Practice exercise – Bayesian:

Covariance and Point Clouds,

Decision boundary in Bayesian classification

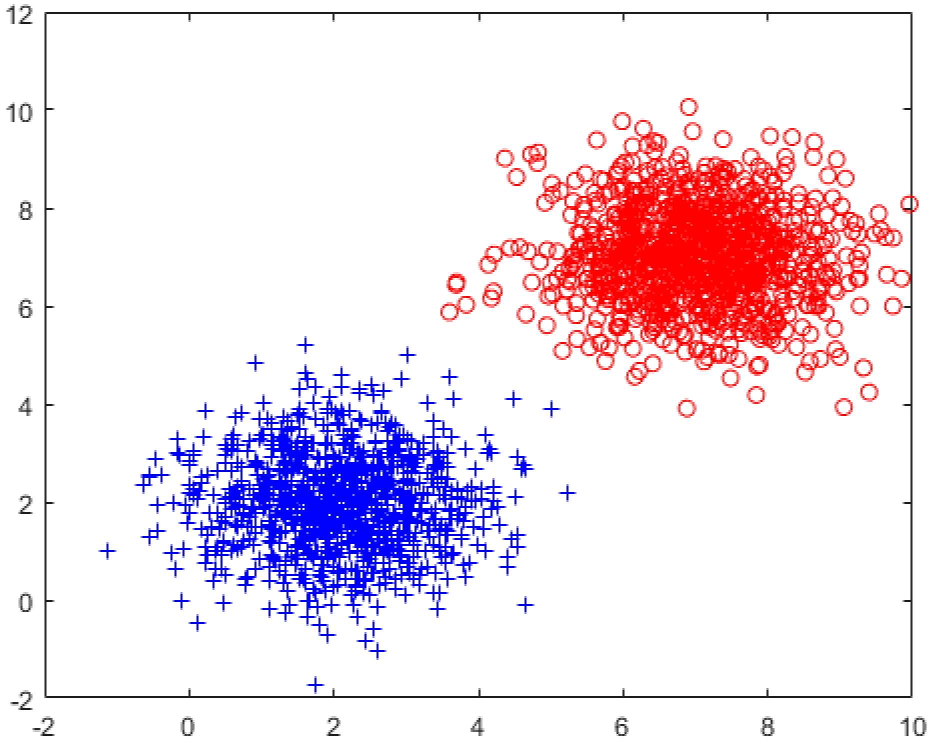
Questions 1

- What is the effect of covariance between attributes and the resultant shape modelled for the point cloud (i.e. the training samples within a class)? You can study the effect under various scenarios: (i) increase in covariance, (ii) decrease in covariance, (iii) positive covariance, (iv) negative covariance, (v) zero covariance, (vi) identical covariance for all classes, (vii) arbitrary (i.e. non-identical) covariance for different classes, etc. You may find the following MATLAB's functions useful to determine the covariance between attributes and studying the effect of covariance on the shape of point cloud: cov, mvnrnd, mvnpdf.

Covariance and point cloud

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



$$\boldsymbol{\mu}_2 = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$$

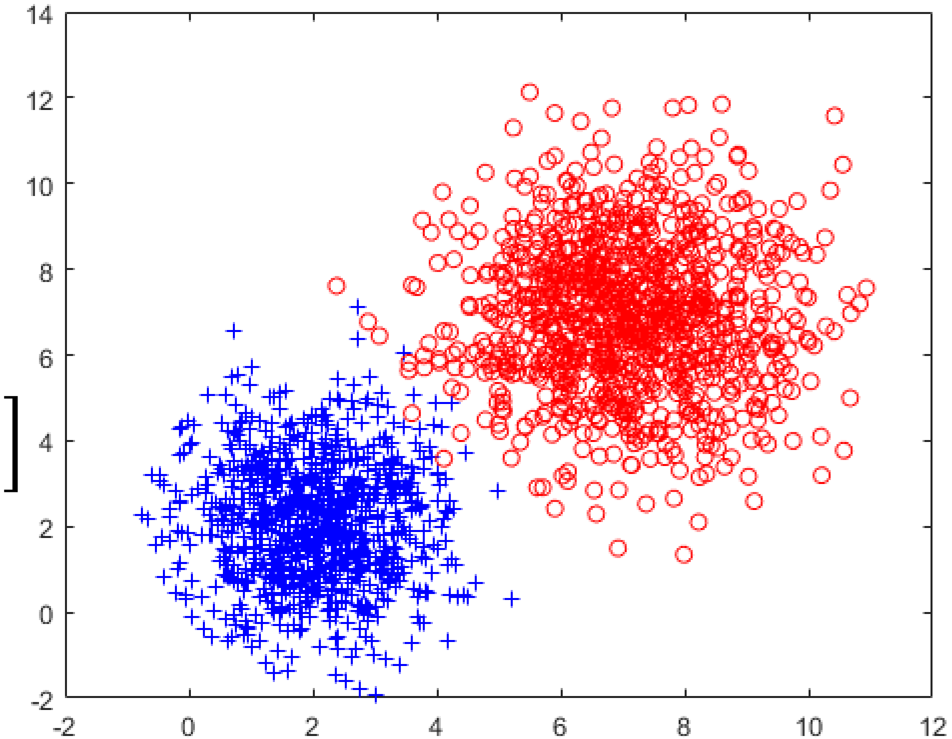
$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- Identical variance
- Zero co-variance

Covariance and point cloud

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$



$$\boldsymbol{\mu}_2 = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$$

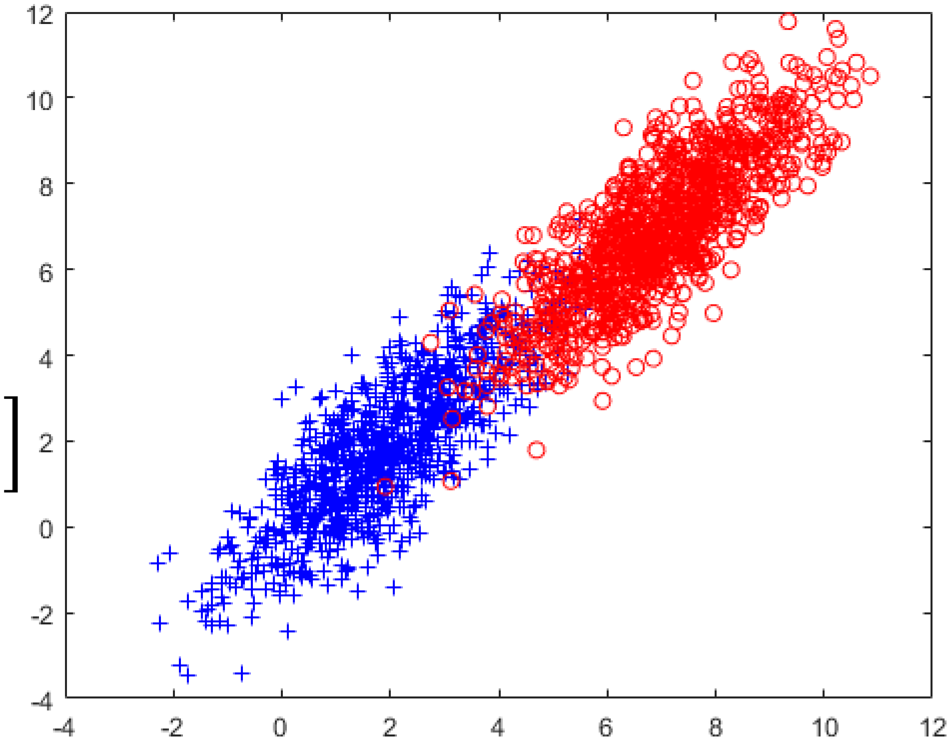
$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

