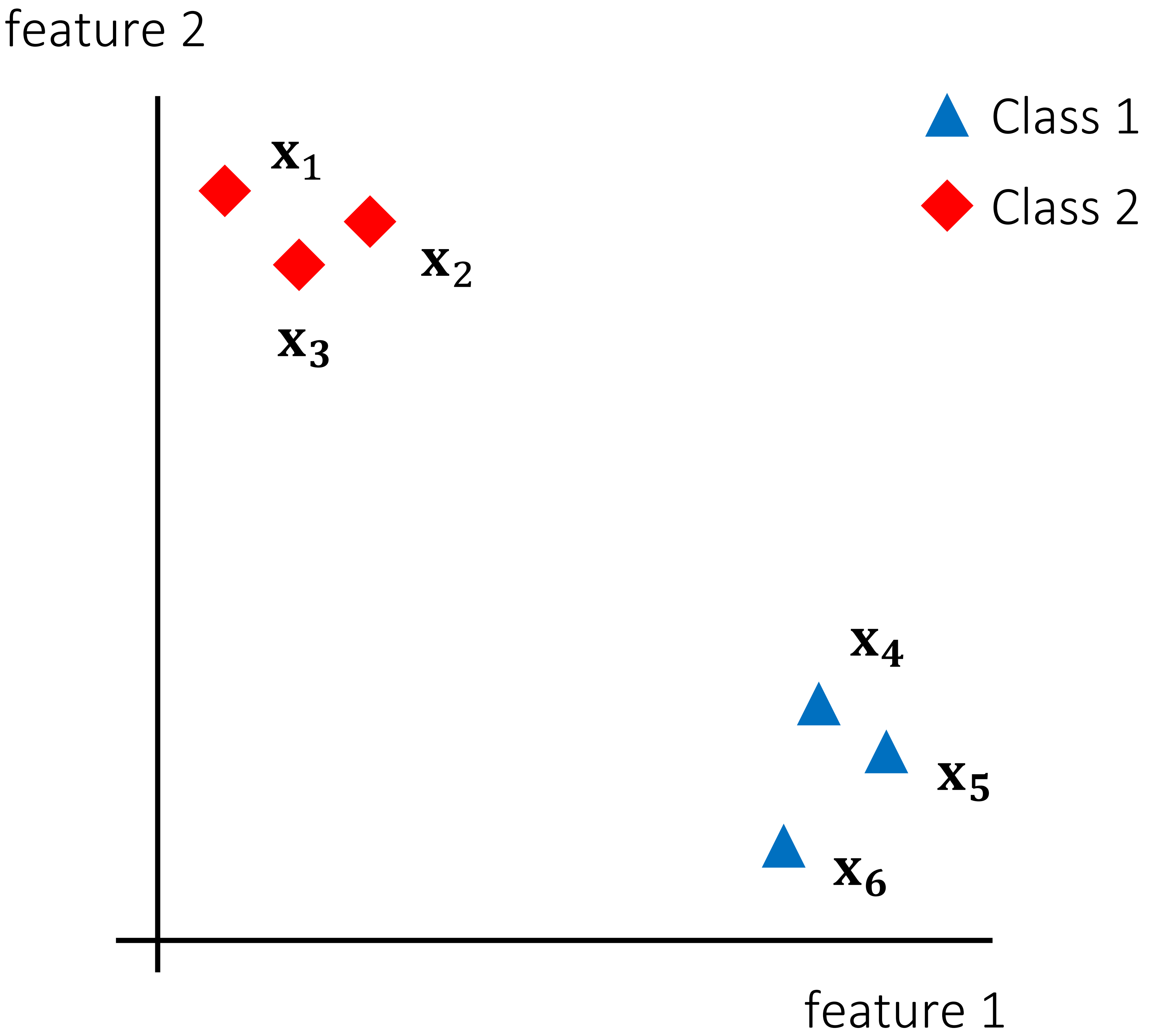
CS4641B Machine Learning

Focus video: Gaussian Naïve Bayes

Rodrigo Borela ▸ rborelav@gatech.edu

Gaussian Naïve Bayes classifier

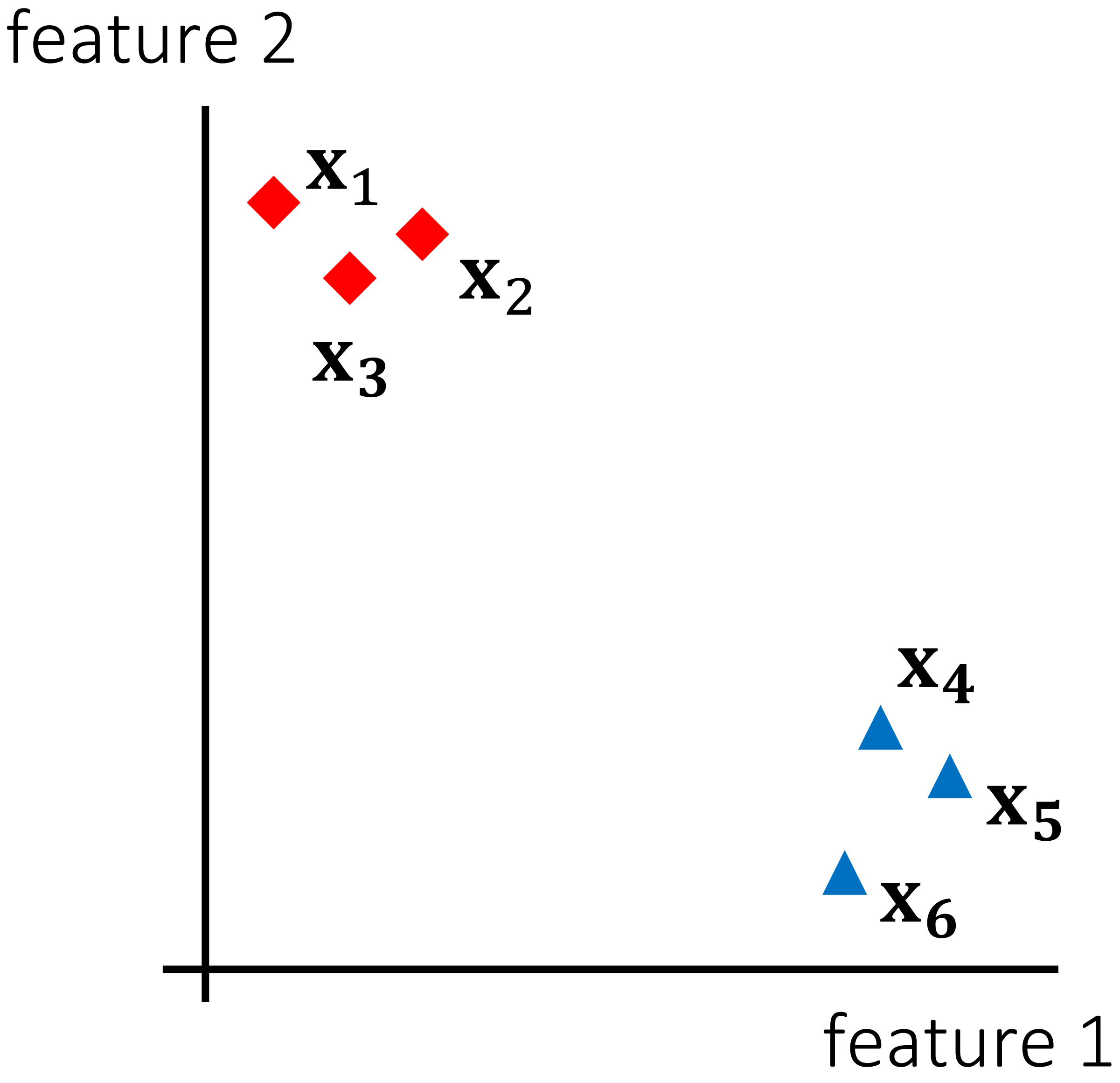


Inference

- Using the Bayes rule, we can compute the conditional probability of $\mathbf{z}$ given $\mathbf{x}$, i.e. probability that a point $\mathbf{x}$ belongs to a class k

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

Training



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

Target values: $\mathbf{t} = \mathbf{z} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}_{N \times K}$

Training

- Mean for class k, $\boldsymbol{\mu}_k$

$$\boldsymbol{\mu}_k = \frac{\sum_{n=1}^{N} z_{nk}\mathbf{x}_n}{\sum_{n=1}^{N} z_{nk}}$$

