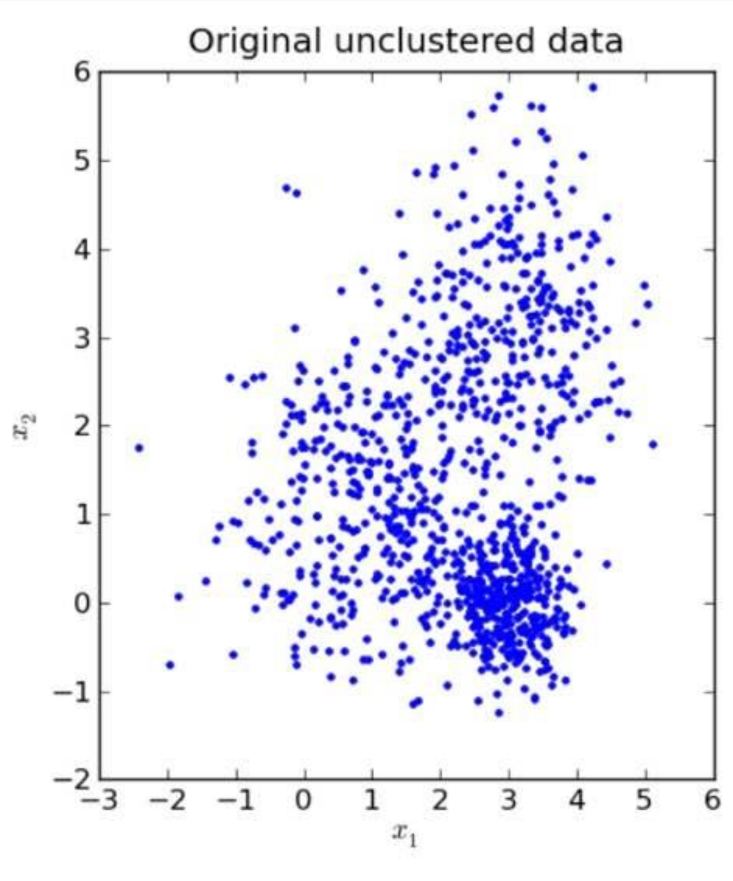
EM GMM

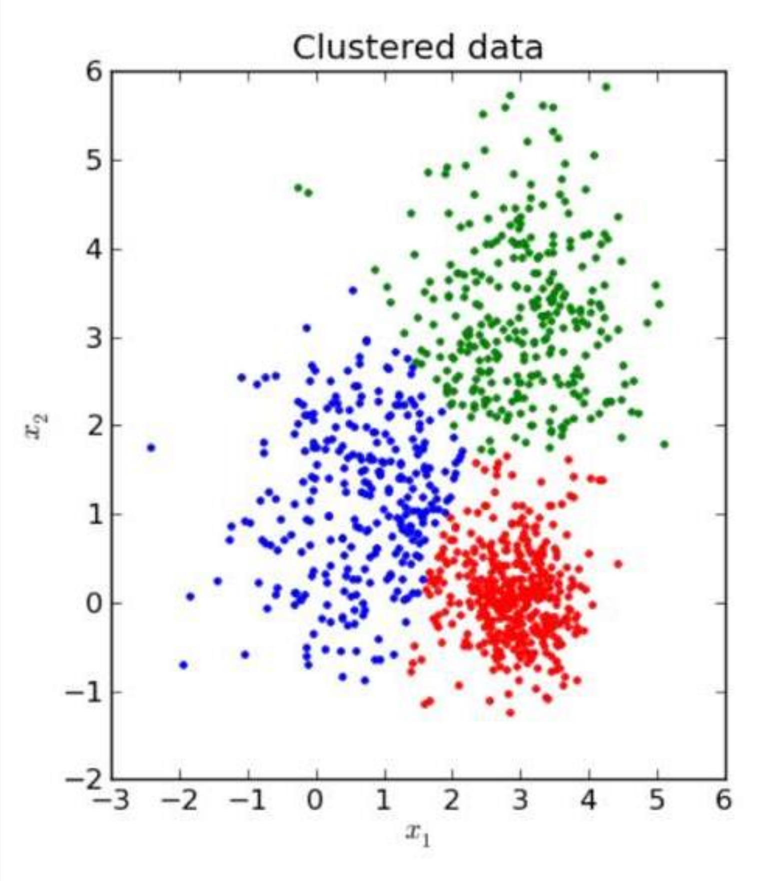
ESC
2020
1학기
PRML
Expectation
Maximization
Gaussian
Mixture
Model

서경덕
강경훈

K-Means Clustering



Clustering



K-Means Clustering

K−means Clustering?

클러스터 내부에 속한 데이터들이 서로 가깝다. (유클리드 거리개념(l_2 $norm$))

가까운 기준?

클러스터의 중심과 가깝다. $data\ set : \{X_1, X_2, \cdots, X_N\}$ -> D dimension

목적?

Data set을 K개의 군집으로 나누자.

$$J = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\|X_n - \mu_k\|^2 \qquad r_{nk} = \begin{cases} 1\ (if\ X_n \text{ is } included\ in\ k_{th}\ cluster\ ) \\ 0\ (\ other\ wise) \end{cases}$$

J를 최소화하는 μ_k, r_{nk}를 구해야한다.

K-Means Clustering

Iterative method(Alternative update)

① Initial μ_k선정

② μ_k고정 후 J를 최소화하는 r_{nk}를 구한다.

