CS4641B Machine Learning

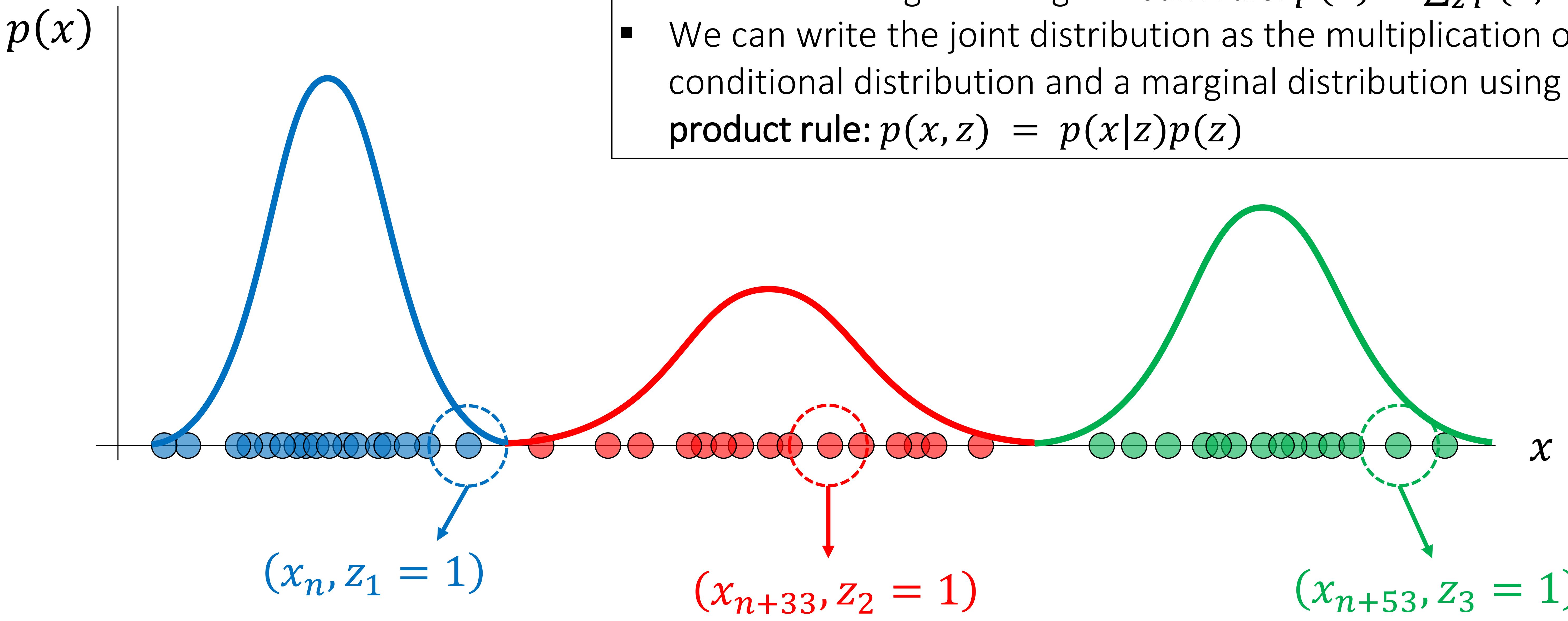
Focus video: GMM

Rodrigo Borela ▸ rborelav@gatech.edu

Defining our problem:

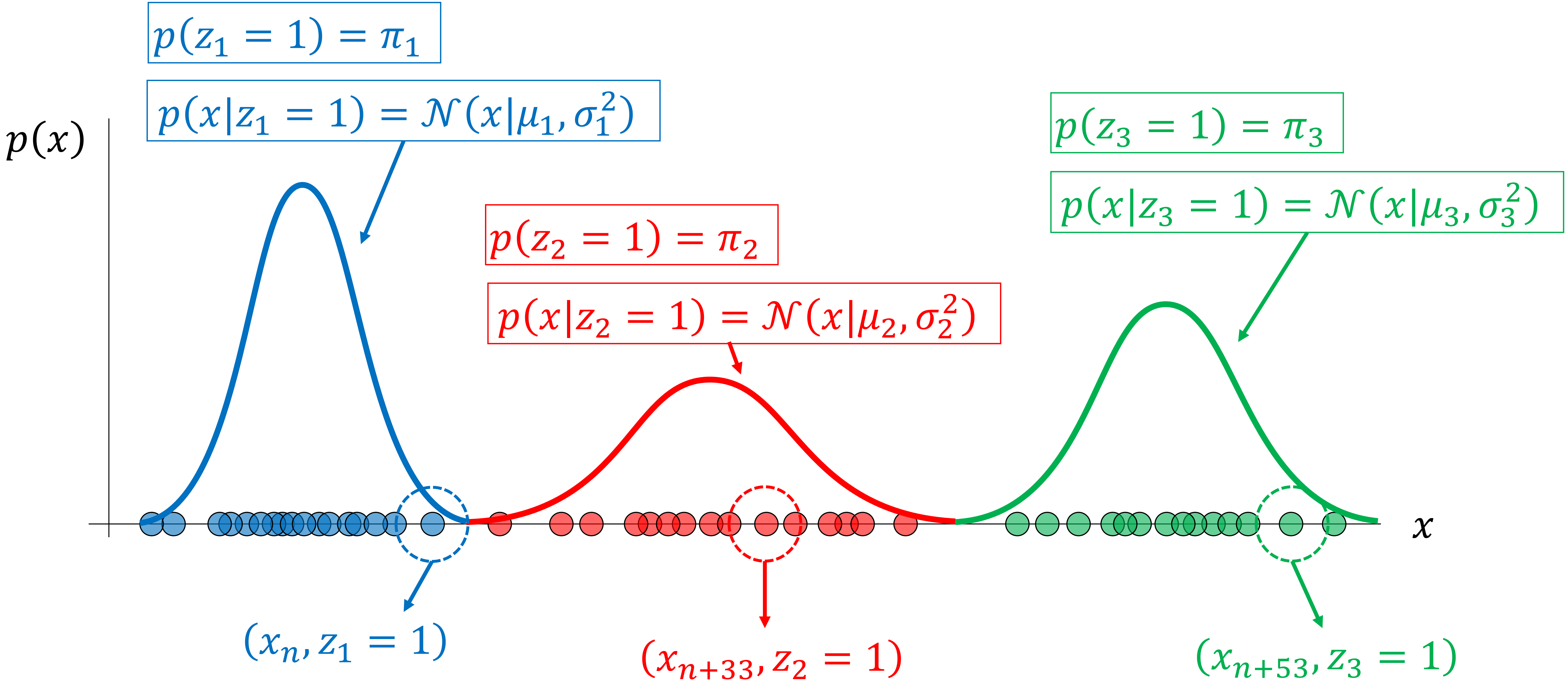
Obtain a marginal probability distribution for continuous variable x.

- If we have a joint distribution over variables x and z, we can obtain the marginal using the **sum rule:** $p(x) = \sum_z p(x, z)$.
- We can write the joint distribution as the multiplication of a conditional distribution and a marginal distribution using the **product rule:** $p(x, z) = p(x|z)p(z)$



