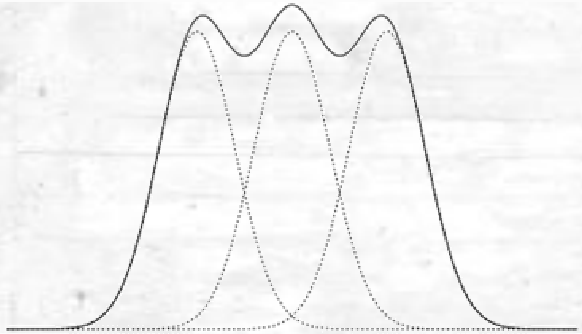
Mixture Model

전체 분포에서 하위 분포가 존재한다고 보는 모델

데이터가 모수를 갖는 여러개의 분포로부터 생성되었다고 가정하는 모델



Gaussian distribution

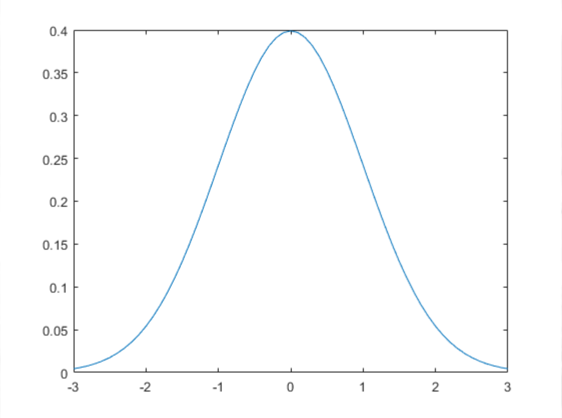


Mixture of Gaussian distributions

Mixture Model

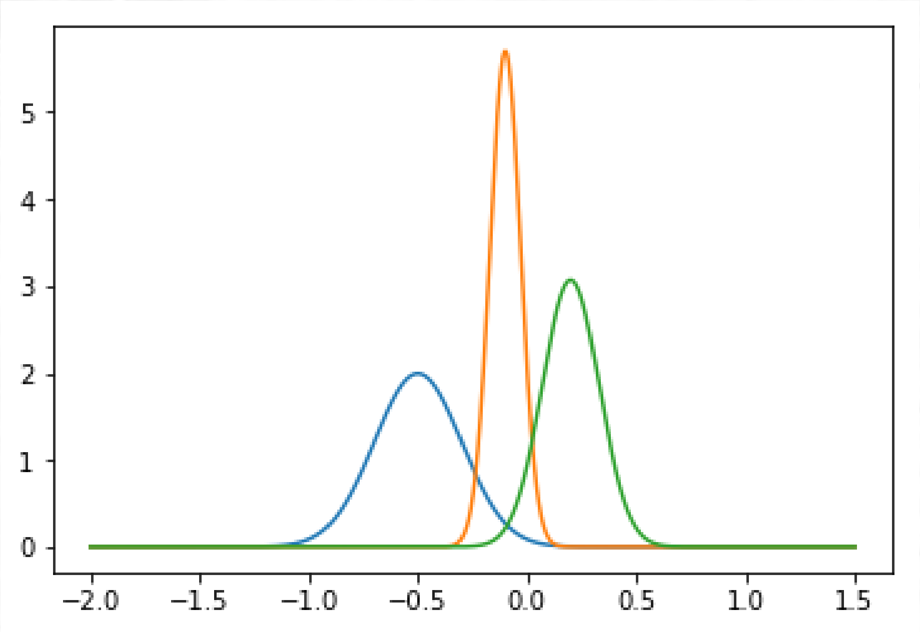