- Non-identical variance
- Zero co-variance

Covariance and point cloud

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \qquad \boldsymbol{\mu}_2 = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix} \qquad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}$$

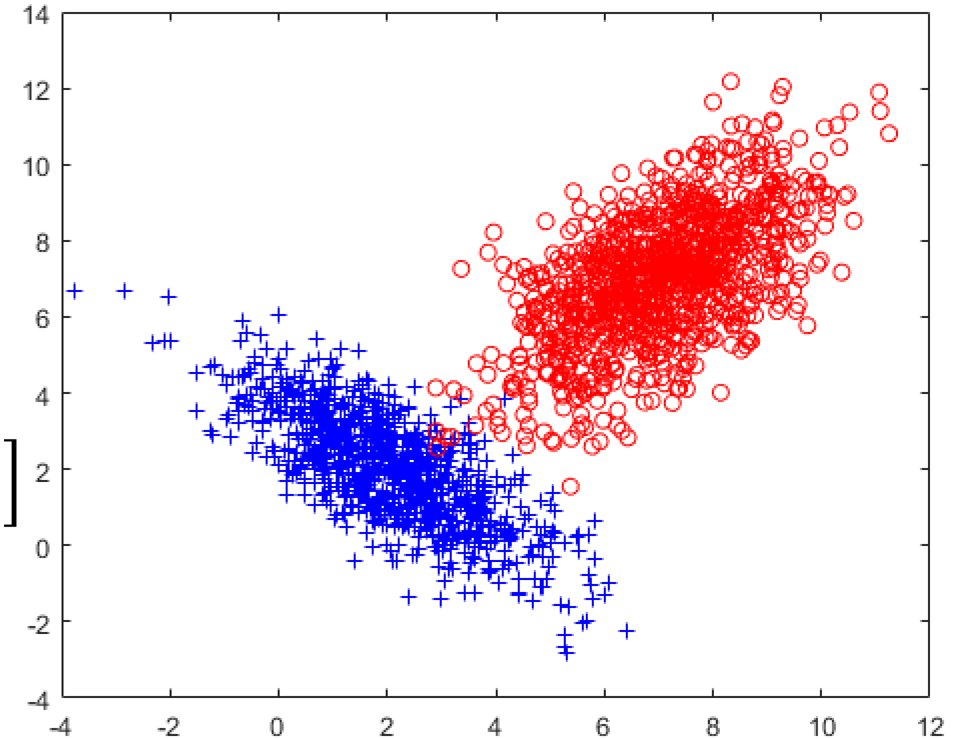


- Non-identical variance
- Identical co-variance

Covariance and point cloud

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 2 & -1.5 \\ -1.5 & 2 \end{bmatrix}$$



$$\boldsymbol{\mu}_2 = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 2 & 1.5 \\ 1.5 & 3 \end{bmatrix}$$

- Non-identical variance
- Non-identical co-variance

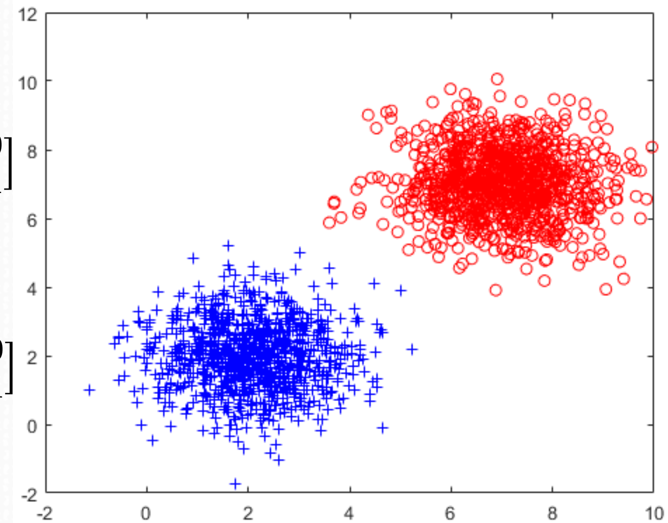
Covariance and point cloud

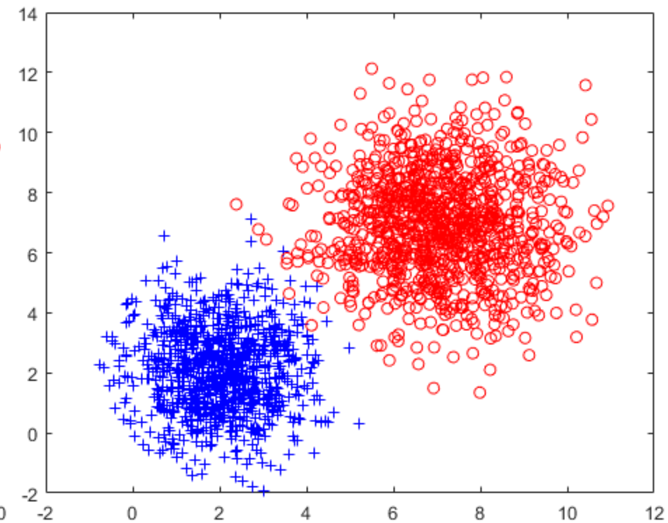
$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$\boldsymbol{\mu}_2 = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$

$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$





$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$

$\boldsymbol{\mu}_2 = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$

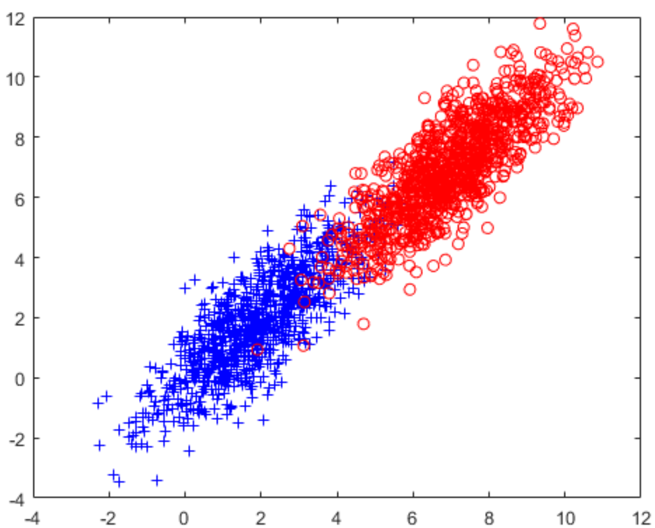
$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$

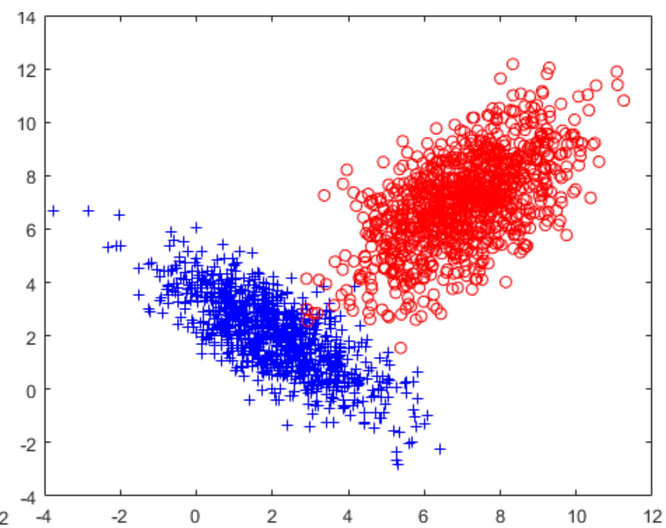
$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}$

$\boldsymbol{\mu}_2 = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$

$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}$





$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 2 & -1.5 \\ -1.5 & 2 \end{bmatrix}$

$\boldsymbol{\mu}_2 = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$

$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 2 & 1.5 \\ 1.5 & 3 \end{bmatrix}$

Question 2

- Consider a 2 class Bayesian classifier which is trained from the past examples to predict the target labels by computing the posterior estimates. However, this classifier can be considered as a model that draws a decision boundary (linear or non-linear) between the classes such that this boundary separates the training samples. Answer the following questions about the Gaussian classifier:
- (i) Is it possible for a Gaussian classifier to implement a non-linear decision boundary? If so, draw an example and suggest the shape of this non-linear decision boundary. If not, explain why not.
- (ii) How about a Gaussian Naive Bayes classifier? Justify your answer.