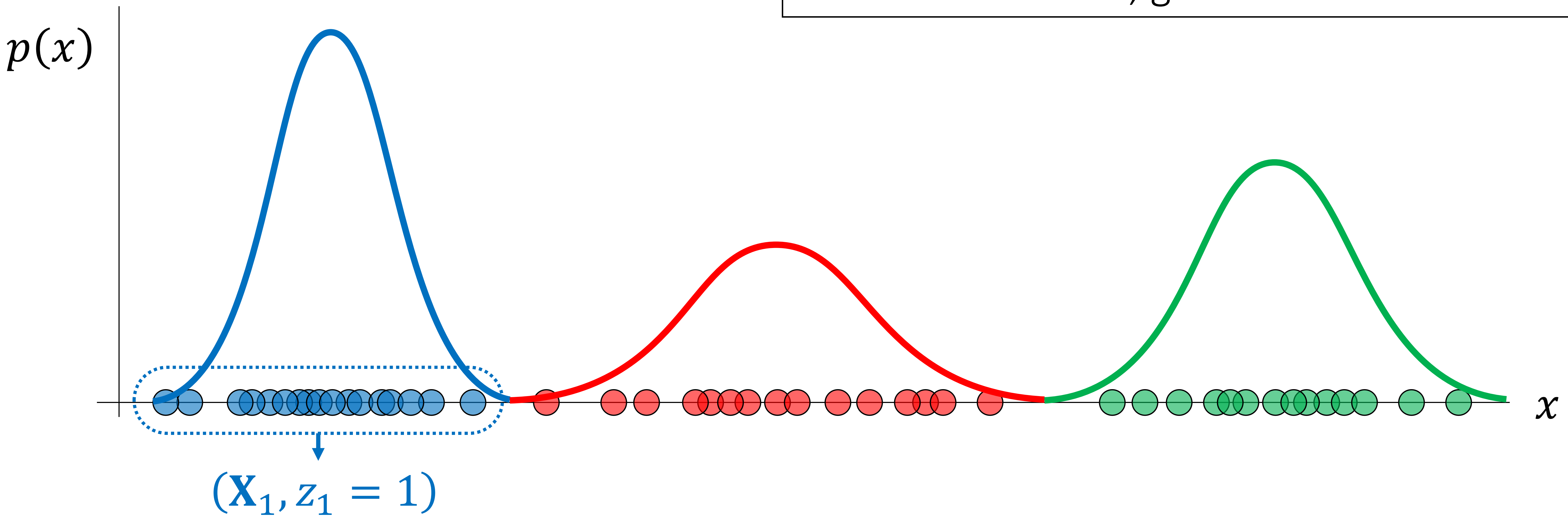
Defining our problem:

$$p(x) = \sum_z p(x|z)p(z) = \sum_k \mathcal{N}(x|\mu_k, \sigma_k^2)\pi_k$$



Defining our problem:

Obtain the parameters for the conditional distributions $p(x|z)$ by maximizing the log-likelihood function, given each value of z.

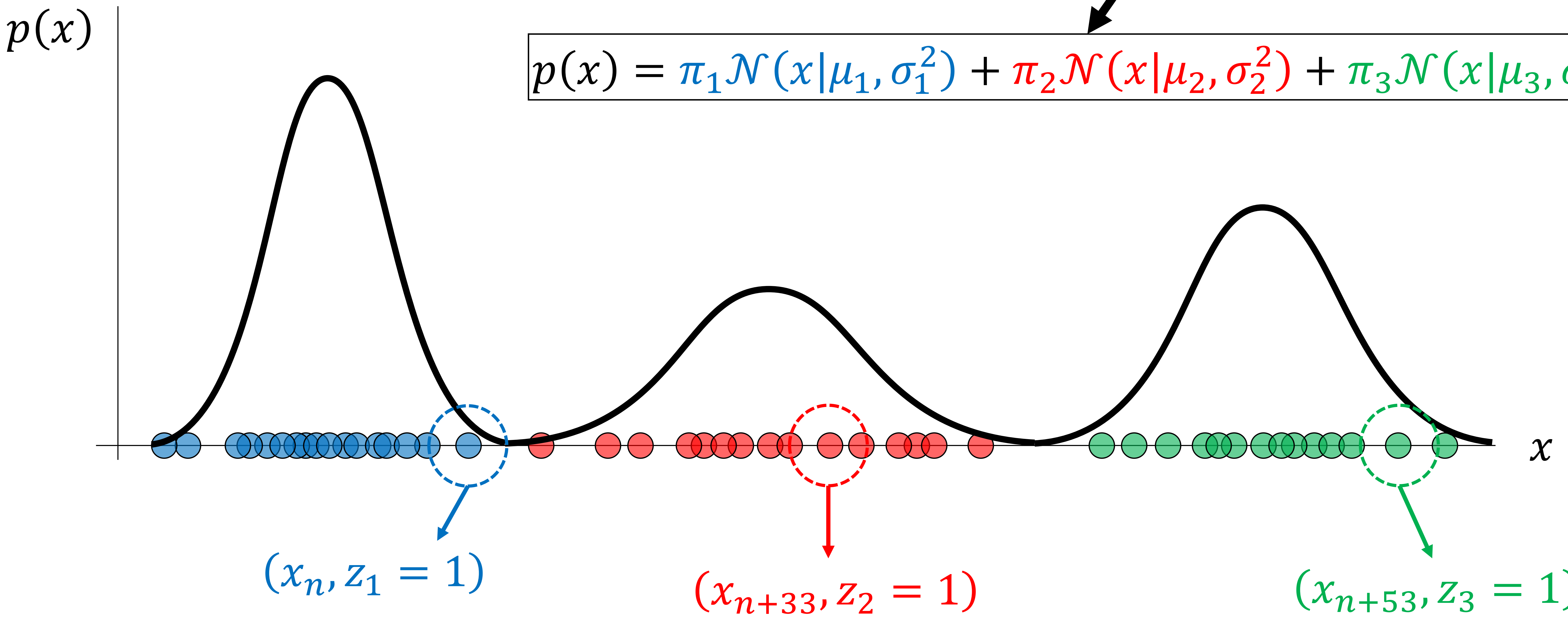


Likelihood function: $L(\mu_1, \sigma_1^2|\mathbf{X}_1) = \mathcal{N}(x_1|\mu_1, \sigma_1^2) \times \cdots \times \mathcal{N}\left(x_{N_1}|\mu_1, \sigma_1^2\right) = \prod_{n=1}^{N_1} \mathcal{N}(x_\mathrm{n}|\mu_1, \sigma_1^2)$

Log-likelihood function: $ll(\mu_1, \sigma_1^2|\mathbf{X}_1) = \sum_{n=1}^{N_1} \ln \mathcal{N}(x_\mathrm{n}|\mu_1, \sigma_1^2)$

Defining our problem: $p(x) = \sum_z p(x|z)p(z) = \sum_k \mathcal{N}(x|\mu_k, \sigma_k^2)\pi_k$

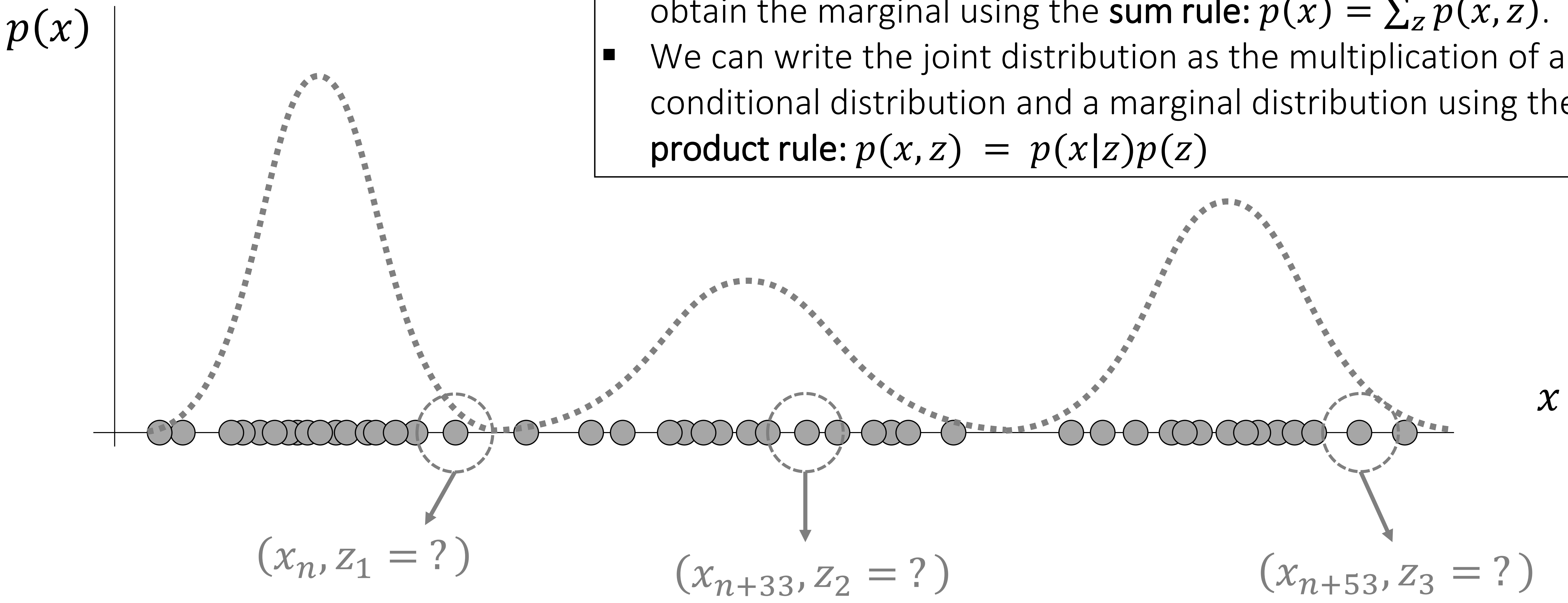
$p(x) = \pi_1 \mathcal{N}(x|\mu_1, \sigma_1^2) + \pi_2 \mathcal{N}(x|\mu_2, \sigma_2^2) + \pi_3 \mathcal{N}(x|\mu_3, \sigma_3^2)$