전체 분포에서 하위 분포가 존재한다고 보는 모델

데이터가 모수를 갖는 여러개의 분포로부터 생성되었다고 가정하는 모델



데이터가 K개의 **정규분포로부터** 생성되었다고 보는 모델

A | Gaussian Mixture Model

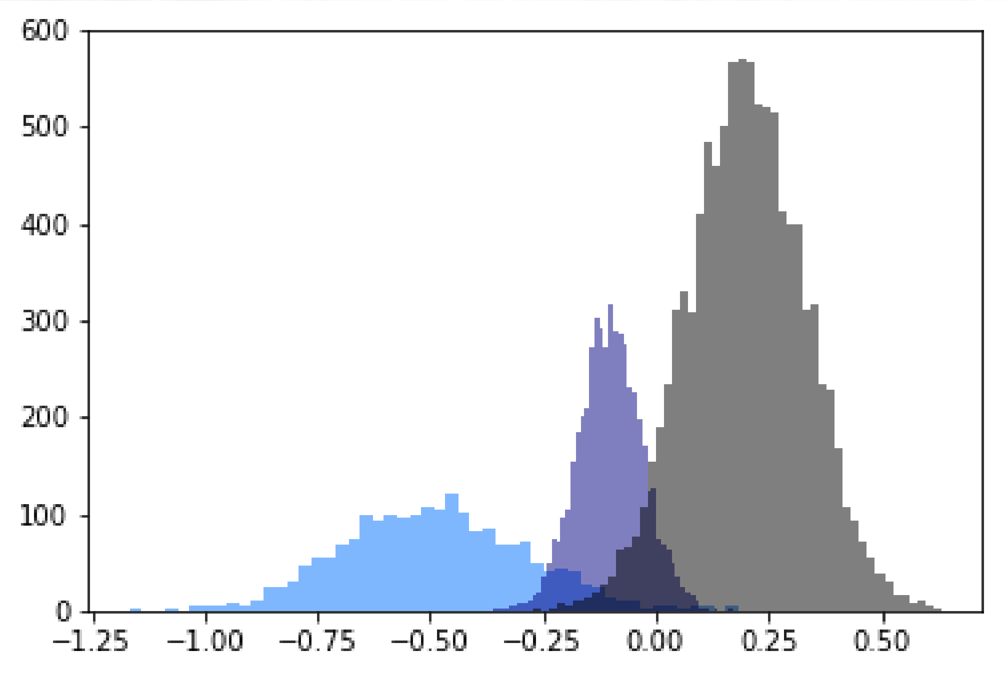


정규분포 1 : 평균 = -0.5, 표준편차 = 0.2 (파란색)

정규분포 2 : 평균 = -0.1, 표준편차 = 0.07 (주황색)

정규분포 3 : 평균 = 0.2, 표준편차 = 0.13 (녹색)