Question 2

Hints: With multivariate Gaussian, the following cases can be considered:

- a) statistically independent attributes, identically distributed Gaussian for each class (i.e. same variance for each class, and 0 entries at the non-diagonal location in the covariance matrix)
- b) identical covariance for each class (i.e. $\mathbf{\Sigma} = \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$)
- c) arbitrary (non-identical) covariance for each class (i.e. $\mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$)

There are two ways to attempt this problem.

- In an informal way, you can attempt this problem by considering the shape of each class (for example by drawing the density contours around training samples) under the influence of covariance matrix.
- In a more formal (and mathematical way), you can consider the maximum a posteriori estimate and determine the shape of the decision boundary in each of the above cases.

Decision boundary in Bayesian classification

- Identical variance
- Zero co-variance

Naïve

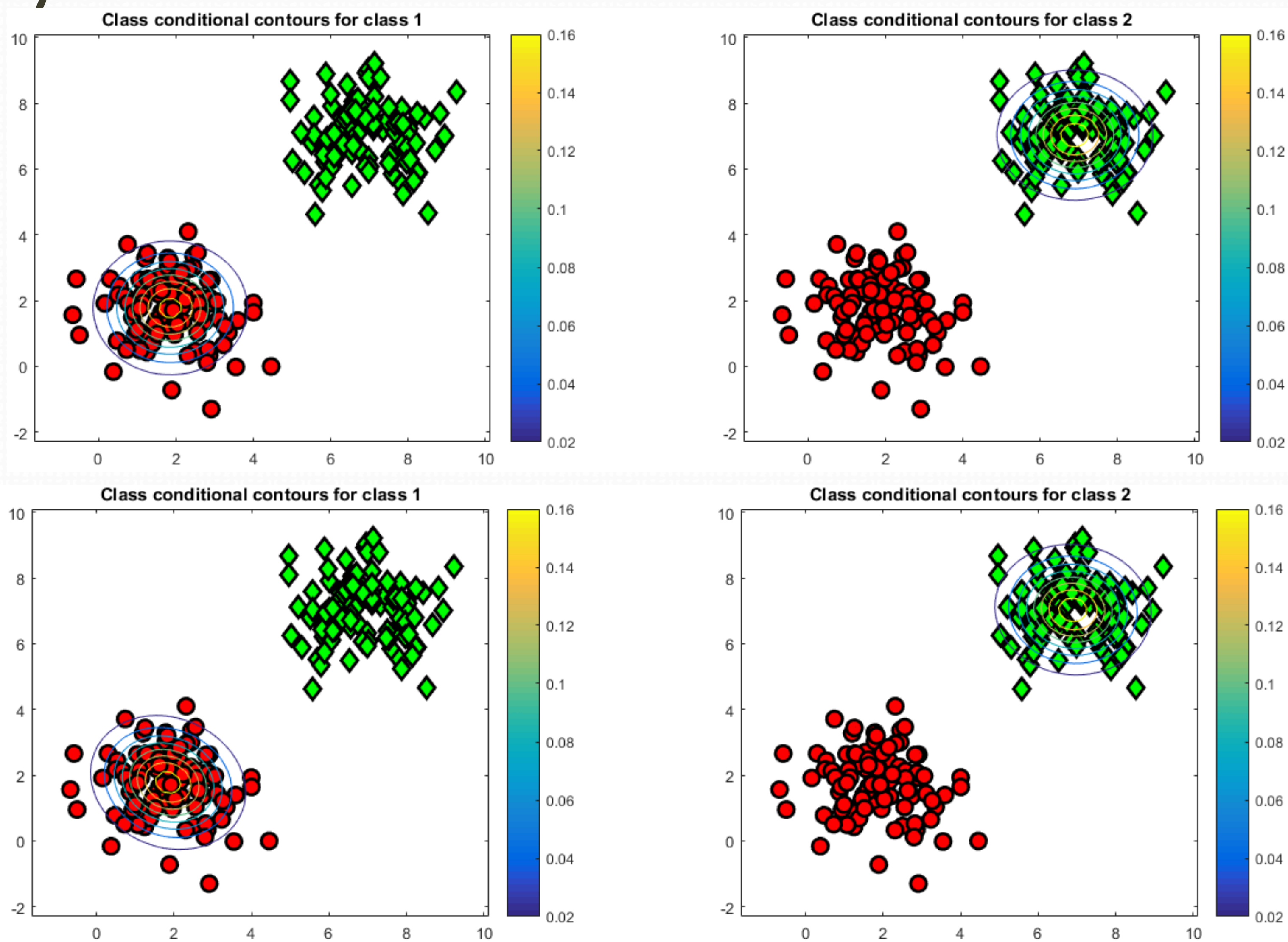
$$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\boldsymbol{\mu}_2 = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Without Naïve