Defining our problem:

Obtain a marginal probability distribution for continuous variable x, when **we don't know z**

- Consider z a latent variable (unobserved)
- If we have a joint distribution over variables x and z, we can obtain the marginal using the **sum rule:** $p(x) = \sum_z p(x,z)$.
- We can write the joint distribution as the multiplication of a conditional distribution and a marginal distribution using the **product rule:** $p(x,z) = p(x|z)p(z)$



GMM: latent variable representation

- Model the probability density distribution of a random variable $\mathbf{x}$ as a mixture of Gaussians

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

- Introducing the binary random variable $\mathbf{z}$ that indicates a possible component from which a data point $\mathbf{x}$ was generated. $\mathbf{z}$ will have a 1-of-K encoding such that $z_k \in \{0,1\}$ and $\sum_k z_k = 1$.

$$p(z_k = 1) = \pi_k, \qquad 0 \le \pi_k \le 1, \qquad \sum_{k=1}^{K} \pi_k = 1$$

$$p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$$

GMM: latent variable representation

- Define a joint distribution $p(\mathbf{x}, \mathbf{z})$ in terms of random variables $\mathbf{x}$ and $\mathbf{z}$
- We can write a joint distribution using the product rule:

$$p(\mathbf{x}, \mathbf{z}) = p(\mathbf{x}|\mathbf{z})p(\mathbf{z})$$

- And the conditional distribution of $\mathbf{x}$ given a value of $\mathbf{z}$ (i.e. the conditional probability of $\mathbf{x}$ given that it was generated by component k)

$$p(\mathbf{x}|z_k = 1) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

$$p(\mathbf{x}|\mathbf{z}) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_k}$$

GMM: latent variable representation

- Using the sum rule, we can marginalize variable $\mathbf{z}$ (i.e. sum over the possible states of $\mathbf{z}$) to obtain $p(\mathbf{x})$

$$p(\mathbf{x}) = \sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{z}) = \sum_{\mathbf{z}} p(\mathbf{x}|\mathbf{z}) p(\mathbf{z}) = \sum_{\mathbf{z}} \left\{ \prod_{k=1}^{K} \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_k} \prod_{k=1}^{K} \pi_k^{z_k} \right\}$$

$$p(\mathbf{x}) = \sum_{\mathbf{z}} \left\{ \prod_{k=1}^{K} \pi_k^{z_k} \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_k} \right\} = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

GMM: latent variable representation

- Using the Bayes rule, we can compute the conditional probability of $\mathbf{z}$ given $\mathbf{x}$, i.e. responsibility that component k explains the observation $\mathbf{x}$

$$\gamma(z_k) = p(z_k = 1|\mathbf{x}) = \frac{\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}\left(\mathbf{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right)}$$

MLE for GMM

- Probability density function for GMM

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

- Likelihood function of the dataset $\mathbf{X}$

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

- Log-likelihood function of the dataset $\mathbf{X}$

$$ll(\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{X}) = \ln L(\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{X}) = \ln \prod_{n=1}^{N} \sum_{k=1}^{K} \pi_k N(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\}$$

MLE for GMM

- Mean for each component, $\boldsymbol{\mu}_k$

$$\frac{\partial}{\partial \boldsymbol{\mu}_k} ll(\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{X}) = 0 \rightarrow \boldsymbol{\mu}_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})\mathbf{x}_n}{\sum_{n=1}^{N} \gamma(z_{nk})}$$

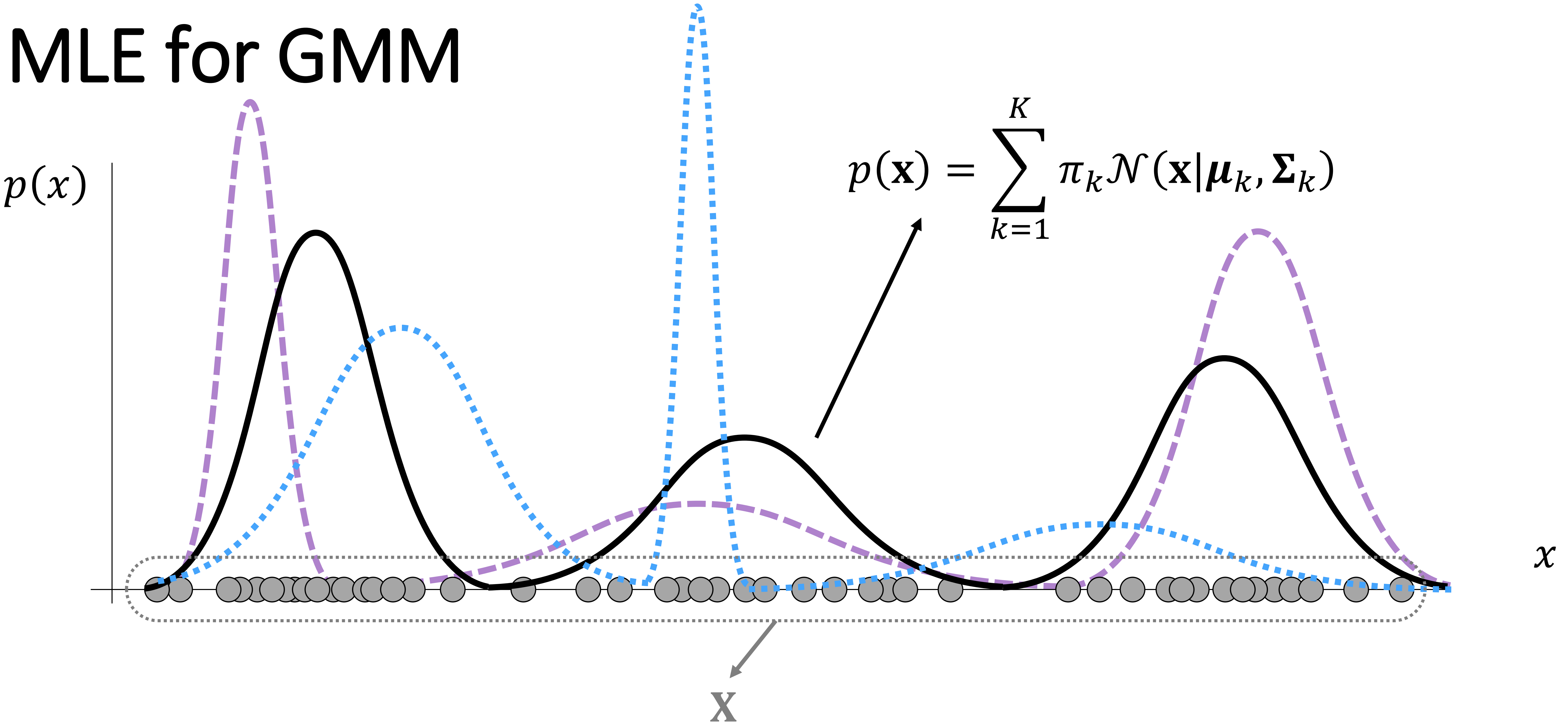
- Covariance for each component, $\boldsymbol{\Sigma}_k$

$$\frac{\partial}{\partial \boldsymbol{\Sigma}_k} ll(\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{X}) = 0 \rightarrow \boldsymbol{\Sigma}_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})(\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^T}{\sum_{n=1}^{N} \gamma(z_{nk})}$$

- Mixing component, π_k

$$\frac{\partial}{\partial \pi_k} ll(\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{X}) = 0 \rightarrow \pi_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})}{N}$$