- Covariance for class k, $\boldsymbol{\Sigma}_k$

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

(making the conditional independence assumption, we are to set the off-diagonal terms to zero)

- Priors, π_k

$$\pi_k = \frac{\sum_{n=1}^{N} z_{nk}}{N}$$

Estimate parameters

- Class $k = 1$

$$N_1 = (1 + 1 + 1) = 3$$

$$\boldsymbol{\mu}_1 = \frac{\sum_{n=1}^{N} z_{n1}\mathbf{x}_n}{N_1} = \begin{bmatrix} 8.50 \\ 1.83 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_1 = \frac{\sum_{n=1}^{N} z_{n1}(\mathbf{x}_n - \boldsymbol{\mu}_1)(\mathbf{x}_n - \boldsymbol{\mu}_1)^T}{N_1} = \begin{bmatrix} 0.21 & 0.14 \\ 0.14 & 0.40 \end{bmatrix} \xrightarrow{cond.\, independence} \begin{bmatrix} 0.21 & 0 \\ 0 & 0.40 \end{bmatrix}$$

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

Training

- Class $k = 2$

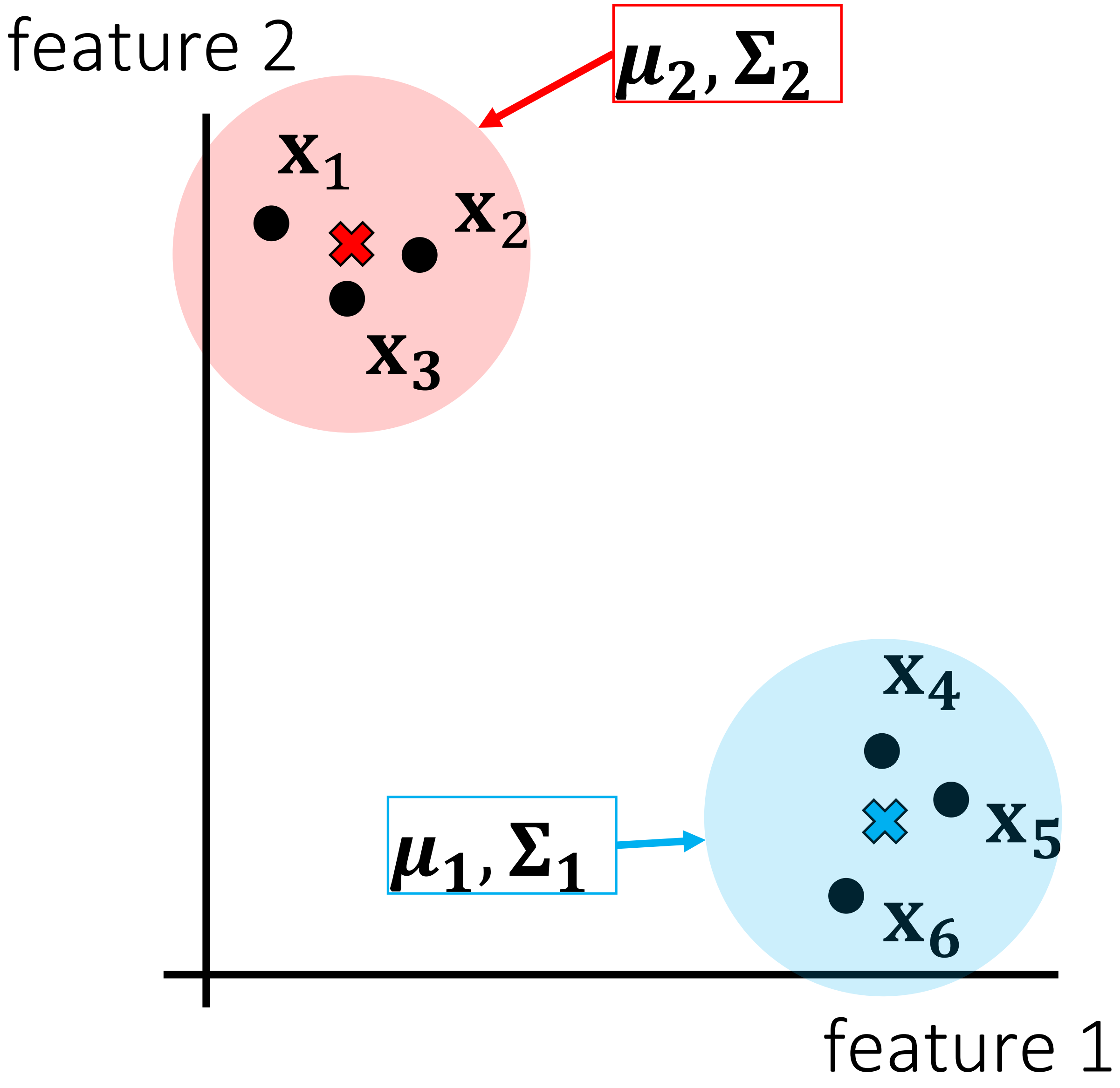
$$N_2 = (1 + 1 + 1) = 3$$

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

$$\boldsymbol{\Sigma}_2 = \frac{\sum_{n=1}^{N} z_{n2}(\mathbf{x}_n - \boldsymbol{\mu}_2)(\mathbf{x}_n - \boldsymbol{\mu}_2)^T}{N_2} = \begin{bmatrix} 0.39 & 0 \\ 0 & 0.17 \end{bmatrix} \xrightarrow{cond.\,independence} \begin{bmatrix} 0.39 & 0 \\ 0 & 0.17 \end{bmatrix}$$

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

Training



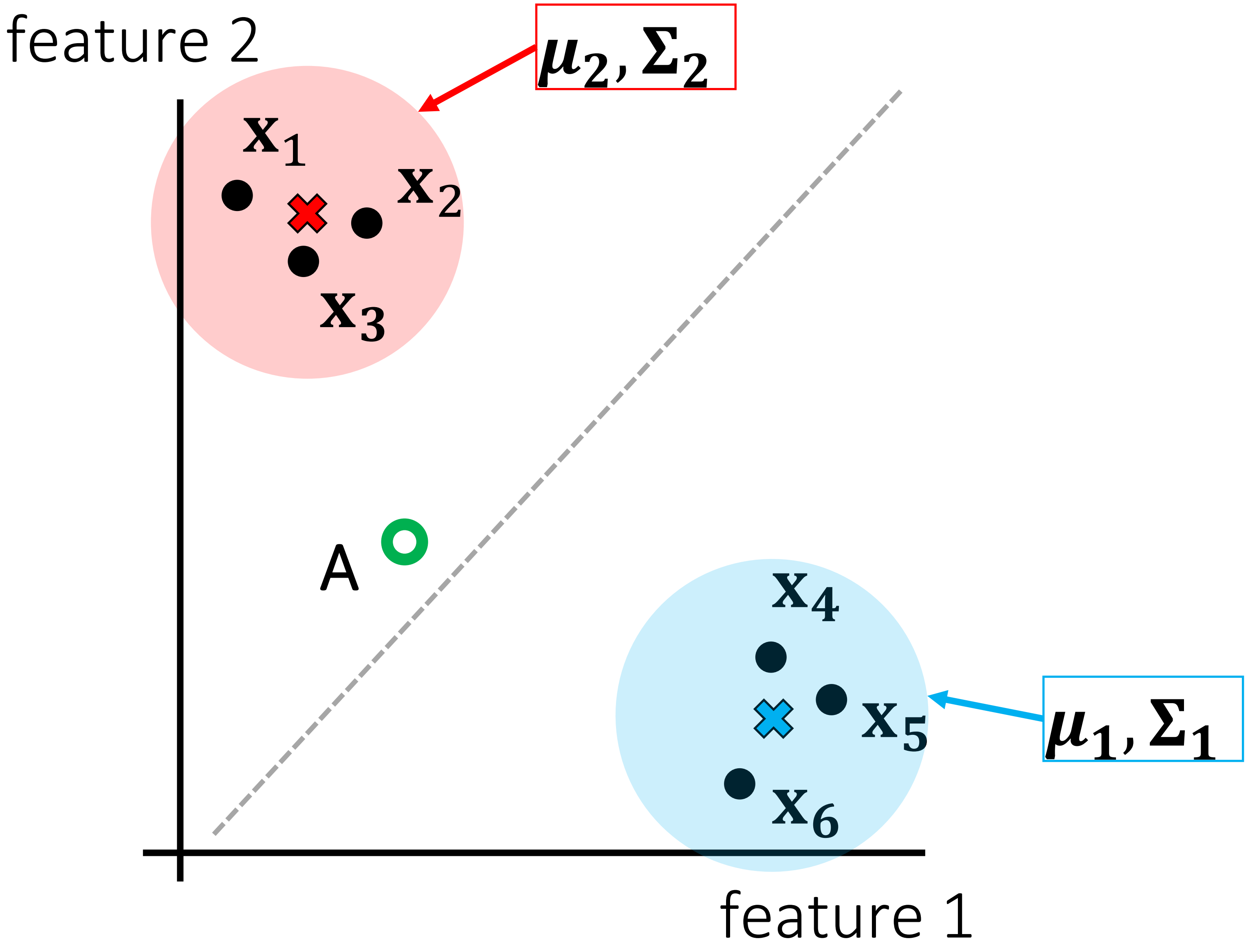
Priors: $\boldsymbol{\pi} = \begin{bmatrix} 0.50 \\ 0.50 \end{bmatrix}_{1 \times K}$