Decision boundary in Bayesian classification

- Identical variance
- Zero co-variance

Naïve

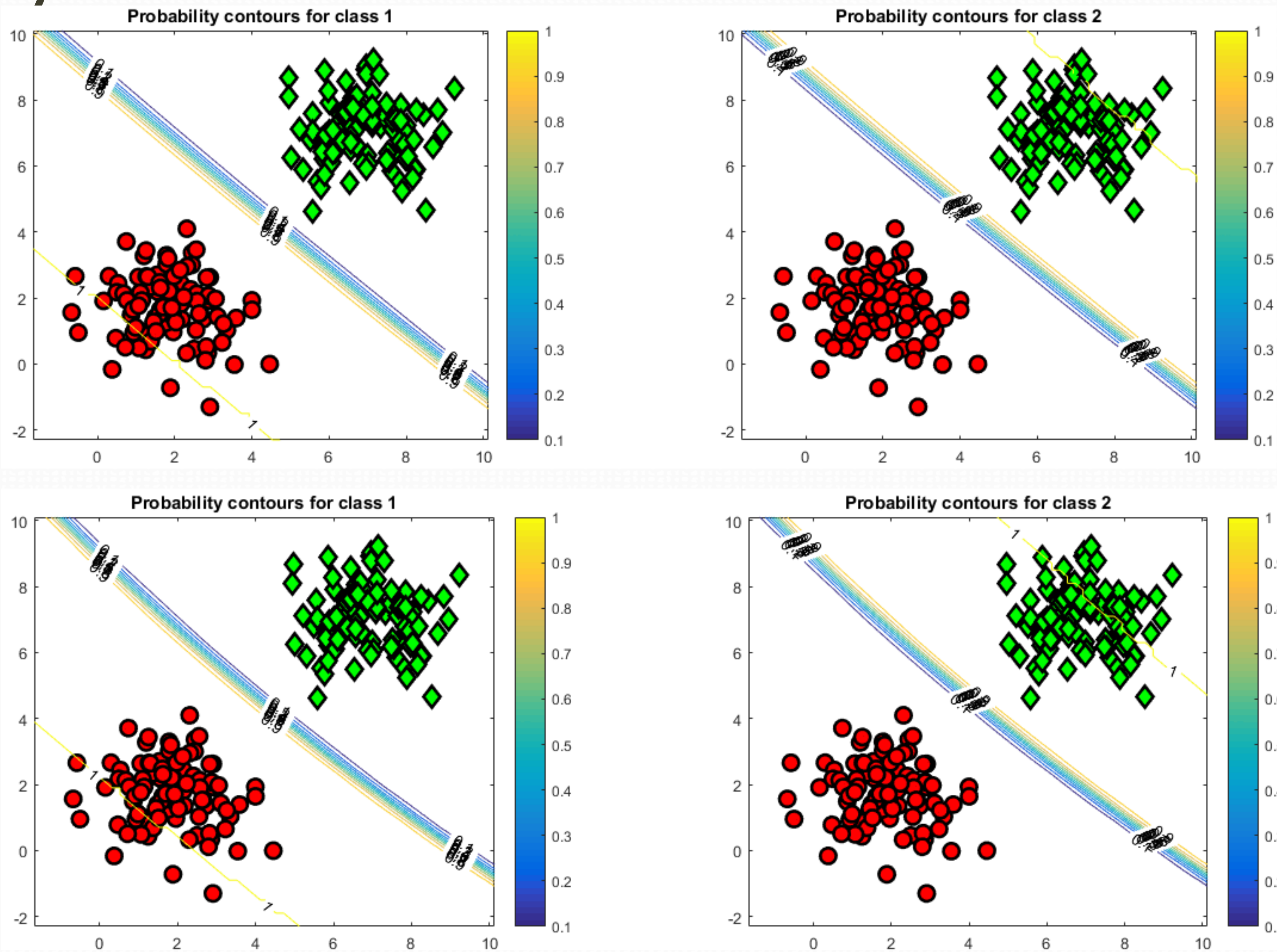
$$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\boldsymbol{\mu}_2 = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Without Naïve



Decision boundary in Bayesian classification

- Non-identical variance
- Zero co-variance

Naïve

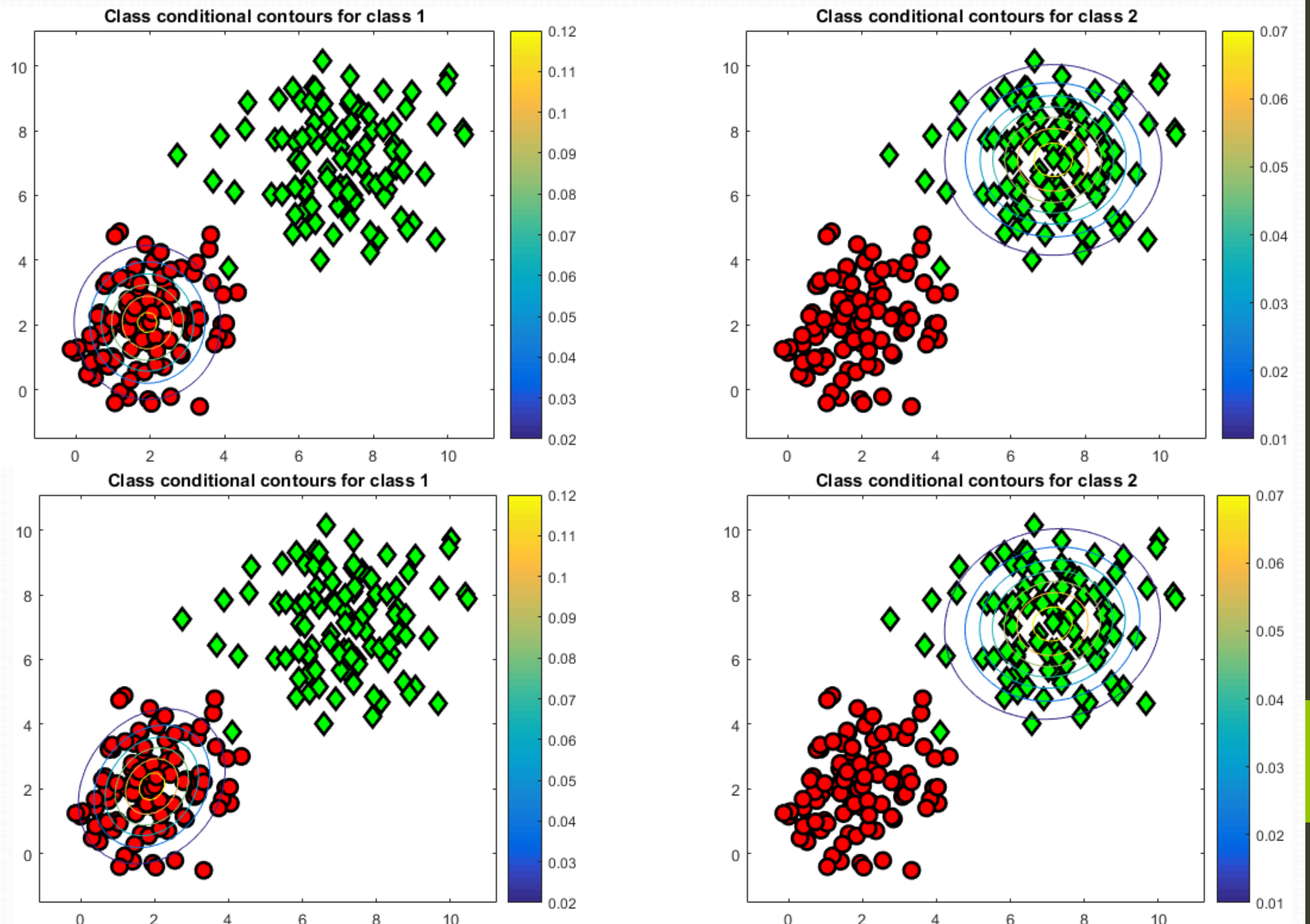
$$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\boldsymbol{\mu}_2 = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