A | Gaussian Mixture Model

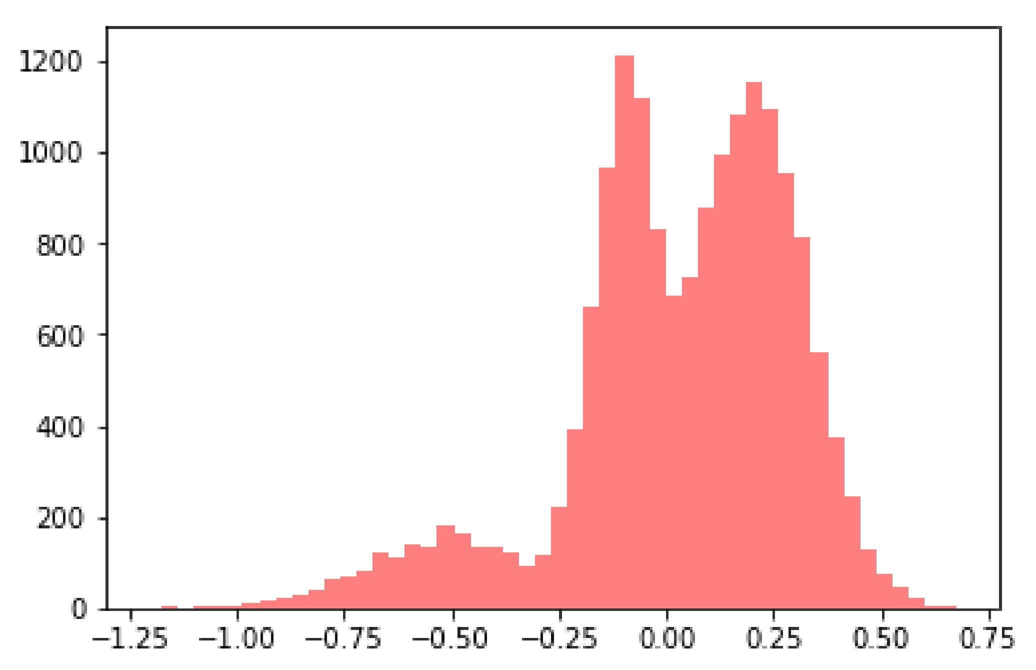


정규분포 1 : 평균 = -0.5, 표준편차 = 0.2 (파란색)

정규분포 2 : 평균 = -0.1, 표준편차 = 0.07 (주황색)

정규분포 3 : 평균 = 0.2, 표준편차 = 0.13 (녹색)

A | Gaussian Mixture Model

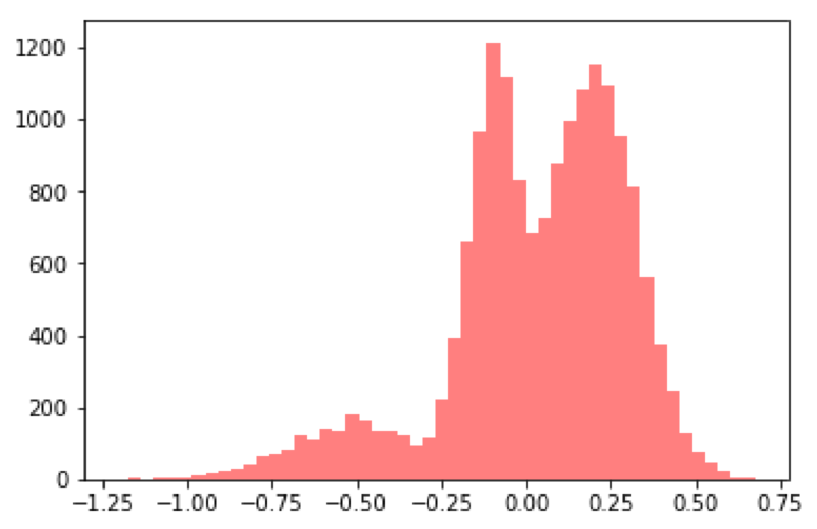


정규분포 1 : 평균 = -0.5, 표준편차 = 0.2 (파란색)

정규분포 2 : 평균 = -0.1, 표준편차 = 0.07 (주황색)

정규분포 3 : 평균 = 0.2, 표준편차 = 0.13 (녹색)