※

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|X_n - \mu_k\|^2$$

$$r_{nk} = \begin{cases} 1 \; (if \; X_n \text{ is } included \; in \; k_{th} \; cluster \, , \quad k = arg \min_j \|X_n - \mu_j\|^2) \\ 0 \; (\, other \; wise) \end{cases}$$

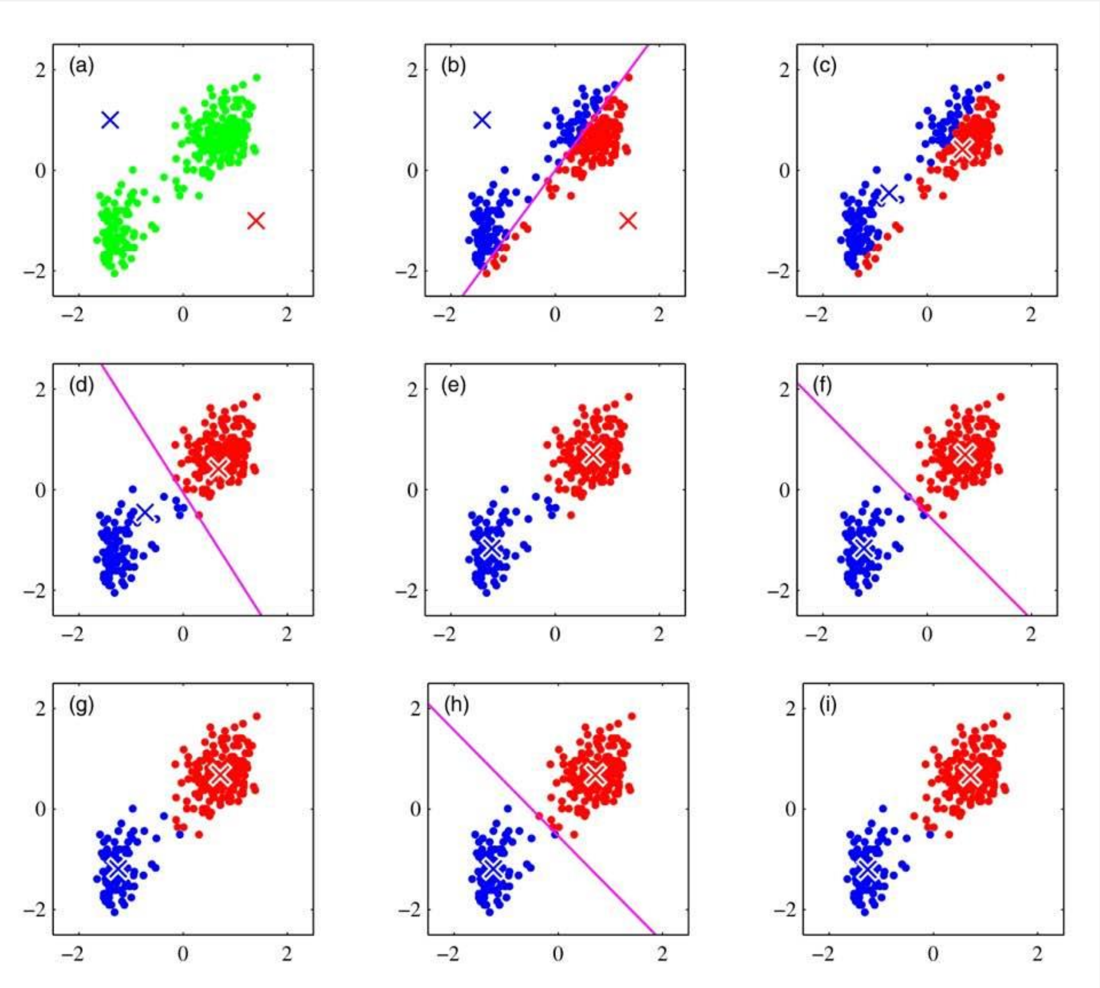
③ r_{nk}고정 후 J를 최소화하는 u_k를 구한다.

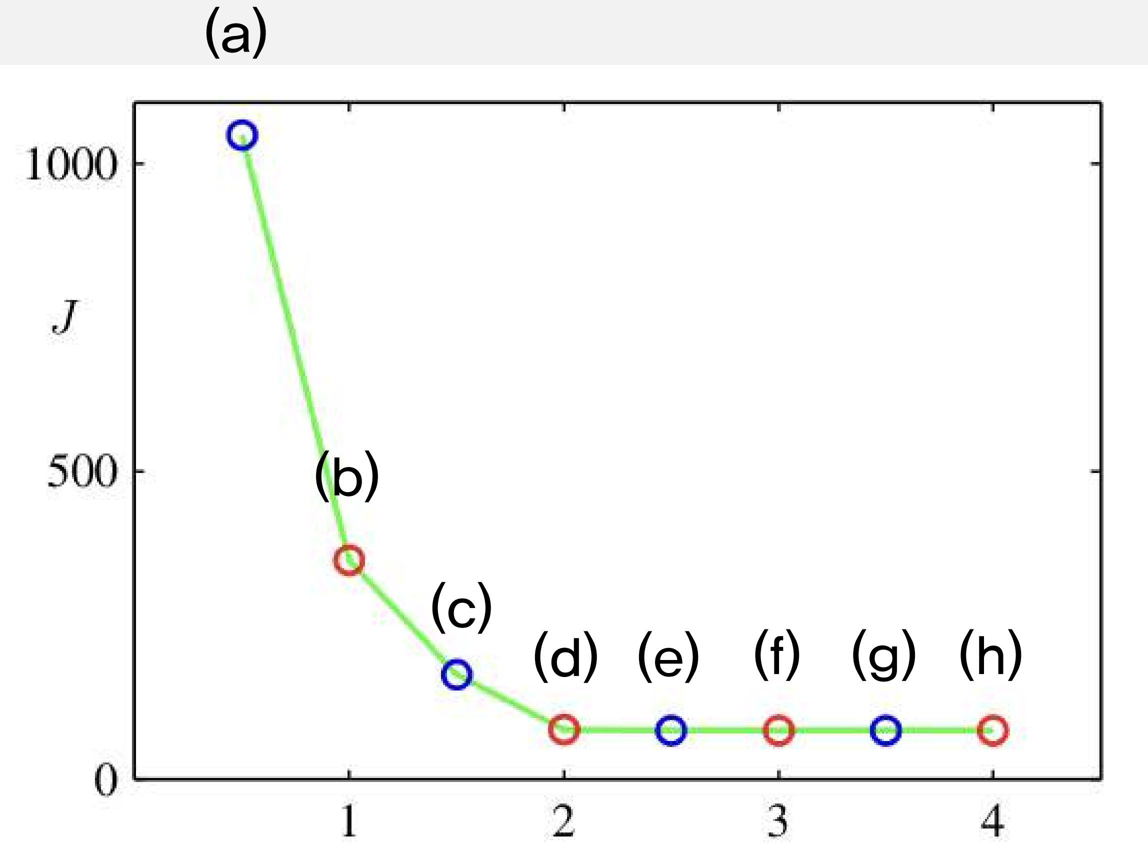
※

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|X_n - \mu_k\|^2$$

$$2 \sum_{n=1}^{N} r_{nk} (X_n - \mu_k) = 0 \, , \mu_k = \frac{\sum_n r_{nk} X_n}{\sum_n r_{nk}}$$