Without Naïve



Decision boundary in Bayesian classification

- Non-identical variance
- Zero co-variance

Naïve

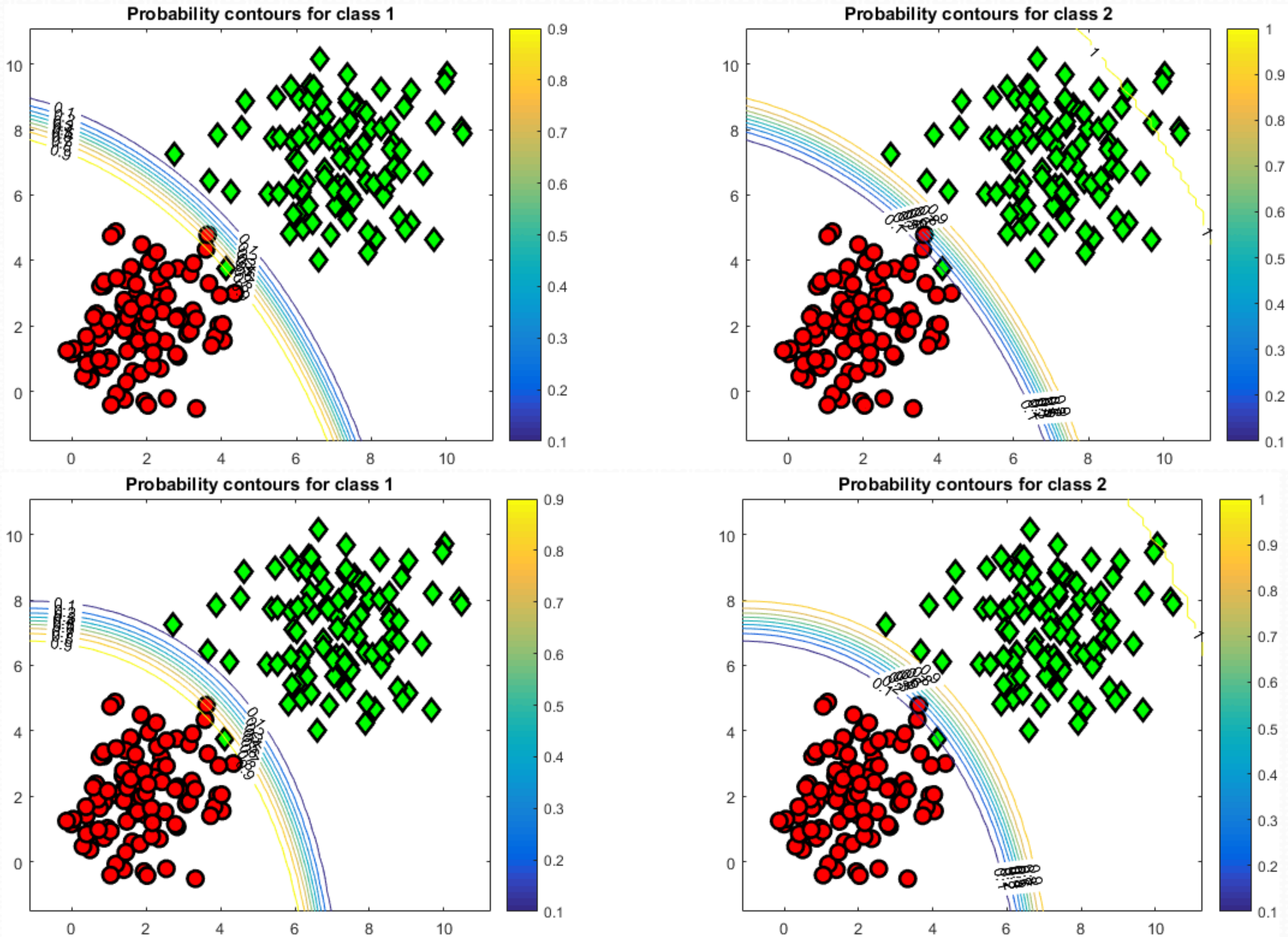
$$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\boldsymbol{\mu}_2 = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

Without Naïve



Decision boundary in Bayesian classification

- Non-identical variance
- Identical co-variance

Naïve

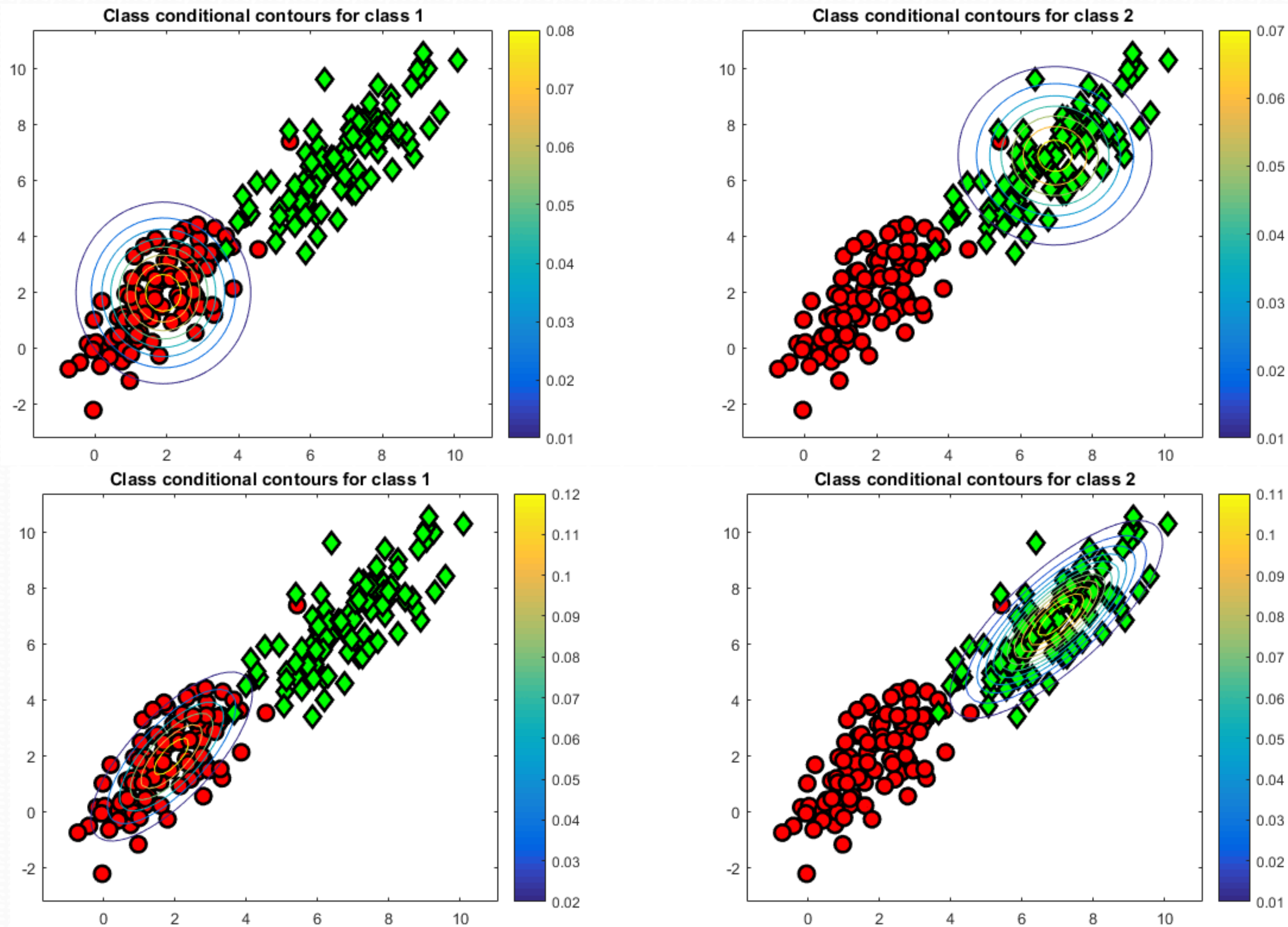
$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}$