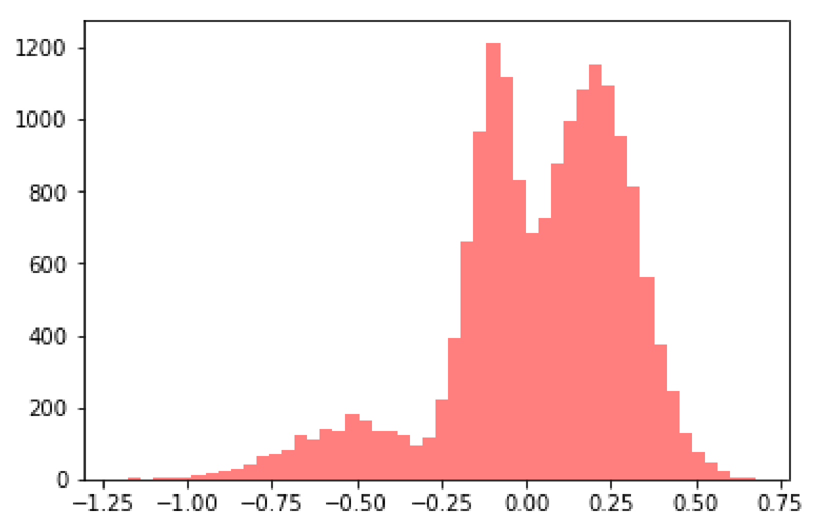
A | Gaussian Mixture Model



Weight : 3가지 정규분포 중 확률적으로 어디에서 속해 있는가를 나타내는 값

Mean, Variance : 모수(평균, 분산)

A | Gaussian Mixture Model



Mean, Variance : EM 알고리즘을 iterative하게 구현하여 모수 추정

A | Gaussian Mixture Model

$$p(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k) \quad (1)$$

$$0 \leq \pi_k \leq 1 \quad (2)$$

$$\sum_{k=1}^{K} \pi_k = 1 \quad (3)$$

적절한 πk,μk,Σk를 추정

A | Gaussian Mixture Model

$$p(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k) \quad (1)$$

$$0 \le \pi_k \le 1 \quad (2)$$

$$\sum_{k=1}^{K} \pi_k = 1 \quad (3)$$

$$\gamma(z_{nk}) = p(z_{nk}=1|x_n) \quad (4)$$

A | Gaussian Mixture Model

$$p(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k) \qquad (1)$$

$$0 \leq \pi_k \leq 1 \qquad (2)$$

$$\sum_{k=1}^{K} \pi_k = 1 \qquad (3)$$

$$\gamma(z_{nk}) = p(z_{nk}=1|x_n) \qquad (4)$$

$$\gamma(z_{nk}) = p(z_{nk}=1|x_n) = \frac{p(z_{nk}=1)p(x_n|z_{nk}=1)}{\sum_{j=1}^{K} p(z_{nj}=1)p(x_n|z_{nj}=1)} = \frac{\pi_k N(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(x_n|\mu_j, \Sigma_j)} \qquad (5)$$

A | EM Algorithm

$$\gamma(z_{nk}) = p(z_{nk}=1|x_n) = \frac{p(z_{nk}=1)p(x_n|z_{nk}=1)}{\sum_{j=1}^{K} p(z_{nj}=1)p(x_n|z_{nj}=1)} = \frac{\pi_k N(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(x_n|\mu_j, \Sigma_j)} \tag{5}$$

$$\mathcal{L}(X;\theta) = \ln p(X|\pi, \mu, \Sigma) = \ln\left\{ \prod_{n=1}^{N} p(x_n|\pi, \mu, \Sigma) \right\} = \sum_{n=1}^{N} \ln\left\{ \sum_{k=1}^{K} \pi_k N(x_n|\mu_k, \Sigma_k) \right\} \tag{6}$$

$$\frac{\partial \mathcal{L}(X;\theta)}{\partial \mu_k} = \sum_{n=1}^{N} \frac{\pi_k N(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(x_n|\mu_j, \Sigma_j)} \Sigma_k^{-1}(x_n - \mu_k) = 0$$

$$\Leftrightarrow \sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k) = 0$$

$$\therefore \mu_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) x_n}{\sum_{n=1}^{N} \gamma(z_{nk})} \tag{7}$$

