

Theoretical exercise 3

Håvard Godal

February 8, 2021

Problem 1

Assume that the underlying a priori probabilities and class conditional probability density functions from problem 2, exercise 2 is unknown. However, we have access to measurements so that we have the following samples from the two categories:

$$\chi_1 = \left\{ \begin{pmatrix} 2 \\ 6 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 3 \\ 8 \end{pmatrix}, \begin{pmatrix} 4 \\ 6 \end{pmatrix} \right\} \quad (1.1)$$

and

$$\chi_2 = \left\{ \begin{pmatrix} 1 \\ -2 \end{pmatrix}, \begin{pmatrix} 2.7 \\ -4 \end{pmatrix}, \begin{pmatrix} 3.3 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 \\ -2 \end{pmatrix} \right\} \quad (1.2)$$

a) - Assuming a Gaussian distribution, use a parametric approach to formulate the Bayes classifier. Apply the maximum-likelihood (ML) method to estimate the required functions.

Finding μ and Σ for each class ω_1 and ω_2 by using the maximum-likelihood-estimators for a multivariate gaussian distribuion:

For μ_i :

$$\begin{aligned}\hat{\mu}_1 &= \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \\ &= \frac{1}{4} \left[\begin{pmatrix} 2 \\ 6 \end{pmatrix} + \begin{pmatrix} 3 \\ 4 \end{pmatrix} + \begin{pmatrix} 3 \\ 8 \end{pmatrix} + \begin{pmatrix} 4 \\ 6 \end{pmatrix} \right] \\ &= \begin{pmatrix} 3 \\ 6 \end{pmatrix}\end{aligned}\tag{1.3}$$

$$\begin{aligned}\hat{\mu}_2 &= \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \\ &= \frac{1}{4} \left[\begin{pmatrix} 1 \\ -2 \end{pmatrix} + \begin{pmatrix} 2.7 \\ -4 \end{pmatrix} + \begin{pmatrix} 3.3 \\ 0 \end{pmatrix} + \begin{pmatrix} 5 \\ -2 \end{pmatrix} \right] \\ &= \begin{pmatrix} 3 \\ -2 \end{pmatrix}\end{aligned}\tag{1.4}$$

For Σ_i :

$$\begin{aligned}\hat{\Sigma}_1 &= \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \hat{\mu}_1)(\mathbf{x}_i - \hat{\mu}_1)^T \\ &= \frac{1}{4} \left[\begin{pmatrix} -1 \\ 0 \end{pmatrix} \begin{pmatrix} -1 & 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -2 \end{pmatrix} \begin{pmatrix} 0 & -2 \end{pmatrix} + \begin{pmatrix} 0 \\ 2 \end{pmatrix} \begin{pmatrix} 0 & 2 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} \right] \\ &= \frac{1}{4} \left[\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right] \\ &= \frac{1}{4} \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 2 \end{pmatrix}\end{aligned}\tag{1.5}$$

$$\begin{aligned}\hat{\Sigma}_2 &= \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \hat{\mu}_2)(\mathbf{x}_i - \hat{\mu}_2)^T \\ &= \frac{1}{4} \left[\begin{pmatrix} -2 \\ 0 \end{pmatrix} \begin{pmatrix} -2 & 0 \end{pmatrix} + \begin{pmatrix} -0.3 \\ -2 \end{pmatrix} \begin{pmatrix} -0.3 & -2 \end{pmatrix} + \begin{pmatrix} 0.3 \\ 2 \end{pmatrix} \begin{pmatrix} 0.3 & 2 \end{pmatrix} + \begin{pmatrix} 2 \\ 0 \end{pmatrix} \begin{pmatrix} 2 & 0 \end{pmatrix} \right] \\ &= \frac{1}{4} \left[\begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0.09 & 0.6 \\ 0.6 & 4 \end{pmatrix} + \begin{pmatrix} 0.09 & 0.6 \\ 0.6 & 4 \end{pmatrix} + \begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix} \right] \\ &= \frac{1}{4} \begin{pmatrix} 8.18 & 1.2 \\ 1.2 & 8 \end{pmatrix} \\ &= \begin{pmatrix} 2.045 & 0.3 \\ 0.3 & 2 \end{pmatrix}\end{aligned}\tag{1.6}$$

Based on the number of observations of each class, the following priors are determined:

$$P(\omega_1) = \frac{4}{8} = \frac{1}{2} \quad (1.7)$$

$$P(\omega_2) = \frac{4}{8} = \frac{1}{2} \quad (1.8)$$

Bayes decision rule is as follows:

$$\text{Decide } \begin{cases} \omega_1 & \text{if } P(\omega_1|\mathbf{x}) > P(\omega_2|\mathbf{x}) \\ \omega_2 & \text{otherwise} \end{cases} \quad (1.9)$$