$\boldsymbol{\mu}_2 = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$

$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}$

Without Naïve



Decision boundary in Bayesian classification

- Non-identical variance
- Identical co-variance

Naïve

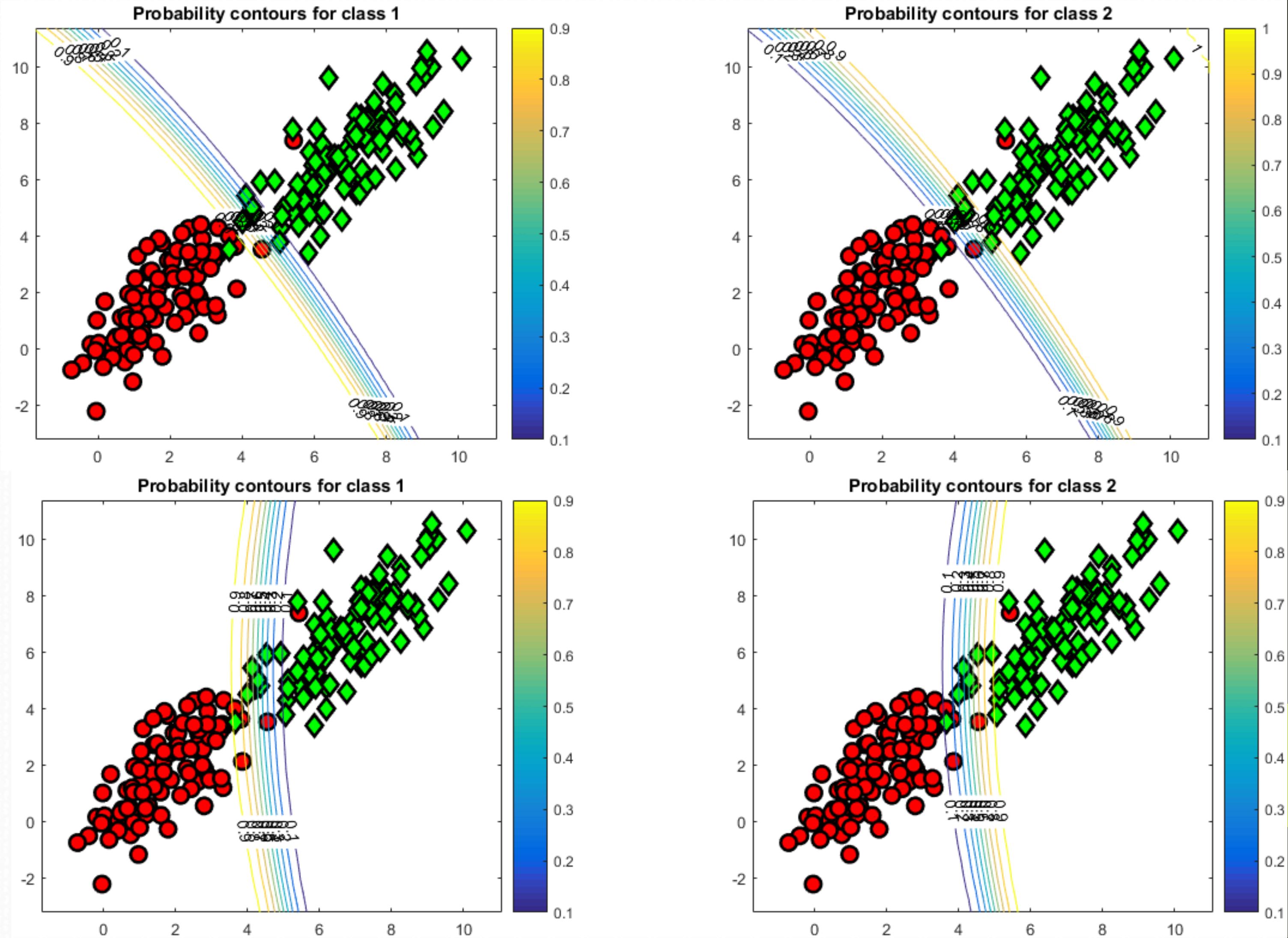
$$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}$$

$$\boldsymbol{\mu}_2 = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}$$

Without Naïve



Decision boundary in Bayesian classification

- Non-identical variance
- Non-identical co-variance

Naïve

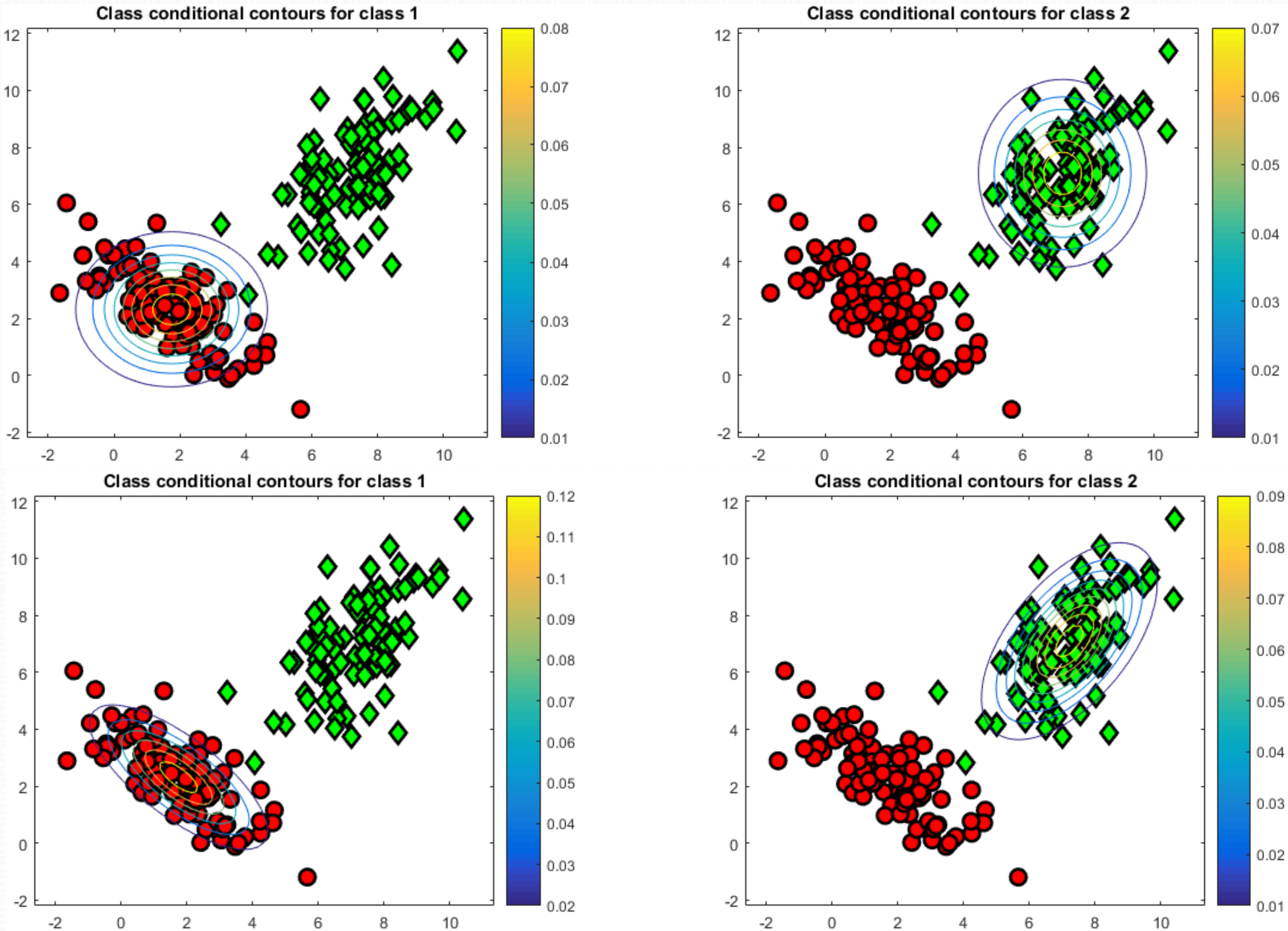
$$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 2 & -1.5 \\ -1.5 & 2 \end{bmatrix}$$

$$\boldsymbol{\mu}_2 = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 2 & 1.5 \\ 1.5 & 3 \end{bmatrix}$$