MLE for GMM



$$L(\pi, \mu, \sigma^2|\mathbf{X}) = \prod_{n=1}^{N} \sum_{k=1}^{K} \pi_k \mathcal{N}(x_{\mathrm{n}}|\mu_k, \sigma_k^2)$$

EM-algorithm for GMM

- **E-step:** evaluate the responsibilities using current parameters

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

- **M-step**: re-estimate the parameters using the current responsibilities

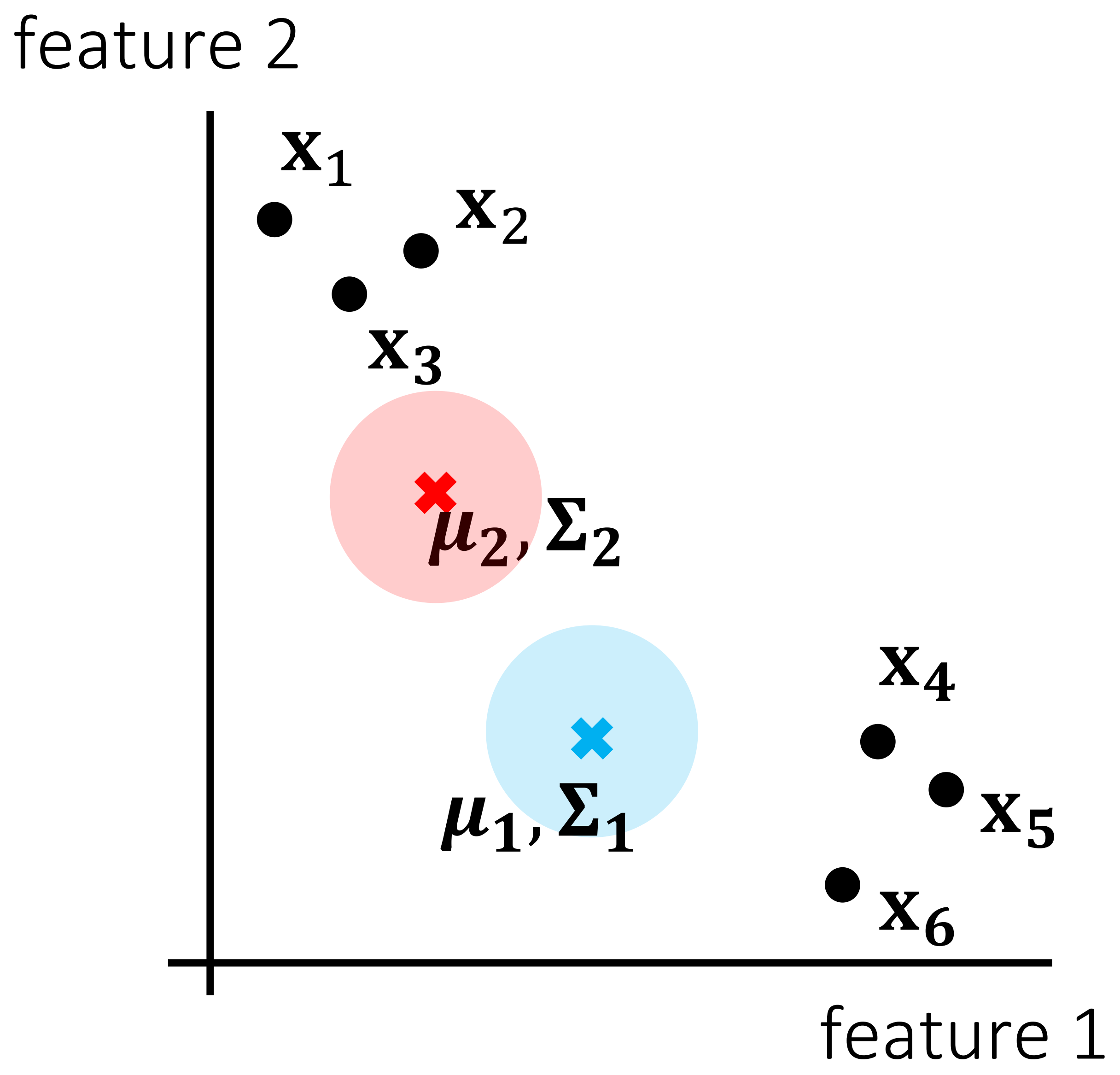
$$\boldsymbol{\mu}_k^{new} = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n}{N_k}$$

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

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

Where $N_k = \sum_{n=1}^{N} \gamma(z_{nk})$

EM-algorithm for GMM



Dataset: $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \mathbf{x}_3^T \\ \mathbf{x}_4^T \\ \mathbf{x}_5^T \\ \mathbf{x}_6^T \end{bmatrix}_{N \times D} = \begin{bmatrix} 1.0 & 8.0 \\ 2.5 & 7.5 \\ 2.0 & 7.0 \\ 8.5 & 2.5 \\ 9.0 & 2.0 \\ 8.0 & 1.0 \end{bmatrix}_{N \times D = 6 \times 2}$

<u>Parameter initialization</u>

Mixing components: $\boldsymbol{\pi} = \begin{bmatrix} 0.60 \\ 0.40 \end{bmatrix}_{1 \times K}$

Cluster centers: $\mathbf{M} = \begin{bmatrix} \boldsymbol{\mu}_1^T \\ \boldsymbol{\mu}_2^T \end{bmatrix} = \begin{bmatrix} 4.5 & 2.5 \\ 2.5 & 5.0 \end{bmatrix}_{K \times D = 2 \times 2}$

Covariance matrices: $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{D \times D = 2 \times 2}$

E-step: evaluate responsibilities

- For each datapoint $\mathbf{x}_n$, evaluate $\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{x}_\mathrm{n}|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_\mathrm{n}|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j)}$

$$\mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k) = \frac{1}{(2\pi)^{\frac{D}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}_n - \boldsymbol{\mu}_k)^\mathrm{T}\boldsymbol{\Sigma}^{-\mathbf{1}}(\mathbf{x}_n - \boldsymbol{\mu}_k)\right\}$$

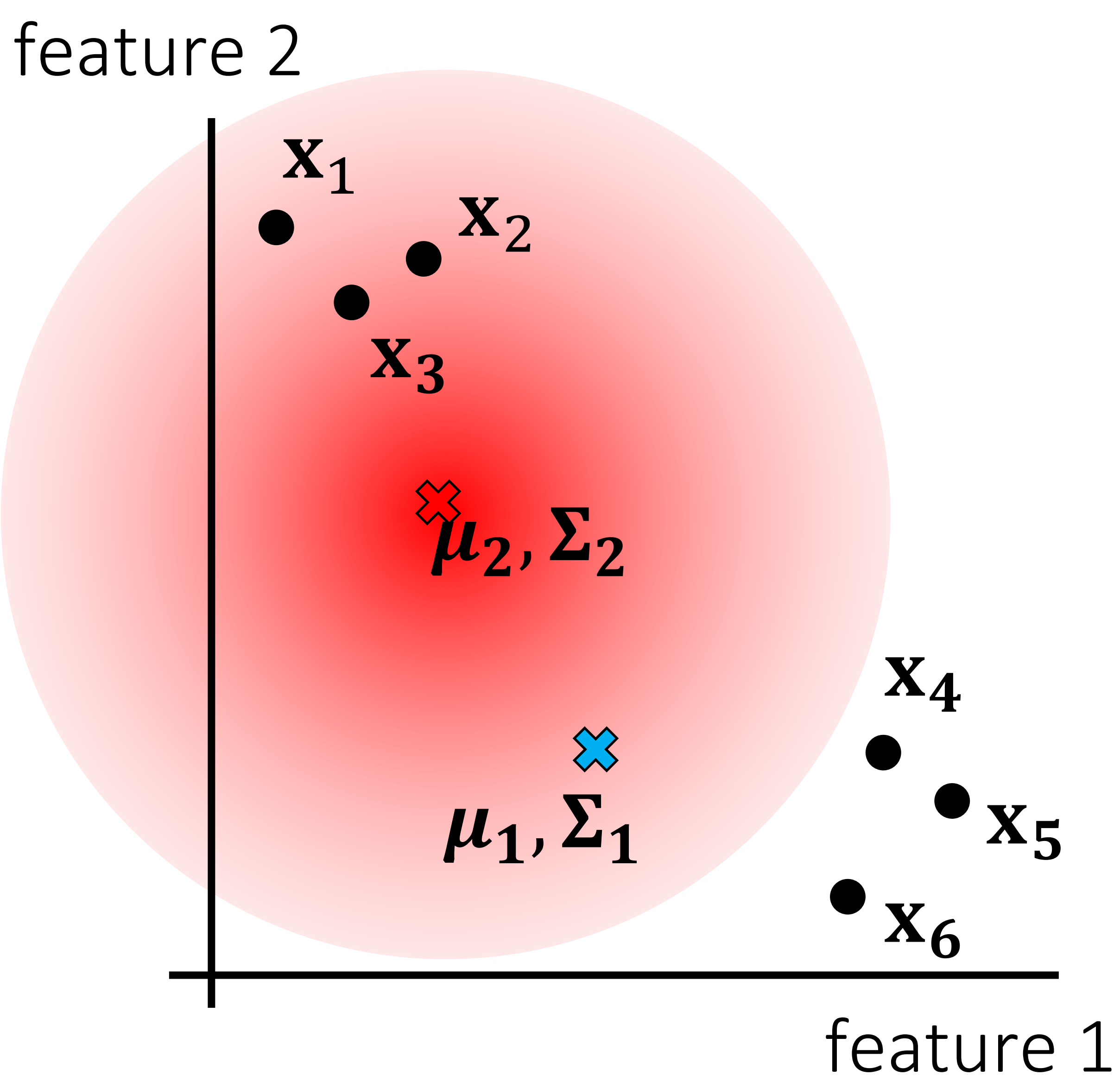
- Let's consider $\mathbf{x}_1^T = [1.0 \quad 8.0]$
  - For $\mathbf{z}^T = [1 \quad 0]$ (component $k = 1$)