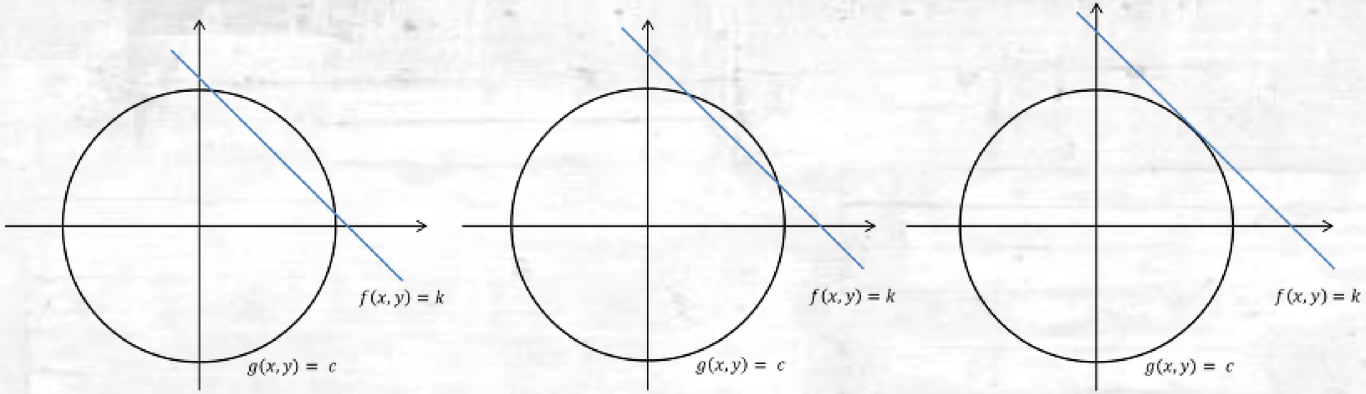
A | EM Algorithm

$$\frac{\partial \mathcal{L}(X;\theta)}{\partial \Sigma_k} = \sum_{n=1}^{N} \frac{\pi_k N(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(x_n|\mu_j, \Sigma_j)} \left\{ \frac{1}{2} \Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T \Sigma_k^{-1} - \frac{1}{2} \Sigma_k^{-1} \right\} = 0$$

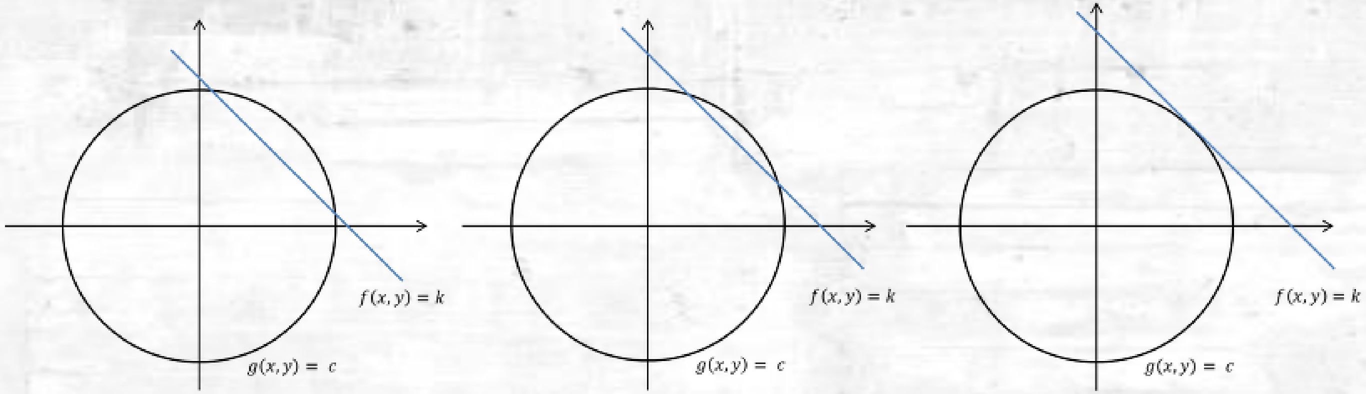
$$\Leftrightarrow \sum_{n=1}^{N} \gamma(z_{nk}) \{ \Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T - 1 \} = 0$$

$$\therefore \Sigma_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})(x_n-\mu_k)(x_n-\mu_k)^T}{\sum_{n=1}^{N} \gamma(z_{nk})} \tag{8}$$