Without Naïve



Decision boundary in Bayesian classification

- Non-identical variance
- Non-identical co-variance

Naïve

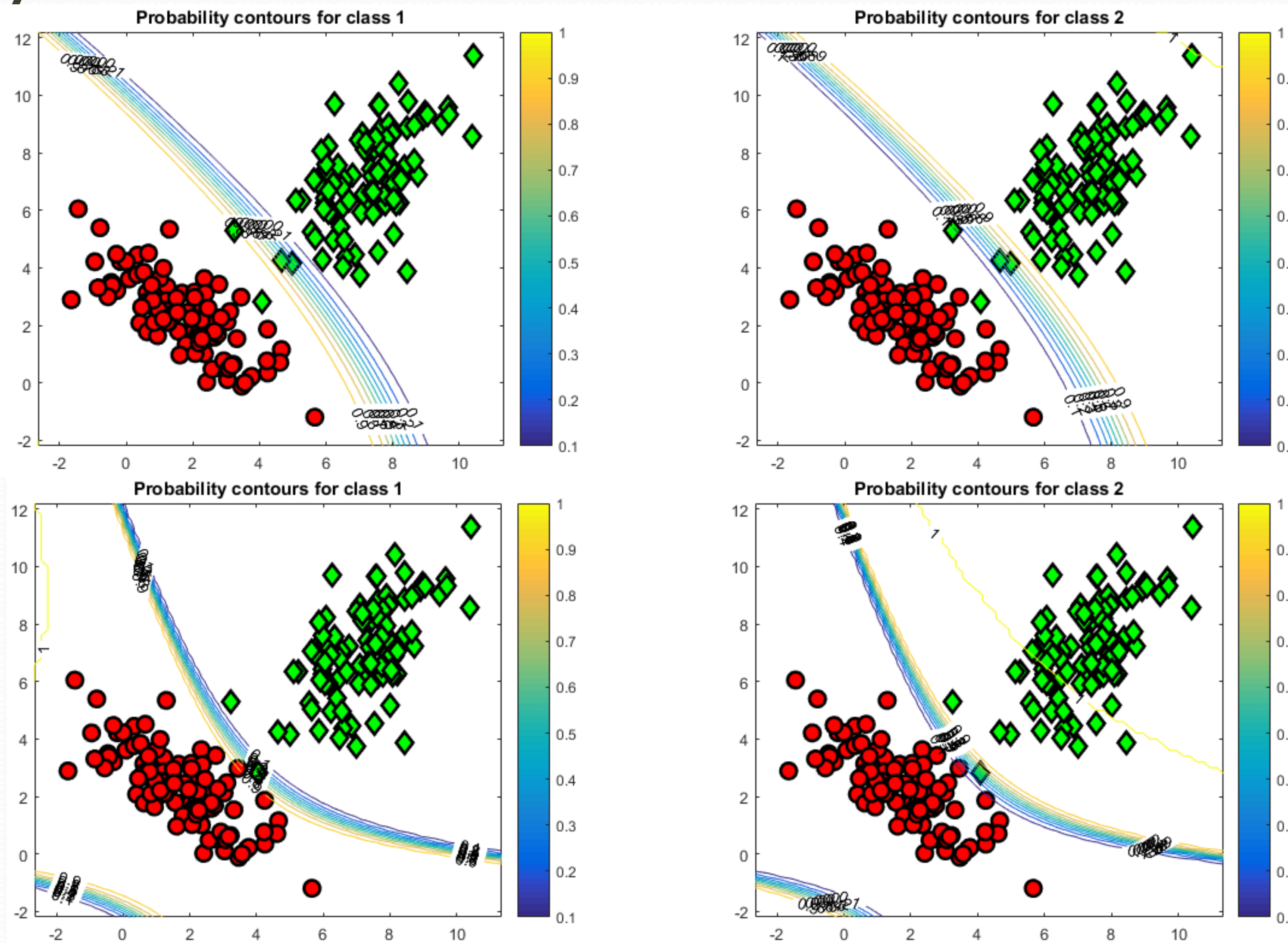
$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 2 & -1.5 \\ -1.5 & 2 \end{bmatrix}$

$\boldsymbol{\mu}_2 = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$

$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 2 & 1.5 \\ 1.5 & 3 \end{bmatrix}$

Without Naïve



Decision boundary in Bayesian classification

- The maximum a posteriori (MAP) estimate:

$$p(c|\boldsymbol{x}) = p(\boldsymbol{x}|c)p(c)$$

- Modelling class conditional likelihood with Gaussian:

$$p(c|\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}_c|}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_c)^T{\boldsymbol{\Sigma}_c}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_c)\right\}p(c)$$

- Let's consider the log of posterior:

$$\begin{aligned}g_c(\boldsymbol{x}) &= \ln\big(p(c|\boldsymbol{x})\big)\\ &= \ln\left(\frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}_c|}}\right) + \ln\left(\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_c)^T{\boldsymbol{\Sigma}_c}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_c)\right\}\right)\\ &\quad + \ln\big(p(c)\big)\end{aligned}$$

Decision boundary in Bayesian classification

$$
\begin{aligned}
g_c(\boldsymbol{x}) &= \ln\big(p(c|\boldsymbol{x})\big) \\
&= \ln\left(\frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}_c|}}\right) \\
&\quad + \ln\left(\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_c)^T\boldsymbol{\Sigma}_c{}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_c)\right\}\right) + \ln\big(p(c)\big)
\end{aligned}
$$

$$
\begin{aligned}
g_c(\boldsymbol{x}) &= \ln\left((2\pi)^{-d/2}|\boldsymbol{\Sigma}_c|^{-1/2}\right) - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_c)^T\boldsymbol{\Sigma}_c{}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_c) \\
&\quad + \ln\big(p(c)\big)
\end{aligned}
$$

$$
\begin{aligned}
g_c(\boldsymbol{x}) &= -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_c)^T\boldsymbol{\Sigma}_c{}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_c) - \frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln(|\boldsymbol{\Sigma}_c|) \\
&\quad + \ln\big(p(c)\big)
\end{aligned}
$$

Decision boundary in Bayesian classification

- Case I: identical variance, zero co-variance case
  - $\mathbf{\Sigma}_c = \mathbf{\Sigma} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \sigma^2 I, \qquad \forall c \in 1,2,\dots,C$

$$g_c(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_c)^T \mathbf{\Sigma}_c{}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_c) - \frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln(|\mathbf{\Sigma}_c|) + \ln\big(p(c)\big)$$

becomes

$$g_c(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_c)^T \frac{1}{\sigma^2} I(\boldsymbol{x}-\boldsymbol{\mu}_c) - \frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2|I|) + \ln\big(p(c)\big)$$

- Omitting class independent terms $\frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2|I|)$ and expanding:

$$g_c(\boldsymbol{x}) = -\frac{1}{2\sigma^2}(\boldsymbol{x}^T\boldsymbol{x} - 2\boldsymbol{\mu}_c^T\boldsymbol{x} + \boldsymbol{\mu}_c^T\boldsymbol{\mu}_c) + \ln\big(p(c)\big)$$