$$\mathcal{N}(\mathbf{x}_1|\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1) = \frac{1}{2\pi \det\left(\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}[1.0-4.5 \quad 8.0-2.5]\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}1.0-4.5\\8.0-2.5\end{bmatrix}\right\} = 9.40 \times 10^{-11}$$

  - For $\mathbf{z}^T = [0 \quad 1]$ (component $k = 2$)

$$\mathcal{N}(\mathbf{x}_1|\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2) = \frac{1}{2\pi \det\left(\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}[1.0-2.5 \quad 8.0-5.0]\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}1.0-2.5\\8.0-5.0\end{bmatrix}\right\} = 5.74 \times 10^{-4}$$

EM-algorithm for GMM



Dataset: $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \mathbf{x}_3^T \\ \mathbf{x}_4^T \\ \mathbf{x}_5^T \\ \mathbf{x}_6^T \end{bmatrix}_{N \times D} = \begin{bmatrix} 1.0 & 8.0 \\ 2.5 & 7.5 \\ 2.0 & 7.0 \\ 8.5 & 2.5 \\ 9.0 & 2.0 \\ 8.0 & 1.0 \end{bmatrix}_{N \times D = 6 \times 2}$

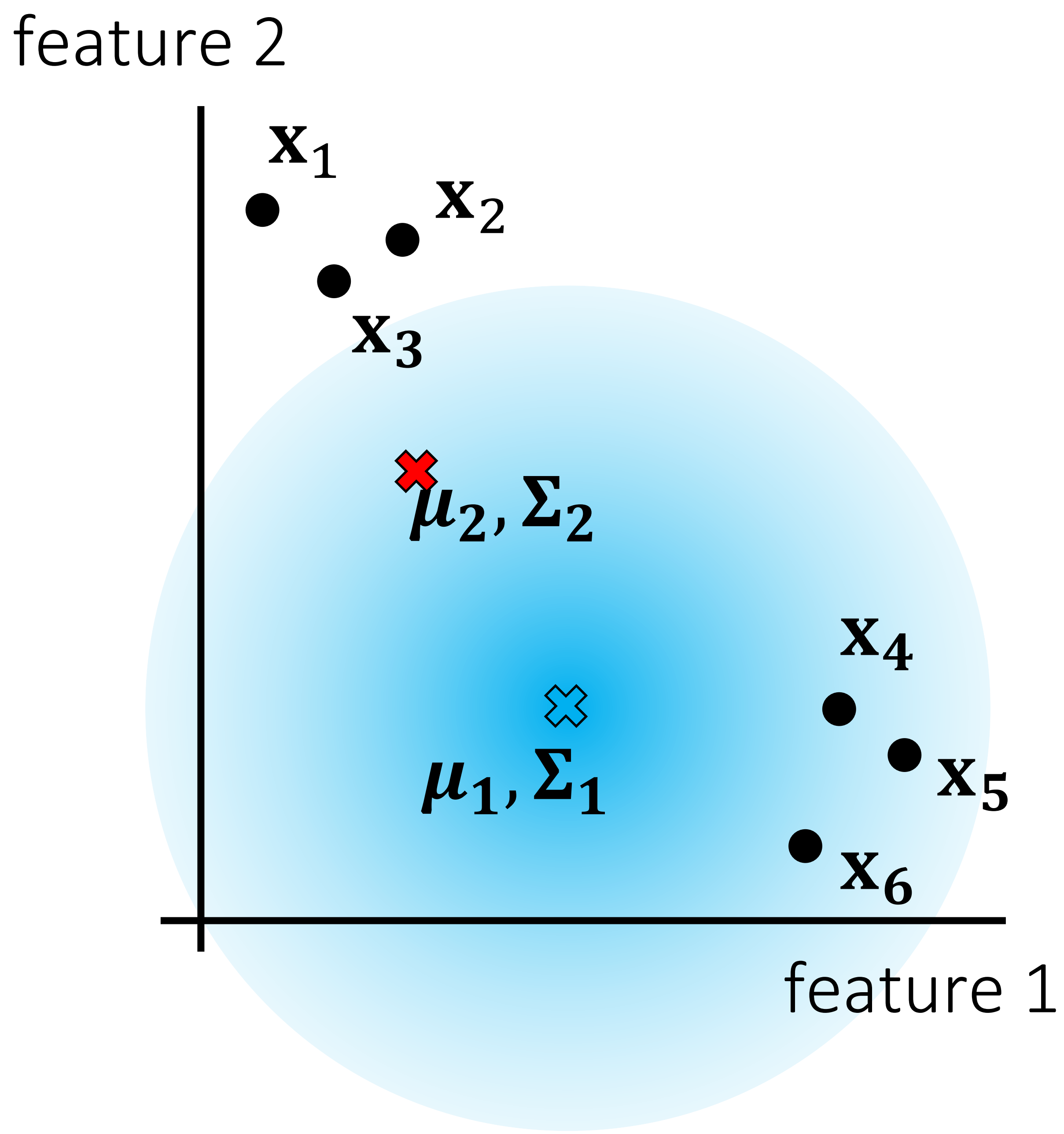
<u>Parameter initialization</u>

Mixing components: $\boldsymbol{\pi} = \begin{bmatrix} 0.60 \\ 0.40 \end{bmatrix}_{1 \times K}$

Cluster centers: $\mathbf{M} = \begin{bmatrix} \boldsymbol{\mu}_1^T \\ \boldsymbol{\mu}_2^T \end{bmatrix} = \begin{bmatrix} 4.5 & 2.5 \\ 2.5 & 5.0 \end{bmatrix}_{K \times D = 2 \times 2}$

Covariance matrices: $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{D \times D = 2 \times 2}$

EM-algorithm for GMM



Dataset: $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \mathbf{x}_3^T \\ \mathbf{x}_4^T \\ \mathbf{x}_5^T \\ \mathbf{x}_6^T \end{bmatrix}_{N \times D} = \begin{bmatrix} 1.0 & 8.0 \\ 2.5 & 7.5 \\ 2.0 & 7.0 \\ 8.5 & 2.5 \\ 9.0 & 2.0 \\ 8.0 & 1.0 \end{bmatrix}_{N \times D = 6 \times 2}$

<u>Parameter initialization</u>

Mixing components: $\boldsymbol{\pi} = \begin{bmatrix} 0.60 \\ 0.40 \end{bmatrix}_{1 \times K}$