A | 라그랑주 승수법



A | 라그랑주 승수법



$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) \qquad (1)$$

$$\nabla f = \lambda \nabla g \qquad (2)$$

$$L(x, y, \lambda) = f(x, y) - \lambda(g(x, y) - c) \qquad (3)$$

$$L(x, y, \lambda_1, \lambda_2, \lambda_2, \ldots, \lambda_n) = f(x, y) - \sum_{i=1}^{N} \lambda_i (g_i(x, y) - c_i) \qquad (4)$$

A | EM Algorithm

$$\frac{\partial \mathcal{L}(X;\theta)}{\partial \Sigma_k} = \sum_{n=1}^{N} \frac{\pi_k N(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(x_n|\mu_j, \Sigma_j)} \left\{ \frac{1}{2} \Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T \Sigma_k^{-1} - \frac{1}{2} \Sigma_k^{-1} \right\} = 0$$

$$\Leftrightarrow \sum_{n=1}^{N} \gamma(z_{nk}) \{ \Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T - 1 \} = 0$$

$$\therefore \Sigma_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})(x_n-\mu_k)(x_n-\mu_k)^T}{\sum_{n=1}^{N} \gamma(z_{nk})} \tag{8}$$

A | EM Algorithm

$$\frac{\partial \mathcal{L}(X;\theta)}{\partial \Sigma_k} = \sum_{n=1}^{N} \frac{\pi_k N(x_n|\mu_k, \ \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(x_n|\mu_j , \ \Sigma_j)} \left\{ \frac{1}{2} \Sigma_k^{-1}(x_n - \mu_k)(x_n - \mu_k)^T \Sigma_k^{-1} - \frac{1}{2} \Sigma_k^{-1} \right\} = 0$$

$$\Leftrightarrow \sum_{n=1}^{N} \gamma(z_{nk}) \{ \Sigma_k^{-1}(x_n - \mu_k)(x_n - \mu_k)^T - 1 \} = 0$$

$$\therefore \ \Sigma_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k)(x_n - \mu_k)^T}{\sum_{n=1}^{N} \gamma(z_{nk})} \qquad (8)$$

$$J(X;\theta, \ \lambda) = \sum_{n=1}^{N} \ln \sum_{k=1}^{K} \pi_k N(x_n|\mu_k, \ \Sigma_k) + \lambda \left(1 - \sum_{k=1}^{K} \pi_k \right) \qquad (9)$$

A | EM Algorithm

$$J(X;\theta,\ \lambda) = \sum_{n=1}^{N} \ln \sum_{k=1}^{K} \pi_k N(x_n|\mu_k,\ \Sigma_k) + \lambda\left(1 - \sum_{k=1}^{K} \pi_k\right) \qquad (9)$$

$$\frac{\partial J(X;\theta,\ \lambda)}{\partial \pi_k} = \sum_{n=1}^{N} \frac{N(x_n|\mu_k,\ \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(x_n|\mu_j,\ \Sigma_j)} - \lambda = 0$$

$$\Leftrightarrow \sum_{k=1}^{K} \sum_{n=1}^{N} \frac{\pi_k N(x_n|\mu_k,\ \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(x_n|\mu_j,\ \Sigma_j)} - \lambda \sum_{k=1}^{K} \pi_k = 0$$

$$\Leftrightarrow \sum_{k=1}^{K} \sum_{n=1}^{N} \gamma(z_{nk}) - \lambda = 0 \qquad \left(\because \sum_{k=1}^{K} \pi_k = 1\right)$$

$$\therefore \lambda = N \qquad \left(\because \sum_{k=1}^{K} \gamma(z_{nk}) = 1\right) \qquad (10)$$