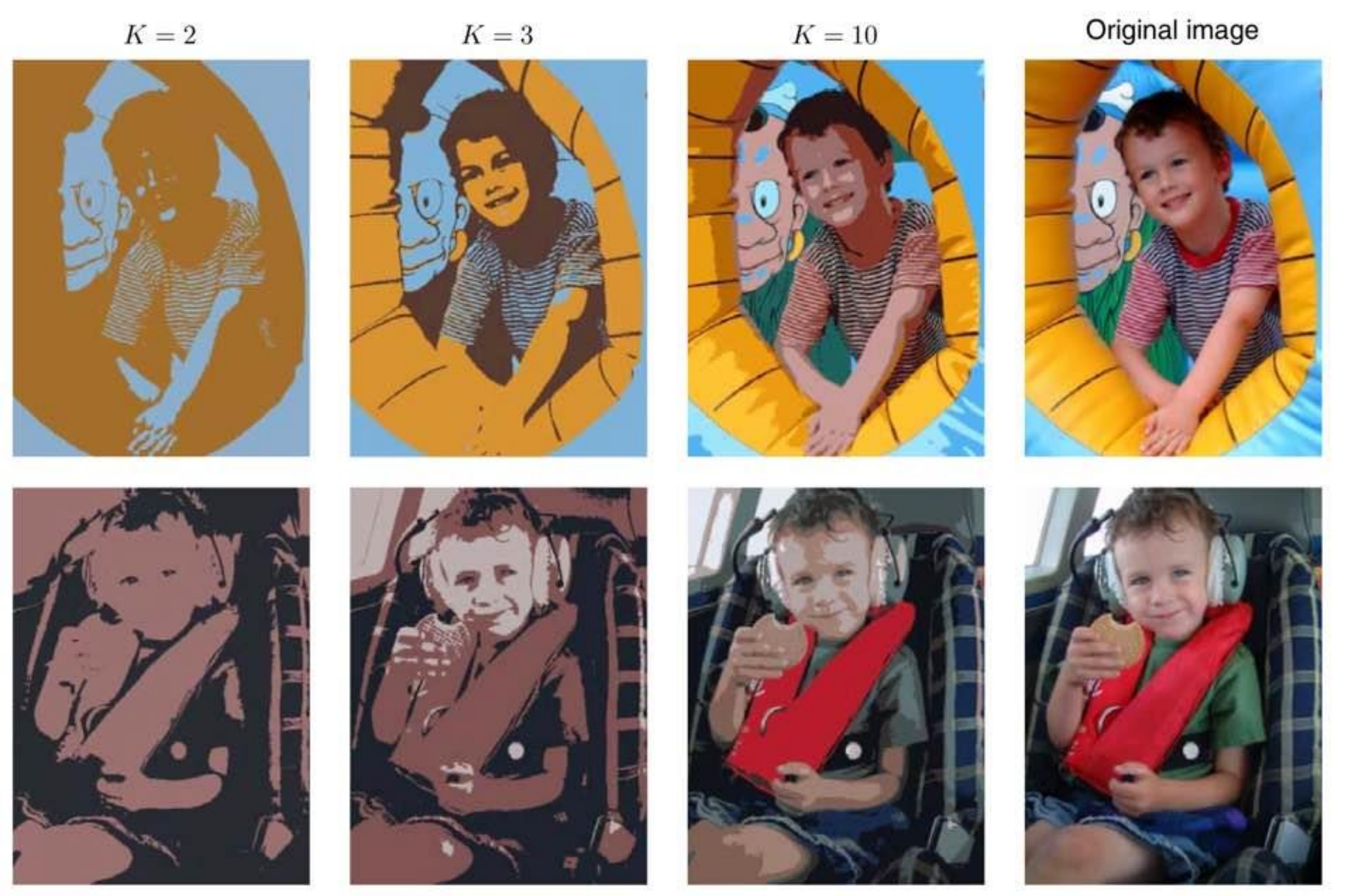
K-Means Clustering





K-Means Clustering

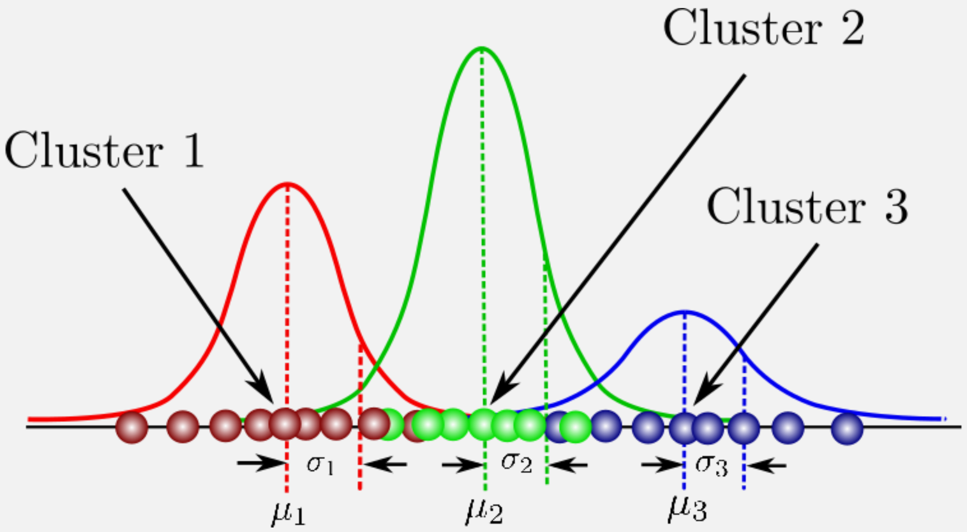
Image segmentation and compression

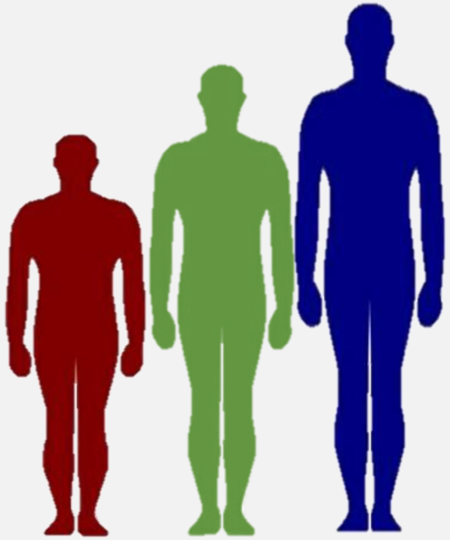


$$data\ set = \{p_1, p_2, \cdots, p_N\}$$

$$\left\{\begin{pmatrix} R_1 \\ G_1 \\ B_1 \end{pmatrix}, \begin{pmatrix} R_2 \\ G_2 \\ B_2 \end{pmatrix}, \cdots, \begin{pmatrix} R_N \\ G_N \\ B_N \end{pmatrix}\right\}$$