Cluster centers: $\mathbf{M} = \begin{bmatrix} \boldsymbol{\mu}_1^T \\ \boldsymbol{\mu}_2^T \end{bmatrix} = \begin{bmatrix} 4.5 & 2.5 \\ 2.5 & 5.0 \end{bmatrix}_{K \times D = 2 \times 2}$

Covariance matrices: $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{D \times D = 2 \times 2}$

E-step: evaluate responsibilities

- For each datapoint $\mathbf{x}_n$, evaluate $\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{x}_\mathrm{n}|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_\mathrm{n}|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j)}$

$$\mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{(2\pi)^{\frac{D}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}_n - \boldsymbol{\mu}_k)^\mathrm{T}\boldsymbol{\Sigma}^{-\mathbf{1}}(\mathbf{x}_n - \boldsymbol{\mu}_k)\right\}$$

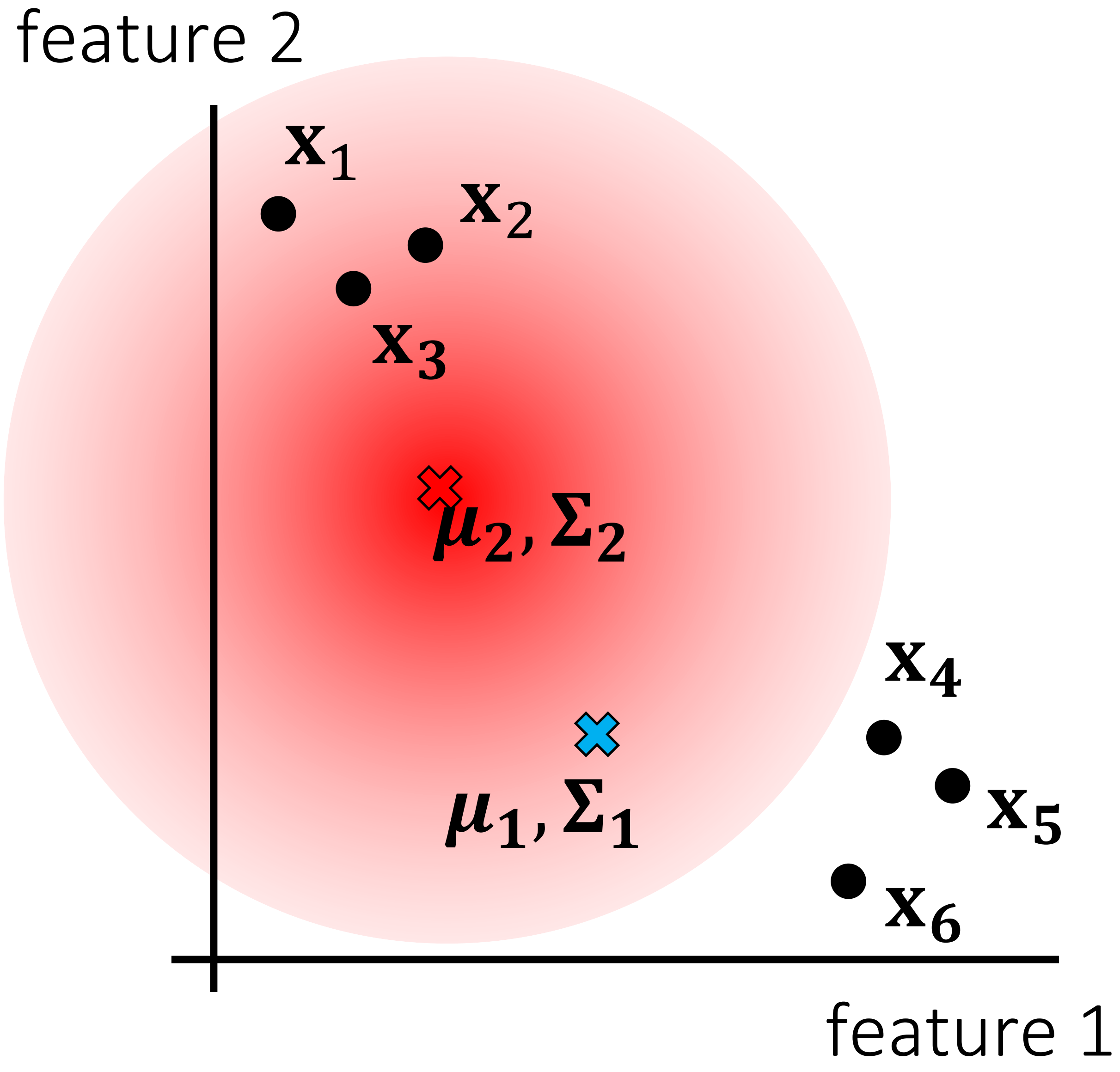
- Let’s consider $\mathbf{x}_5^T = [9.0 \quad 2.0]$
  - For $\mathbf{z}^T = [1 \quad 0]$ (component $k = 1$)

$$\mathcal{N}(\mathbf{x}_5|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) = \frac{1}{2\pi \det\left(\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}[9.0 - 4.5 \quad 2.0 - 2.5]\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}9.0 - 4.5\\2.0 - 2.5\end{bmatrix}\right\} = 5.63 \times 10^{-6}$$

  - For $\mathbf{z}^T = [0 \quad 1]$ (component $k = 2$)

$$\mathcal{N}(\mathbf{x}_5|\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) = \frac{1}{2\pi \det\left(\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}[9.0 - 2.5 \quad 2.0 - 5.0]\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}9.0 - 2.5\\2.0 - 5.0\end{bmatrix}\right\} = 1.18 \times 10^{-12}$$

EM-algorithm for GMM



Dataset: $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \mathbf{x}_3^T \\ \mathbf{x}_4^T \\ \mathbf{x}_5^T \\ \mathbf{x}_6^T \end{bmatrix}_{N \times D} = \begin{bmatrix} 1.0 & 8.0 \\ 2.5 & 7.5 \\ 2.0 & 7.0 \\ 8.5 & 2.5 \\ 9.0 & 2.0 \\ 8.0 & 1.0 \end{bmatrix}_{N \times D = 6 \times 2}$