A | EM Algorithm

$$\frac{\partial J(X;\theta,\ \lambda)}{\partial \pi_k} = \sum_{n=1}^{N} \frac{N(x_n|\mu_k,\ \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(x_n|\mu_j,\ \Sigma_j)} - N = 0$$

$$\Leftrightarrow \sum_{n=1}^{N} \frac{\pi_k N(x_n|\mu_k,\ \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(x_n|\mu_j,\ \Sigma_j)} - N\pi_k = 0$$

$$\therefore\ \pi_k = \frac{1}{N} \sum_{n=1}^{N} \gamma(z_{nk}) \qquad (11)$$

A | Gaussian Mixture Model

Algorithm 1: EM algorithm for GMM

Input : a given data $X = \{x_1, x_2, \dots, x_n\}$

Output : $\pi = \{\pi_1, \pi_2, \dots, \pi_K\}$,
$\mu = \{\mu_1, \mu_2, \dots, \mu_K\}$,
$\Sigma = \{\Sigma_1, \Sigma_2, \dots, \Sigma_K\}$

Randomly initialize π, μ, Σ
for $t = 1 : T$ **do**
  // E-step
  for $n = 1 : N$ **do**
    for $k = 1 : K$ **do**
      $\gamma(z_{nk}) = \frac{\pi_k N(x_n | \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(x_n | \mu_j, \Sigma_j)}$
    end
  end
  // M-step
  for $k = 1 : K$ **do**
    $\mu_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) x_n}{\sum_{n=1}^{N} \gamma(z_{nk})}$
    $\Sigma_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k)(x_n - \mu_k)^T}{\sum_{n=1}^{N} \gamma(z_{nk})}$
    $\pi_k = \frac{1}{N} \sum_{n=1}^{N} \gamma(z_{nk})$
  end
end

Algorithm 2: GMM classification

Input : a given data $X = \{x_1, x_2, \dots, x_n\}$,
$\pi = \{\pi_1, \pi_2, \dots, \pi_K\}$,
$\mu = \{\mu_1, \mu_2, \dots, \mu_K\}$,
$\Sigma = \{\Sigma_1, \Sigma_2, \dots, \Sigma_K\}$

Output : class labels $y = \{y_1, y_2, \dots, y_N\}$ for X

for $n = 1 : N$ **do**
  $y_n = \arg\max_k \gamma(z_{nk})$
end

A | Gaussian Mixture Model

(π= $initialprobability$,it should be [1/3, 1/3, 1/3].

μ=it is diagonal matrix and diagonal entry is random on initial stage.

its shape is 3 x n matrix each row means each cluster, select 3 data randomly and use it as mean of each cluster.

Σ= it is diagonal matrix and diagonal entry is random on initial stage, its shape is 3 x n x n matrix, first dimension means cluster, others means covariancematrix of feature, N = Data Point, K=Num of Gaussian distribution, T = iteration)

Algorithm 1: EM algorithm for GMM

Input : a given data $X = \{x_1, x_2, \ldots, x_N\}$
Output: $\pi = \{\pi_1, \pi_2, ..., \pi_K\}$,
$\mu = \{\mu_1, \mu_2, ..., \mu_K\}$,
$\Sigma = \{\Sigma_1, \Sigma_2, ..., \Sigma_K\}$

Randomly initialize π, μ, Σ
for $t = 1 : T$ **do**
  // E-step
  for $n = 1 : N$ **do**
    for $k = 1 : K$ **do**
      $\gamma(z_{nk}) = \frac{\pi_k N(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(x_n|\mu_j, \Sigma_j)}$
    end
  end
  // M-step
  for $k = 1 : K$ **do**
    $\mu_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) x_n}{\sum_{n=1}^{N} \gamma(z_{nk})}$
    $\Sigma_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k)(x_n - \mu_k)^T}{\sum_{n=1}^{N} \gamma(z_{nk})}$
    $\pi_k = \frac{1}{N} \sum_{n=1}^{N} \gamma(z_{nk})$
  end
end