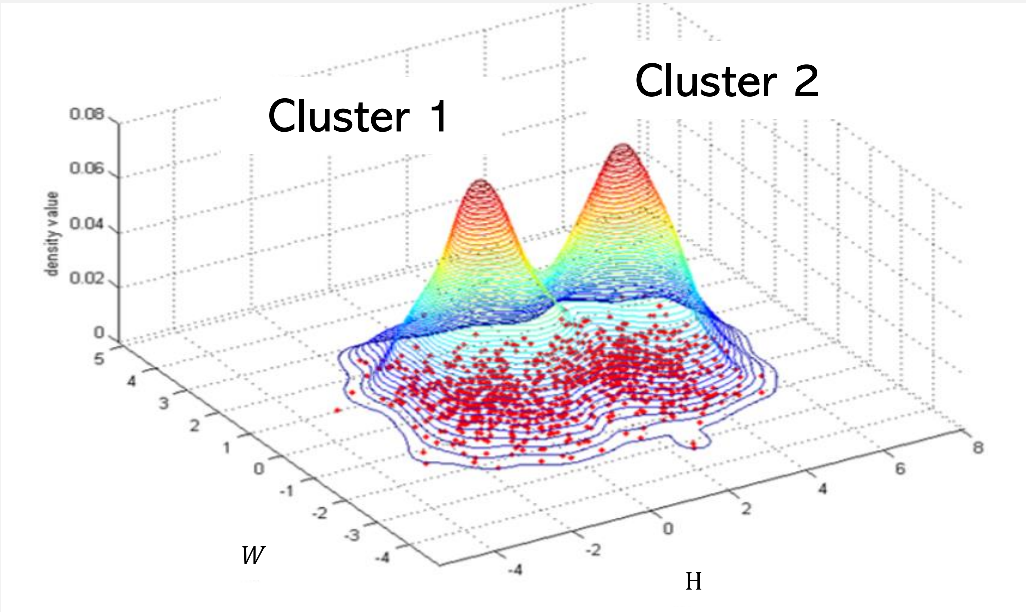
Mixtures of Gaussians - Definition





$$X = (X_1, X_2, \cdots, X_{150})^T$$

$$X|Z = 1 \sim N(\mu_1, \sigma_1^2)$$
$$X|Z = 2 \sim N(\mu_2, \sigma_2^2)$$
$$X|Z = 3 \sim N(\mu_3, \sigma_3^2)$$



$$X = \left(\begin{pmatrix} H_1 \\ W_1 \end{pmatrix}, \begin{pmatrix} H_2 \\ W_2 \end{pmatrix}, \cdots, \begin{pmatrix} H_{100} \\ W_{100} \end{pmatrix} \right)^T$$

$$X|Z = 1 \sim N(\mu_1, \Sigma_1)$$
$$X|Z = 2 \sim N(\mu_2, \Sigma_2)$$

Mixtures of Gaussians - Definition

$$Z = \{e_1, e_2, \cdots, e_K\} = \{\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}, \begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix}, \cdots, \begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}\}$$

k = 1일 때, $Z = (1,0,\cdots,0)$

$cluster\ index$

$$Given\ Z = e_k, X \sim N(\mu_k, \Sigma_k)$$

$where\ \mu_1, \cdots, \mu_K \in \mathbb{R}^d$

$c_1, \cdots, c_K \in d \times d\ cov\ matrix$

$$P(X|Z = e_k) = N(X|\mu_k, \Sigma_k)$$

$$P(Z = e_k) = \pi_k$$

$$p = \frac{1}{\sqrt{2\pi|\Sigma_k|}} \exp(-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k))$$

각 Cluster의 분포가 정규분포를 따른다고 가정한다.

$$P(X, Z) = P(Z) \times P(X|Z) = \pi_k N(X|\mu_k, \Sigma_k)$$

$$P(X) = \sum_Z P(Z) \times P(X|Z) = \sum_{k=1}^{K} \pi_k N(X|\mu_k, \Sigma_k)$$

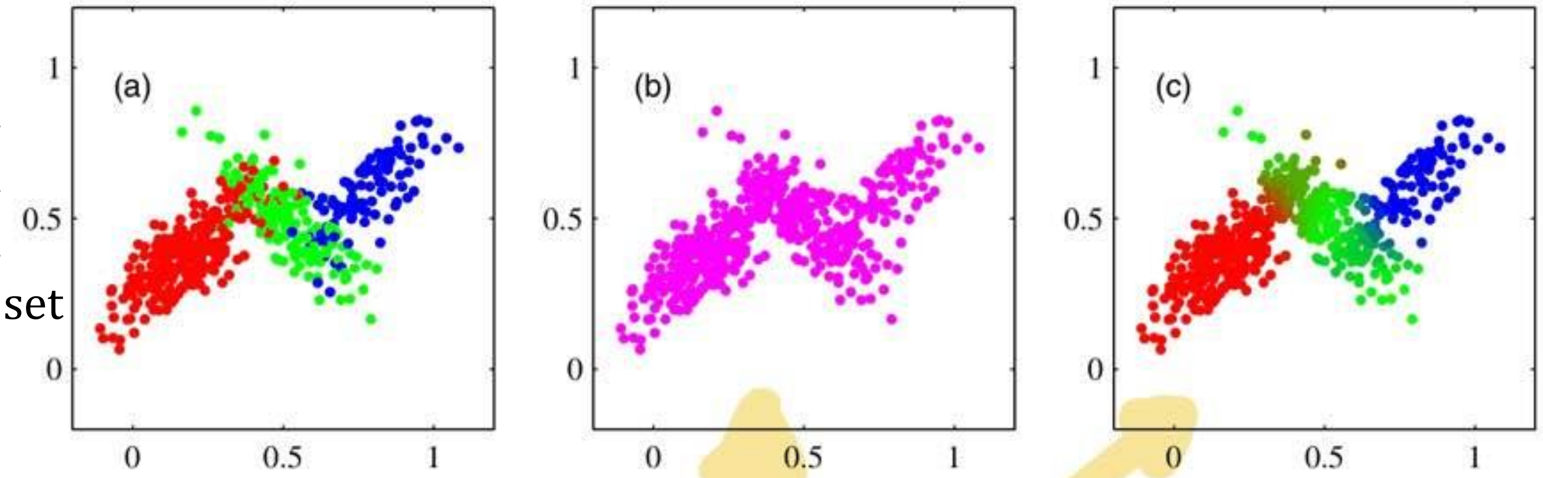
이 때 $\sum \pi_k = 1, \pi_k$는 non−negativ하므로 P(X)는 Convex combination이다.