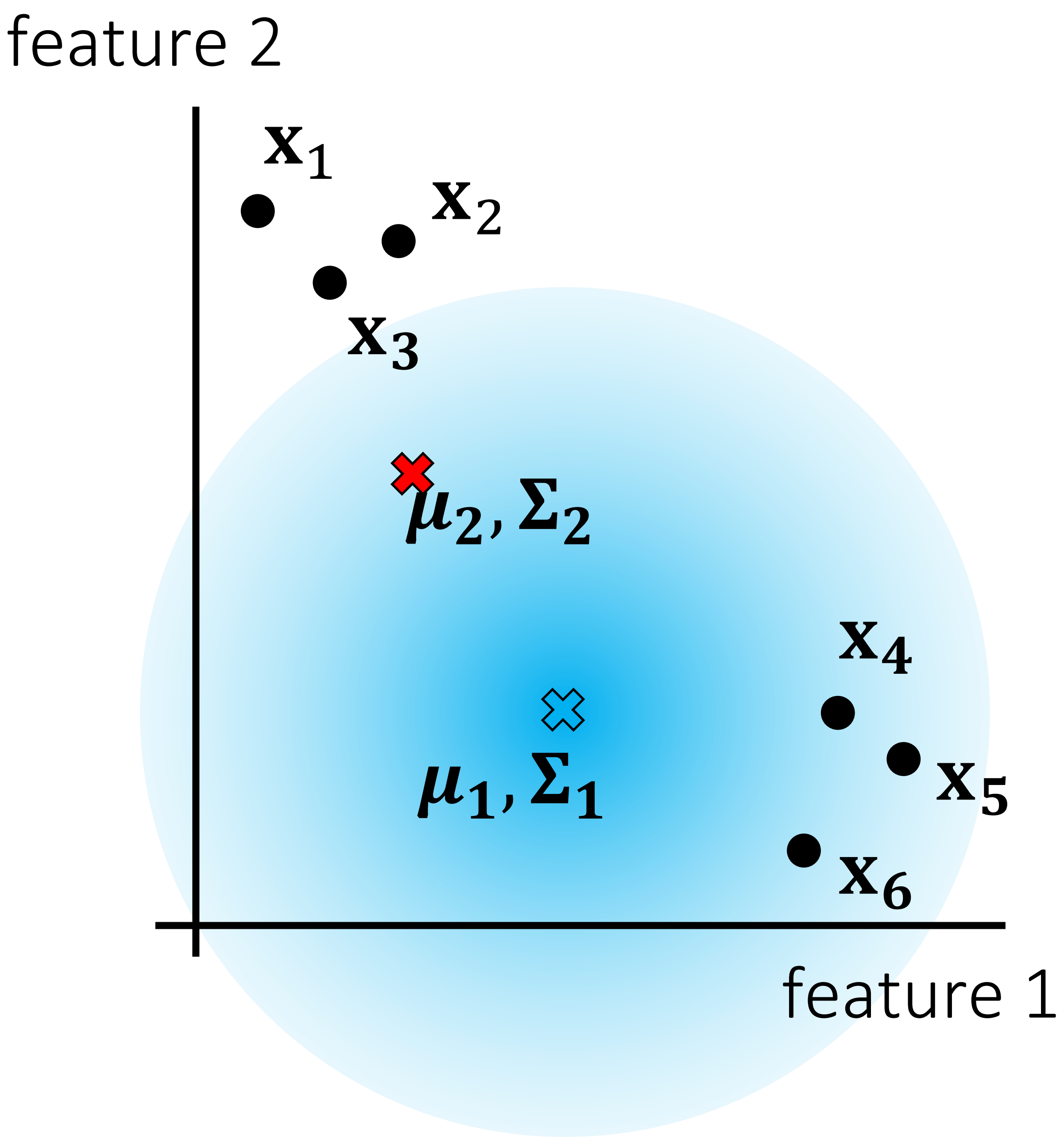
<u>Parameter initialization</u>

Mixing components: $\boldsymbol{\pi} = \begin{bmatrix} 0.60 \\ 0.40 \end{bmatrix}_{1 \times K}$

Cluster centers: $\mathbf{M} = \begin{bmatrix} \boldsymbol{\mu}_1^T \\ \boldsymbol{\mu}_2^T \end{bmatrix} = \begin{bmatrix} 4.5 & 2.5 \\ 2.5 & 5.0 \end{bmatrix}_{K \times D = 2 \times 2}$

Covariance matrices: $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{D \times D = 2 \times 2}$

EM-algorithm for GMM



Dataset: $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \mathbf{x}_3^T \\ \mathbf{x}_4^T \\ \mathbf{x}_5^T \\ \mathbf{x}_6^T \end{bmatrix}_{N \times D} = \begin{bmatrix} 1.0 & 8.0 \\ 2.5 & 7.5 \\ 2.0 & 7.0 \\ 8.5 & 2.5 \\ 9.0 & 2.0 \\ 8.0 & 1.0 \end{bmatrix}_{N \times D = 6 \times 2}$

Parameter initialization

Mixing components: $\boldsymbol{\pi} = \begin{bmatrix} 0.60 \\ 0.40 \end{bmatrix}_{1 \times K}$

Cluster centers: $\mathbf{M} = \begin{bmatrix} \boldsymbol{\mu}_1^T \\ \boldsymbol{\mu}_2^T \end{bmatrix} = \begin{bmatrix} 4.5 & 2.5 \\ 2.5 & 5.0 \end{bmatrix}_{K \times D = 2 \times 2}$

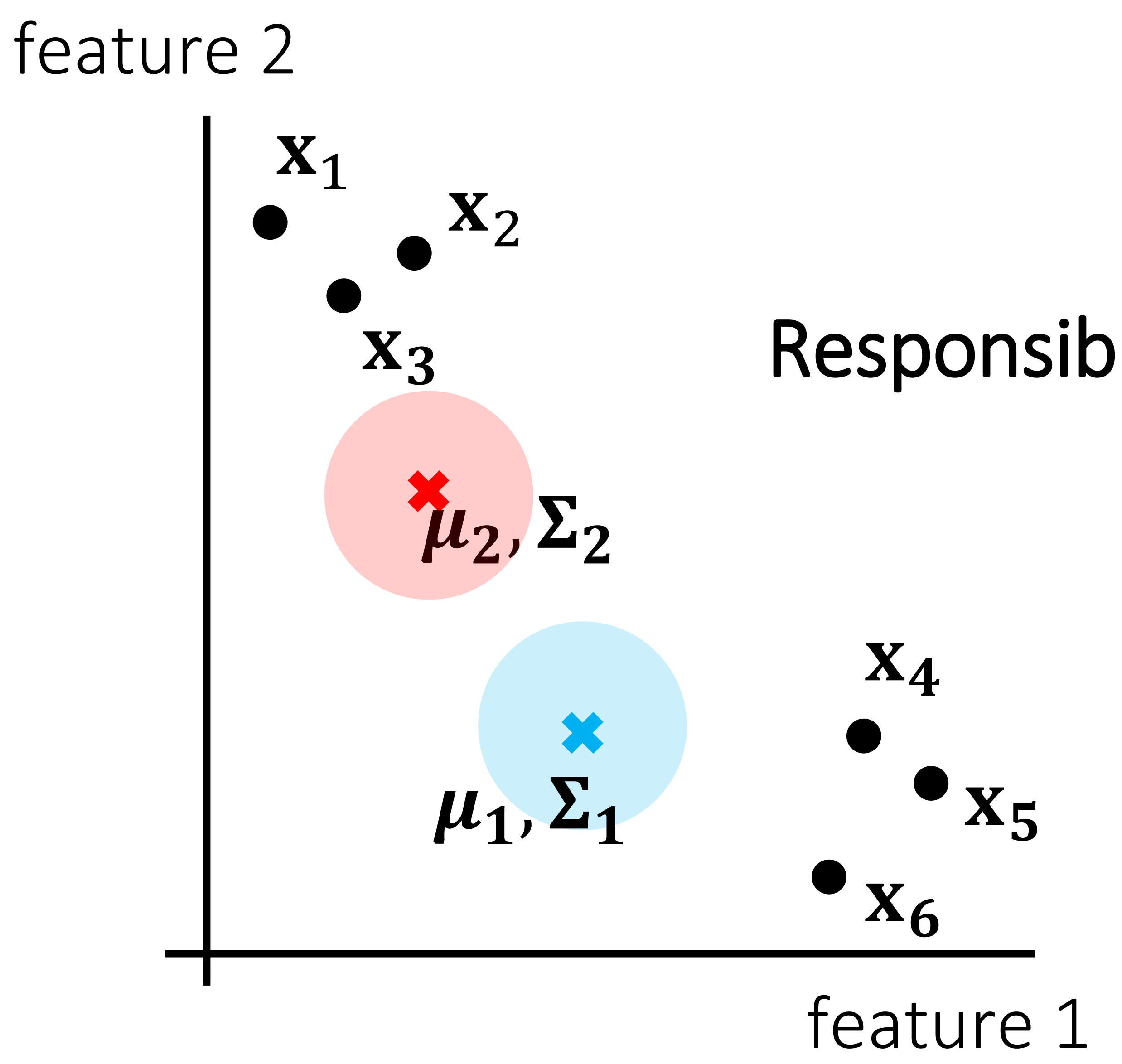
Covariance matrices: $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{D \times D = 2 \times 2}$

E-step: evaluate responsibilities

- For $\mathbf{x}_1^T = [1.0 \quad 8.0]$, $\mathcal{N}(\mathbf{x_1}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) = 9.4 \times 10^{-11}$, $\mathcal{N}(\mathbf{x_1}|\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) = 5.74 \times 10^{-4}$
  - Component $k = 1$: $\gamma(z_{11}) = \frac{0.60 \times 9.4 \times 10^{-11}}{(0.60 \times 9.4 \times 10^{-11}) + (0.40 \times 5.74 \times 10^{-4})} = 2.46 \times 10^{-7}$
  - Component $k = 2$: $\gamma(z_{12}) = \frac{0.40 \times 5.74 \times 10^{-4}}{(0.60 \times 9.4 \times 10^{-11}) + (0.40 \times 5.74 \times 10^{-4})} = {\sim}1$