Class centers: $\mathbf{M} = \begin{bmatrix} \boldsymbol{\mu}_1^T \\ \boldsymbol{\mu}_2^T \end{bmatrix} = \begin{bmatrix} 8.5 & 1.83 \\ 1.83 & 6.84 \end{bmatrix}_{K \times D = 2 \times 2}$

Covariance matrices: $\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.21 & 0 \\ 0 & 0.40 \end{bmatrix}$

$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 0.39 & 0 \\ 0 & 0.17 \end{bmatrix}$

Inference



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

Class centers: $\mathbf{M} = \begin{bmatrix} \boldsymbol{\mu}_1^T \\ \boldsymbol{\mu}_2^T \end{bmatrix} = \begin{bmatrix} 8.5 & 1.83 \\ 1.83 & 6.84 \end{bmatrix}_{K\times D = 2\times 2}$

Covariance matrices: $\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.21 & 0 \\ 0 & 0.40 \end{bmatrix}$

$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 0.39 & 0 \\ 0 & 0.17 \end{bmatrix}$$

Inference

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

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

- Let's consider $\mathbf{x}_A^T = [3.0 \quad 4.0]$
  - For $\mathbf{z}^T = [1 \quad 0]$ (class $k = 1$)

$$\mathcal{N}(\mathbf{x}_A|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) = \frac{1}{2\pi \det\left(\begin{bmatrix} 0.21 & 0 \\ 0 & 0.4 \end{bmatrix}\right)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}[3.0 - 8.5 \quad 4.0 - 1.83]\begin{bmatrix} 0.21 & 0 \\ 0 & 0.4 \end{bmatrix}\begin{bmatrix} 3.0 - 8.5 \\ 4.0 - 1.83 \end{bmatrix}\right\} = 8.95 \times 10^{-3}$$

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

$$\mathcal{N}(\mathbf{x}_A|\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) = \frac{1}{2\pi \det\left(\begin{bmatrix} 0.39 & 0 \\ 0 & 0.17 \end{bmatrix}\right)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}[3.0 - 1.83 \quad 4.0 - 6.84]\begin{bmatrix} 0.39 & 0 \\ 0 & 0.17 \end{bmatrix}\begin{bmatrix} 3.0 - 1.83 \\ 4.0 - 6.84 \end{bmatrix}\right\} = 0.238$$

Inference

- Probability that it belongs to class $k = 1$

$$p(z_1 = 1|\mathbf{x}_n) = \frac{(0.5 \times 8.95 \times 10^{-3})}{(0.5 \times 8.95 \times 10^{-3}) + (0.5 \times 0.238)} = 0.036$$

- Probability that it belongs to class $k = 2$

$$p(z_2 = 1|\mathbf{x}_n) = \frac{(0.5 \times 0.238)}{(0.5 \times 8.95 \times 10^{-3}) + (0.5 \times 0.238)} = 0.964$$

We classify point A as pertaining to class $k = 2$