Mixtures of Gaussians - Definition

$$P(Z = e_k|X) = \frac{P(Z) \times P(X|Z)}{P} = \frac{\pi_k N(X|\mu_k, \Sigma_k)}{\sum_{k=1}^{K} \pi_k N(X|\mu_k, \Sigma_k)} = r(Z_{nk})$$

X incomplete data set

$r(Z_{n1}) = 1$
$r(Z_{n2}) = 1$
$r(Z_{n3}) = 1$
(X, Z) complete data set



$r(Z_{n2}) = 0.5$
$r(Z_{n3}) = 0.5$

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

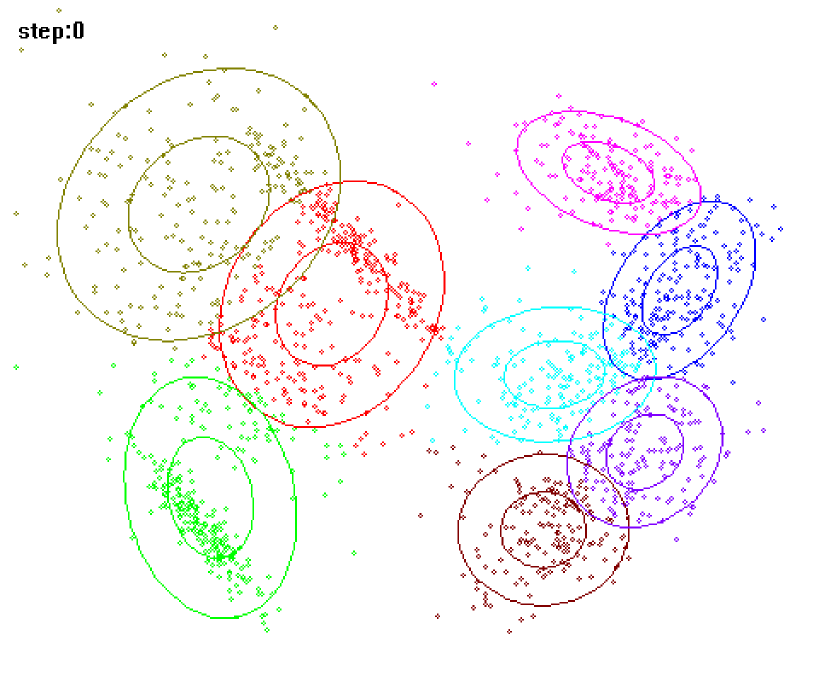
Mixtures of Gaussians - Definition

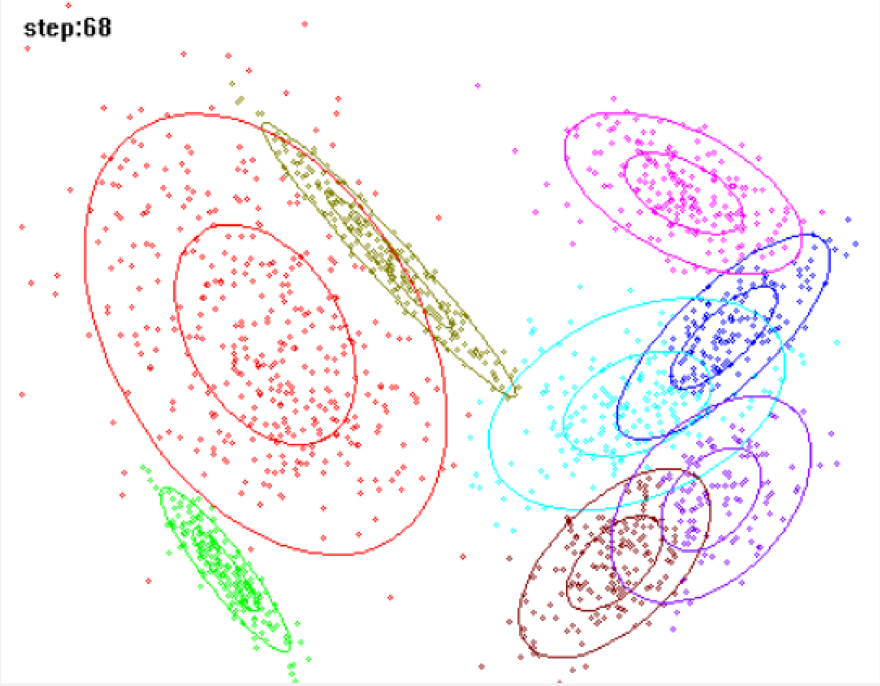
예시

$$x_0 = \begin{bmatrix} 1.2 \\ -3.6 \\ 8.3 \end{bmatrix}$$

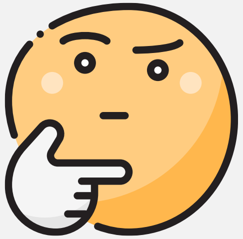
$$p(x_0, Z=1) = 0.2$$
$$p(x_0, Z=2) = 3.7$$
$$p(x_0, Z=3) = 0.5$$

$$p(Z=1|x) = \frac{0.2}{0.2+3.7+0.5}$$





Mixtures of Gaussians - Likelihood



결국 값을 구하려면 모수를
추정해야하는데... How??