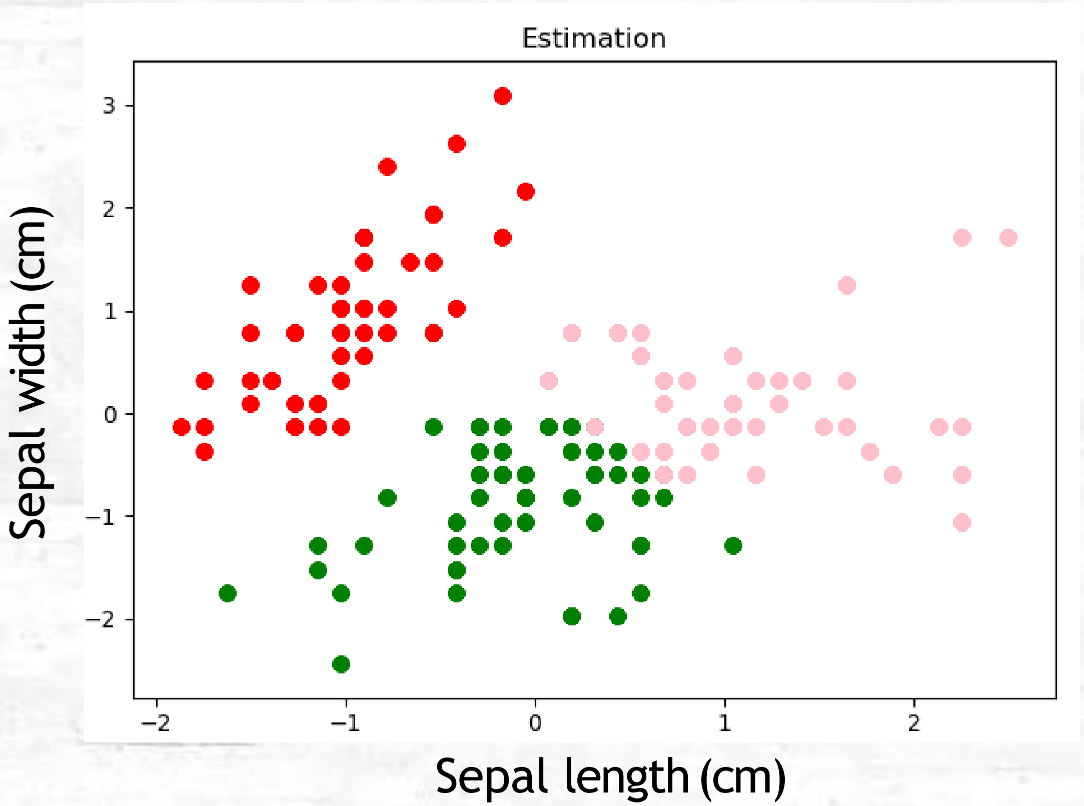
For a given data x, GMM expresses the probability that x will occur as the sum of several Gaussian probability density functions as shownin [Equation 1].

$$p(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k) \tag{1}$$

Learning GMM is equivalent to estimating the appropriate πk, μk, Σk for the given data X = {x1, x2, ..., xN}.

A | Gaussian Mixture Model

After build all of function, you can see below result from python console when you compile “main.py”
GMM – EM algorithm can reach local optimal, sometime but it mostly shows this result.



Feature (Z-score) normalization

Example of
PDF normalization =

PDF1	PDF2
0.001	0.001
0.002	0.002
0.001	0.001
0.002	0.002
0.001	0.001
0.002	0.002
0.001	0.001
0.002	0.002

PDF1	PDF2
0.5	0.5
0.5	0.5
0.5	0.5
0.5	0.5
0.5	0.5
0.5	0.5
0.5	0.5
0.5	0.5

DeepUser

Gaussian Mixture Model

THANKS