Decision boundary in Bayesian classification

- Case I: identical variance, zero co-variance case

$$g_c(\boldsymbol{x}) = -\frac{1}{2\sigma^2}(\boldsymbol{x}^T\boldsymbol{x} - 2\boldsymbol{\mu}_c^T\boldsymbol{x} + \boldsymbol{\mu}_c^T\boldsymbol{\mu}_c) + \ln\big(p(c)\big)$$

$$g_c(\boldsymbol{x}) = -\frac{1}{2\sigma^2}\boldsymbol{x}^T\boldsymbol{x} + \frac{1}{\sigma^2}\boldsymbol{\mu}_c^T\boldsymbol{x} - \frac{1}{2\sigma^2}\boldsymbol{\mu}_c^T\boldsymbol{\mu}_c + \ln\big(p(c)\big)$$

Omitting class independent term $-\frac{1}{2\sigma^2}\boldsymbol{x}^T\boldsymbol{x}$:

$$g_c(\boldsymbol{x}) = \frac{1}{\sigma^2}\boldsymbol{\mu}_c^T\boldsymbol{x} - \frac{1}{2\sigma^2}\boldsymbol{\mu}_c^T\boldsymbol{\mu}_c + \ln\big(p(c)\big)$$

which can be considered as:

$g_c(\boldsymbol{x}) = \boldsymbol{w}_c^T\boldsymbol{x} + w_{c,0}$ i.e. linear boundary

where

$\boldsymbol{w}_c^T = \frac{1}{\sigma^2}\boldsymbol{\mu}_c^T$ and $w_{c,0} = -\frac{1}{2\sigma^2}\boldsymbol{\mu}_c^T\boldsymbol{\mu}_c + \ln\big(p(c)\big)$

Decision boundary in Bayesian classification

- Case II: identical co-variance for each class
  - $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_c = \boldsymbol{\Sigma}, \ \forall c \in 1,2,\ldots,C$

$$g_c(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_c)^T {\boldsymbol{\Sigma}_c}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_c) - \frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln(|\boldsymbol{\Sigma}_c|) + \ln\big(p(c)\big)$$

becomes

$$g_c(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_c)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_c) - \frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln(|\boldsymbol{\Sigma}|) + \ln\big(p(c)\big)$$

- Omitting class independent terms $\frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln(|\boldsymbol{\Sigma}|)$ and expanding:

$$g_c(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} - 2(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_c)^T\boldsymbol{x} + \boldsymbol{\mu}_c^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_c) + \ln\big(p(c)\big)$$

Decision boundary in Bayesian classification

- Case II: identical co-variance for each class

$$g_c(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} - 2(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_c)^T\boldsymbol{x} + \boldsymbol{\mu}_c^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_c) + \ln\big(p(c)\big)$$

$$g_c(\boldsymbol{x}) = -\frac{1}{2}\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} + (\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_c)^T\boldsymbol{x} - \frac{1}{2}\boldsymbol{\mu}_c^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_c + \ln\big(p(c)\big)$$

Omitting class independent term $-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}$:

$$g_c(\boldsymbol{x}) = (\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_c)^T\boldsymbol{x} - \frac{1}{2}\boldsymbol{\mu}_c^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_c + \ln\big(p(c)\big)$$

which can be considered as:

$g_c(\boldsymbol{x}) = \boldsymbol{w}_c^T\boldsymbol{x} + w_{c,0}$ i.e. linear boundary

where

$\boldsymbol{w}_c^T = (\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_c)^T$ and $w_{c,0} = -\frac{1}{2}\boldsymbol{\mu}_c^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_c + \ln\big(p(c)\big)$

Decision boundary in Bayesian classification

- Case III: non-identical co-variance for each class
  - $\mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2 \neq \mathbf{\Sigma}_c \neq \mathbf{\Sigma}, \quad \forall c \in 1,2,\ldots,C$

$$g_c(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_c)^T \mathbf{\Sigma}_c{}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_c) - \frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln(|\mathbf{\Sigma}_c|) + \ln\big(p(c)\big)$$

becomes

$$g_c(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_c)^T \mathbf{\Sigma}_c{}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_c) - \frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln(|\mathbf{\Sigma}_c|) + \ln\big(p(c)\big)$$

- Omitting class independent terms $\frac{d}{2}\ln(2\pi)$ and expanding:

$$g_c(\boldsymbol{x}) = -\frac{1}{2}\Big(\boldsymbol{x}^T\mathbf{\Sigma}_c{}^{-1}\boldsymbol{x} - 2\big(\mathbf{\Sigma}_c{}^{-1}\boldsymbol{\mu}_c\big)^T\boldsymbol{x} + \boldsymbol{\mu}_c^T\mathbf{\Sigma}_c{}^{-1}\boldsymbol{\mu}_c\Big) - \frac{1}{2}\ln(|\mathbf{\Sigma}_c|) + \ln\big(p(c)\big)$$

Decision boundary in Bayesian classification

- Case III: non-identical co-variance for each class

$$g_c(\boldsymbol{x}) = -\frac{1}{2}\left(\boldsymbol{x}^T\boldsymbol{\Sigma}_c{}^{-1}\boldsymbol{x} - 2\left(\boldsymbol{\Sigma}_c{}^{-1}\boldsymbol{\mu}_c\right)^T\boldsymbol{x} + \boldsymbol{\mu}_c^T\boldsymbol{\Sigma}_c{}^{-1}\boldsymbol{\mu}_c\right) - \frac{1}{2}\ln(|\boldsymbol{\Sigma}_c|) + \ln\left(p(c)\right)$$

$$g_c(\boldsymbol{x}) = -\frac{1}{2}\boldsymbol{x}^T\boldsymbol{\Sigma}_c{}^{-1}\boldsymbol{x} + \left(\boldsymbol{\Sigma}_c{}^{-1}\boldsymbol{\mu}_c\right)^T\boldsymbol{x} - \frac{1}{2}\boldsymbol{\mu}_c^T\boldsymbol{\Sigma}_c{}^{-1}\boldsymbol{\mu}_c - \frac{1}{2}\ln(|\boldsymbol{\Sigma}_c|) + \ln\left(p(c)\right)$$

$$g_c(\boldsymbol{x}) = \boldsymbol{x}^T\left(\frac{-\boldsymbol{\Sigma}_c{}^{-1}}{2}\right)\boldsymbol{x} + \left(\boldsymbol{\Sigma}_c{}^{-1}\boldsymbol{\mu}_c\right)^T\boldsymbol{x} - \frac{1}{2}\boldsymbol{\mu}_c^T\boldsymbol{\Sigma}_c{}^{-1}\boldsymbol{\mu}_c - \frac{1}{2}\ln(|\boldsymbol{\Sigma}_c|) + \ln\left(p(c)\right)$$

which can be considered as:

$g_c(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{w}_{c,2}^T\boldsymbol{x} + \boldsymbol{w}_{c,1}^T\boldsymbol{x} + w_{c,0}$ i.e. quadratic boundary

where $\boldsymbol{w}_{c,2}^T = \frac{-\boldsymbol{\Sigma}_c{}^{-1}}{2}$, $\boldsymbol{w}_{c,1}^T = \left(\boldsymbol{\Sigma}_c{}^{-1}\boldsymbol{\mu}_c\right)^T$ and

$w_{c,0} = -\frac{1}{2}\boldsymbol{\mu}_c^T\boldsymbol{\Sigma}_c{}^{-1}\boldsymbol{\mu}_c - \frac{1}{2}\ln(|\boldsymbol{\Sigma}_c|) + \ln\left(p(c)\right)$