각 모수의 MLE를 구하는데
한꺼번에 구하면 힘드니까
Alternative한 방식으로 구하자

EM FOR GMM

Mixtures of Gaussians - Likelihood

Iterative method(Alternative update)

① Initial θ_k선정 $\theta_k = \{\pi_k, \mu_k, \Sigma_k\}$

② $E - step$ 주어진 모수로 $r\{Z_{nk}\}$ 계산

$$P(Z = e_k|X) = \frac{P(Z) \times P(X|Z)}{P} = \frac{\pi_k N(X|\mu_k, \Sigma_k)}{\sum_{k=1}^{K} \pi_k N(X|\mu_k, \Sigma_k)} = r(Z_{nk})$$

③ $M - step$ 주어진 $r\{Z_{nk}\}$로 모수 재추정

$$\mathrm{L}(\theta_k|X) = p(X|\theta_k) = \prod_{n=1}^{N} \sum_{k=1}^{K} \pi_k N(X_n|\mu_k, \Sigma_k)$$

$$\ln p(X|\theta_k) = \sum_{n=1}^{N} \ln \boxed{\sum_{k=1}^{K} \pi_k N(X_n|\mu_k, \Sigma_k)}$$

Two ways to solve this equation.

$\longrightarrow$ *Incomplete data set*

$\longrightarrow$ *complete data set*

로그 안에 시그마가 들어있어서 까다로운 형태.

EM for GMM

$$\ln p(X|\theta_k) = \sum_{n=1}^{N} \ln \sum_{k=1}^{K} \pi_k N(X_n|\mu_k, \Sigma_k)$$

$$\frac{\partial}{\partial \mu_k} \ln p(X|\theta_k) = -\sum_{n=1}^{N} \frac{\pi_k N(X|u_k, \Sigma_k)}{\sum_{k=1}^{K} \pi_k N(X|u_k, \Sigma_k)} \Sigma_k^{-1}(X_n - \mu_k) = 0$$

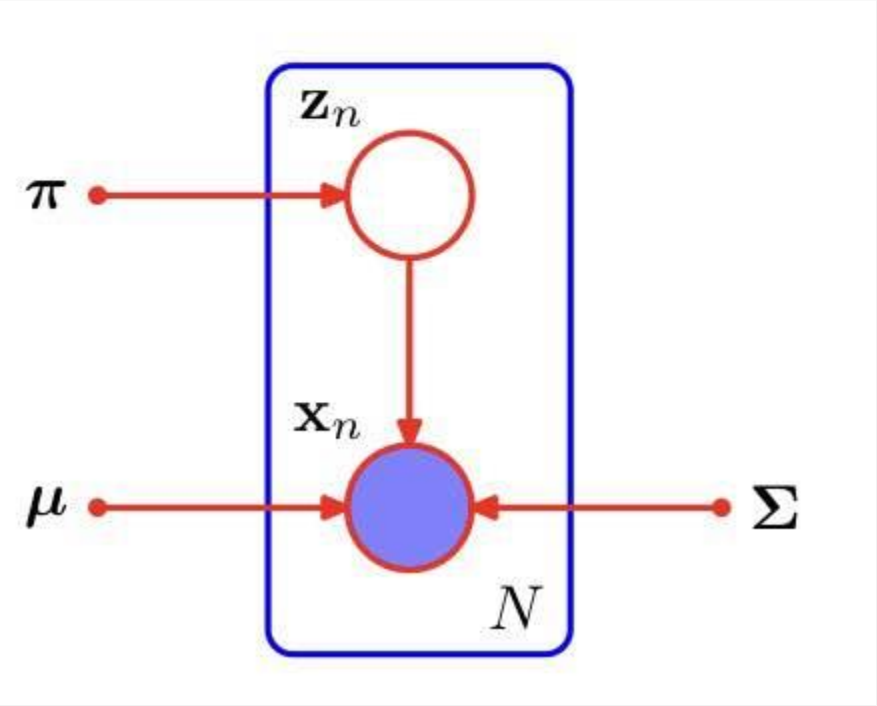
$$\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} r(Z_{nk}) X_n$$

$$\Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} r(Z_{nk})(X_n - \mu_k)(X_n - \mu_k)^T$$

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

$$N_k = \sum_{n=1}^{N} r(Z_{nk})$$

Incomplete data set



EM for GMM

X: Incomplete data set $\rightarrow \ln p(X|\theta) = \ln \sum_{Z} p(X, Z|\theta)$

(X, Z): Complete data set $\rightarrow \ln p(X, Z|\theta)$

실제 데이터에서는 Z가 관찰되지 않기 때문에
$Likelihood$ 대신 $\rightarrow E_Z[\ln p(X, Z|\theta)]$

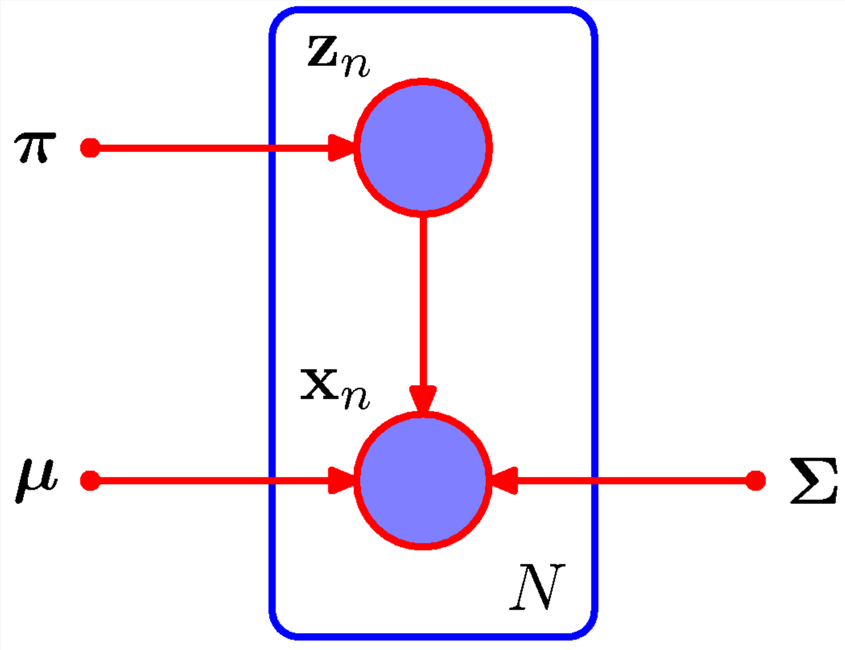
$$E_Z[\ln p(X, Z|\theta)] = \sum_{Z} p(Z|X, \theta) \times \ln p(X, Z|\theta)$$