- For $\mathbf{x}_5^T = [9.0 \quad 2.0]$, $\mathcal{N}(\mathbf{x}_5|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) = 5.63 \times 10^{-6}$, $\mathcal{N}(\mathbf{x}_5|\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) = 1.18 \times 10^{-12}$
  - Component $k = 1$: $\gamma(z_{51}) = \frac{0.60 \times 5.63 \times 10^{-6}}{(0.60 \times 5.63 \times 10^{-6}) + (0.40 \times 1.18 \times 10^{-12})} = {\sim}1$
  - Component $k = 2$: $\gamma(z_{52}) = \frac{0.40 \times 1.18 \times 10^{-12}}{(0.60 \times 5.63 \times 10^{-6}) + (0.40 \times 1.18 \times 10^{-12})} = 1.40 \times 10^{-7}$

E-step: evaluate responsibilities



Responsibilities: $\boldsymbol{\Gamma} = \begin{bmatrix} 2.46 \times 10^{-7} & \sim 1 \\ 1.72 \times 10^{-5} & \sim 1 \\ 0.003 & 0.997 \\ \sim 1 & 1.33 \times 10^{-6} \\ \sim 1 & 1.40 \times 10^{-7} \\ \sim 1 & 8.50 \times 10^{-8} \end{bmatrix}_{N \times K}$

M-step: reevaluate parameters

$$N_k = \sum_{n=1}^{N} \gamma(z_{nk})$$

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

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

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

$$\mathbf{X} = \begin{bmatrix} 1.0 & 8.0 \\ 2.5 & 7.5 \\ 2.0 & 7.0 \\ 8.5 & 2.5 \\ 9.0 & 2.0 \\ 8.0 & 1.0 \end{bmatrix},$$

$$\boldsymbol{\Gamma} = \begin{bmatrix} 2.46 \times 10^{-7} & \sim 1 \\ 1.72 \times 10^{-5} & \sim 1 \\ 0.003 & 0.997 \\ \sim 1 & 1.33 \times 10^{-6} \\ \sim 1 & 1.40 \times 10^{-7} \\ \sim 1 & 8.50 \times 10^{-8} \end{bmatrix}$$

M-step: reevaluate parameters

- Component $k = 1$

$$N_1 = (2.46 \times 10^{-7} + 1.72 \times 10^{-5} + 0.003 + 1 + 1 + 1) \cong 3$$

$$\boldsymbol{\mu}_1^{new} = \frac{\sum_{n=1}^{N} \gamma(z_{n1}) \mathbf{x}_n}{N_1} = \begin{bmatrix} 8.49 \\ 1.84 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_1^{new} = \frac{\sum_{n=1}^{N} \gamma(z_{n1}) (\mathbf{x}_n - \boldsymbol{\mu}_1^{new})(\mathbf{x}_n - \boldsymbol{\mu}_1^{new})^T}{N_1} = \begin{bmatrix} 0.21 & 0.14 \\ 0.14 & 0.40 \end{bmatrix}$$

$$\pi_1^{new} = \frac{N_1}{N} = \frac{3}{6} = 0.5$$

M-step: reevaluate parameters

- Component $k = 2$

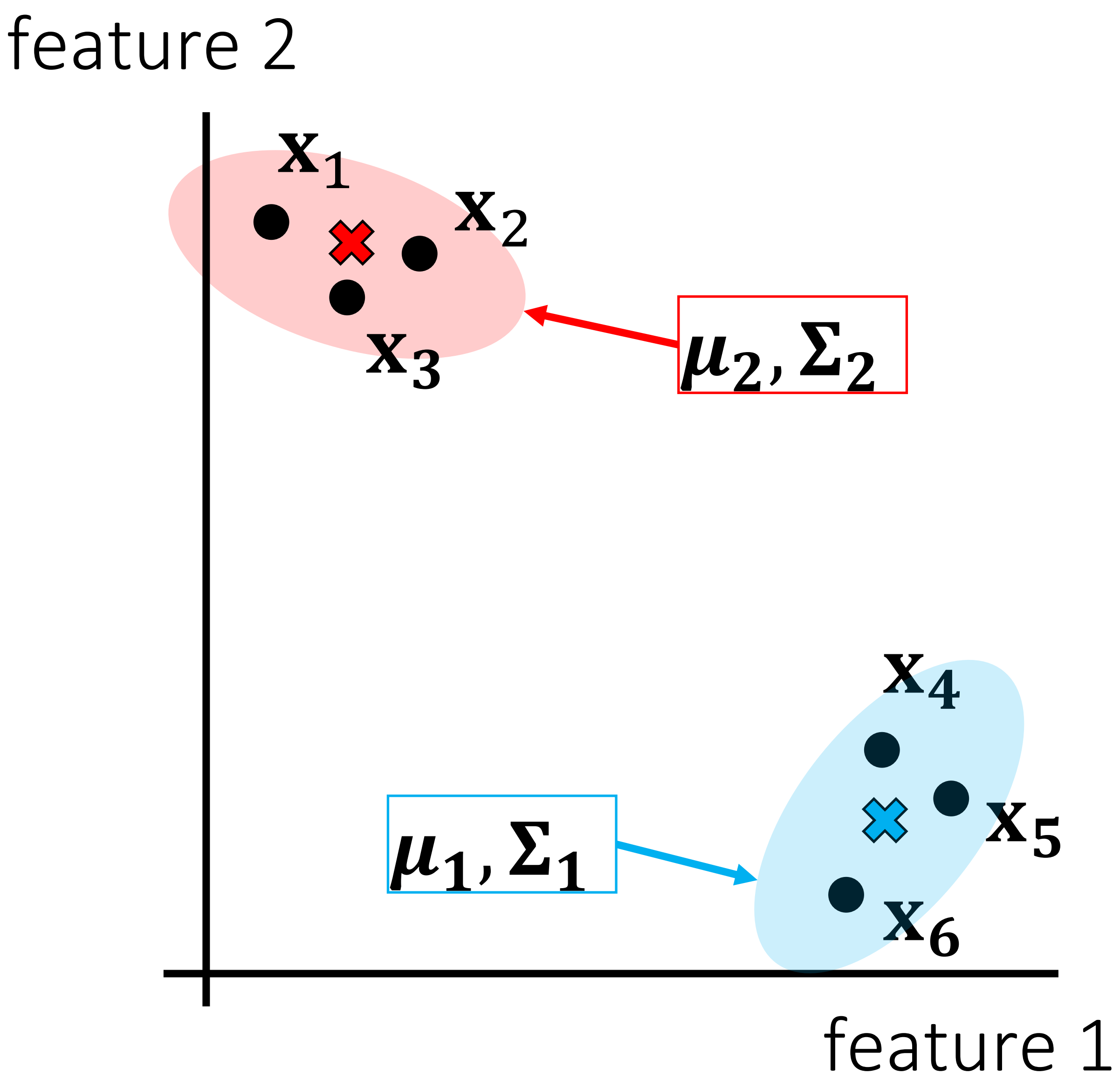
$$N_2 = (1 + 1 + 0.993 + 1.33 \times 10^{-6} + 1.40 \times 10^{-7} + 8.50 \times 10^{-8}) \cong 3$$

$$\boldsymbol{\mu}_2^{new} = \frac{\sum_{n=1}^{N} \gamma(z_{n2}) \mathbf{x}_n}{N_2} = \begin{bmatrix} 1.83 \\ 6.84 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_2^{new} = \frac{\sum_{n=1}^{N} \gamma(z_{n2}) (\mathbf{x}_n - \boldsymbol{\mu}_2^{new})(\mathbf{x}_n - \boldsymbol{\mu}_2^{new})^T}{N_2} = \begin{bmatrix} 0.39 & -0.28 \\ -0.28 & 0.17 \end{bmatrix}$$

$$\pi_2^{new} = \frac{N_2}{N} = \frac{3}{6} = 0.5$$

EM-algorithm for GMM



<u>After one iteration</u>

Mixing components: $\boldsymbol{\pi} = \begin{bmatrix} 0.50 \\ 0.50 \end{bmatrix}_{1 \times K}$

Cluster centers: $\mathbf{M} = \begin{bmatrix} \boldsymbol{\mu}_1^T \\ \boldsymbol{\mu}_2^T \end{bmatrix} = \begin{bmatrix} 8.49 & 1.84 \\ 1.83 & 6.84 \end{bmatrix}_{K \times D = 2 \times 2}$

Covariance matrices: $\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.21 & 0.14 \\ 0.14 & 0.40 \end{bmatrix}$

$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 0.39 & -0.28 \\ -0.28 & 0.17 \end{bmatrix}$$