Z에 대해 우리가 얻을 수 있는 정보

$$E - step: \; Q(\theta, \theta^{old}) = \sum_{Z} p(Z|X, \theta^{old}) \times \ln p(X, Z|\theta)$$

$$M - step: \theta^{new} = arg \max_{\theta} Q(\theta, \theta^{old})$$

complete data set



EM for GMM

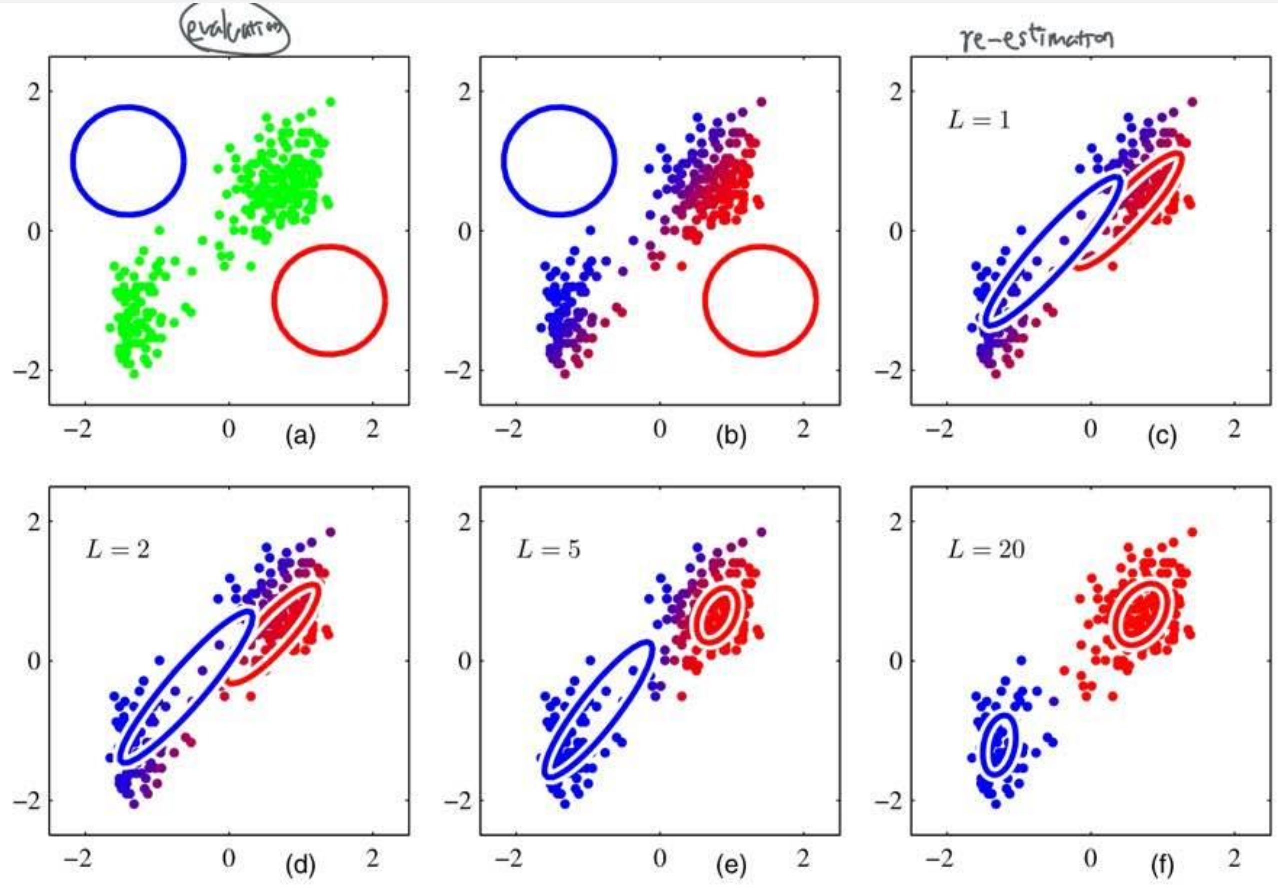
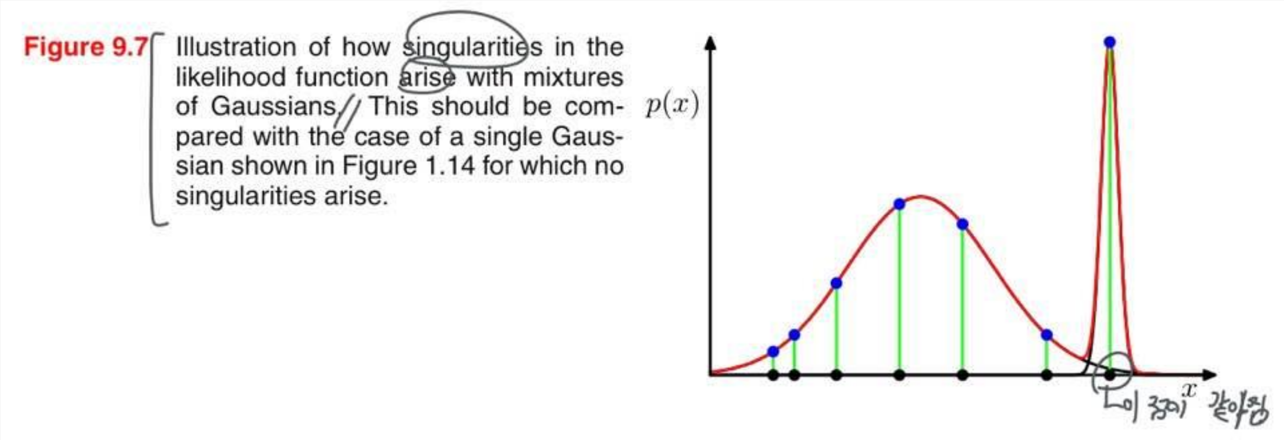


Figure 9.8 Illustration of the EM algorithm using the Old Faithful set as used for the illustration of the K-means algorithm in Figure 9.1. See the text for details.

EM for GMM

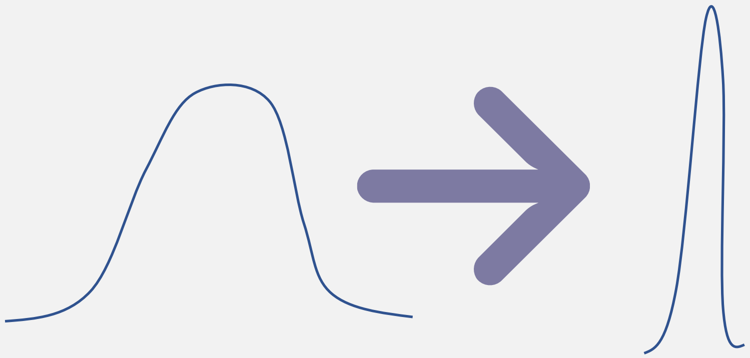
Problem of singularities



$\mu_2 = X_n$이 되버린다.

$$N(X_n|X_n, \sigma_2^2 I) = \frac{1}{\sqrt{2\pi}} \times \frac{1}{\sigma_2} \left(\because \exp\left\{-\frac{1}{2}(x-\mu)\sim\right\}\right)$$

$L(\pi, \mu, \sigma^2)$ 최대화 -> σ_2 최소화



분산이 줄고, 분포가 μ_2값에 수렴한다.

단일 가우시안모델의 경우 likelihood를 0으로 만들어주는 경우가 많이 존재해서 해당 문제가 발생하지 않는다.

문제해결을 위해 값이 붕괴되는 지점을 찾아서 평균값을 다른 값으로 바꿔준다.

Relation to K-means

K$-$*means*: Hard assignment

$K-means$ 에서는 Z만 고려

군집별 데이터의 등분산성을 가정

평균과 거리만 고려, 각 클러스터의 분산 고려X

EM: Soft assignment

$EM\ IN\ GMM$에서는 배정확률값을 고려

$$\gamma(z_{nk}) = \frac{\pi_k \exp\left\{-\frac{\|x_n-\mu_k\|^2}{2\sigma}\right\}}{\sum_j \pi_j \exp\left\{-\frac{\|x_n-\mu_j\|^2}{2\sigma}\right\}}$$

등분산성과 각 분산이 독립이라는 가정을 하자.

$\sigma \to 0$이면, soft에서 hard로 바뀐다.

Bernoulli mixture model

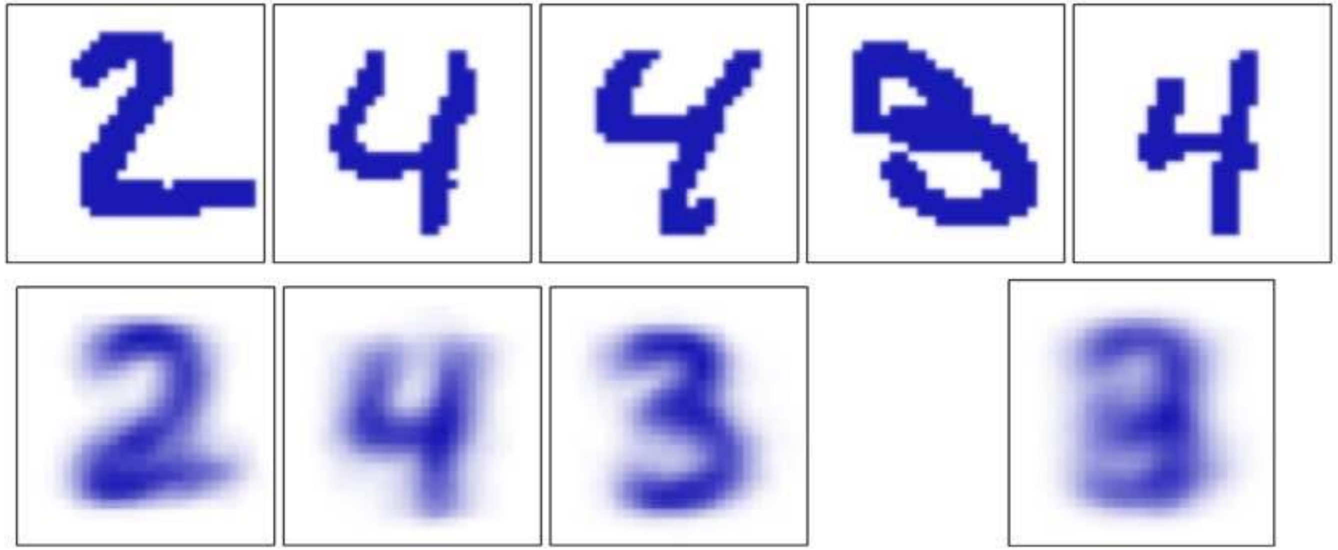


Figure 9.10 Illustration of the Bernoulli mixture model in which the top row shows examples from the digits data set after converting the pixel values from grey scale to binary using a threshold of 0.5. On the bottom row the first three images show the parameters μ_{ki} for each of the three components in the mixture model. As a comparison, we also fit the same data set using a single multivariate Bernoulli distribution, again using maximum likelihood. This amounts to simply averaging the counts in each pixel and is shown by the right-most image on the bottom row.

$$p(X|\mu_k) = \prod_{i=1}^{D} \mu_{ki}^{x_i}(1-\mu_{ki})^{1-x_i}$$

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

$$\ln p(X|\mu,\pi) = \sum_{n=1}^{N} \ln \sum_{k=1}^{K} \pi_k p(X_n|\mu_k)$$

$$E_Z[\ln p(X,Z|\mu,\pi)]$$

이산확률변수에서의 혼합 –> 베르누이 분포의 혼합 –> Handwritten digit recognition
Pixel 여부는 Binary class이고 3개의 군집으로 분류를 하였다.

경청해주셔서
감사합니다

Thank You