Using the same procedure as in *Exercise 2, problem 2* to calculate the discriminant functions $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$:

$$g_i(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\Theta}_i \mathbf{x} + \boldsymbol{\theta}_i^T \mathbf{x} + \theta_{i0} \quad (1.10)$$

where:

$$\begin{aligned} \boldsymbol{\Theta}_i &= -\frac{1}{2} \Sigma_i^{-1} \\ \boldsymbol{\theta}_i &= \Sigma_i^{-1} \boldsymbol{\mu}_i \\ \theta_{i0} &= -\frac{1}{2} \boldsymbol{\mu}_i^T \Sigma_i^{-1} \boldsymbol{\mu}_i - \frac{1}{2} \ln|\Sigma_i| + \ln(P(\omega_i)) \end{aligned} \quad (1.11)$$

Solving for $g_1(\mathbf{x})$:

$$\begin{aligned} \boldsymbol{\Theta}_1 &= -\frac{1}{2} \begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix} \\ &= \begin{bmatrix} -1 & 0 \\ 0 & -1/4 \end{bmatrix} \end{aligned} \quad (1.12)$$

$$\begin{aligned} \boldsymbol{\theta}_1 &= \begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 3 \\ 6 \end{bmatrix} \\ &= \begin{bmatrix} 6 \\ 3 \end{bmatrix} \end{aligned} \quad (1.13)$$

$$\begin{aligned} \theta_{10} &= -\frac{1}{2} \begin{bmatrix} 3 & 6 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 3 \\ 6 \end{bmatrix} - \frac{1}{2} \ln(1/2 \cdot 2 - 0 \cdot 0) + \ln(1/2) \\ &= -18.69314 \end{aligned} \quad (1.14)$$

$$\begin{aligned} g_1(\mathbf{x}) &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -1/4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 6 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - 18.69314 \\ &= -x_1^2 - \frac{1}{4}x_2^2 + 6x_1 + 3x_2 - 18.69314 \end{aligned} \quad (1.15)$$

Solving for $g_2(\mathbf{x})$:

$$\begin{aligned}\Theta_2 &= -\frac{1}{2} \begin{bmatrix} 0.5 & -0.075 \\ -0.075 & 0.51125 \end{bmatrix} \\ &= \begin{bmatrix} -0.25 & 0.0375 \\ 0.0375 & -0.255625 \end{bmatrix}\end{aligned}\quad (1.16)$$

$$\begin{aligned}\theta_2 &= \begin{bmatrix} 0.5 & -0.075 \\ -0.075 & 0.51125 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} \\ &= \begin{bmatrix} 1.65 \\ -1.2475 \end{bmatrix}\end{aligned}\quad (1.17)$$

$$\begin{aligned}\theta_{20} &= -\frac{1}{2} \begin{bmatrix} 3 & -2 \end{bmatrix} \begin{bmatrix} 0.5 & -0.075 \\ -0.075 & 0.51125 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} - \frac{1}{2} \ln(2.045 \cdot 2 - 0.3 \cdot 0.3) + \ln(1/2) \\ &= -5.11379\end{aligned}\quad (1.18)$$

$$\begin{aligned}g_2(\mathbf{x}) &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} -0.25 & 0.0375 \\ 0.0375 & -0.255625 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1.65 & -1.2475 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - 5.11379 \\ &= -0.25x_1^2 - 0.255625x_2^2 + 0.075x_1x_2 + 1.65x_1 - 1.2475x_2 - 5.11379\end{aligned}\quad (1.19)$$

The decision border occurs when $g_1(\mathbf{x}) = g_2(\mathbf{x})$

$$\begin{aligned}g_1(\mathbf{x}) &= g_2(\mathbf{x}) \\ -x_1^2 - 0.25x_2^2 + 6x_1 + 3x_2 - 18.69314 &= \\ -0.25x_1^2 - 0.255625x_2^2 + 0.075x_1x_2 + 1.65x_1 - 1.2475x_2 - 5.11379 &= \\ -0.75x_1^2 + 0.005625x_2^2 + 4.25x_1 + 4.248x_2 - 0.075x_1x_2 - 13.57935 &= 0\end{aligned}\quad (1.20)$$

Using Symbolab to solve for x_2 :

$$x_2 = \frac{-4.248 + 0.075x_1 \pm \sqrt{0.0225x_1^2 - 0.732825x_1 + 18.35103}}{0.01125}\quad (1.21)$$

One of the solutions, $\dots - \sqrt{\dots}$, only provides values of around -600 , which clearly does not conform with the relevant decision boundary. Therefore:

$$x_2 = \frac{-4.248 + 0.075x_1 + \sqrt{0.0225x_1^2 - 0.732825x_1 + 18.35103}}{0.01125}\quad (1.22)$$

b) - Evaluate the decision border. How are the estimated density function oriented in relation to the true density functions?

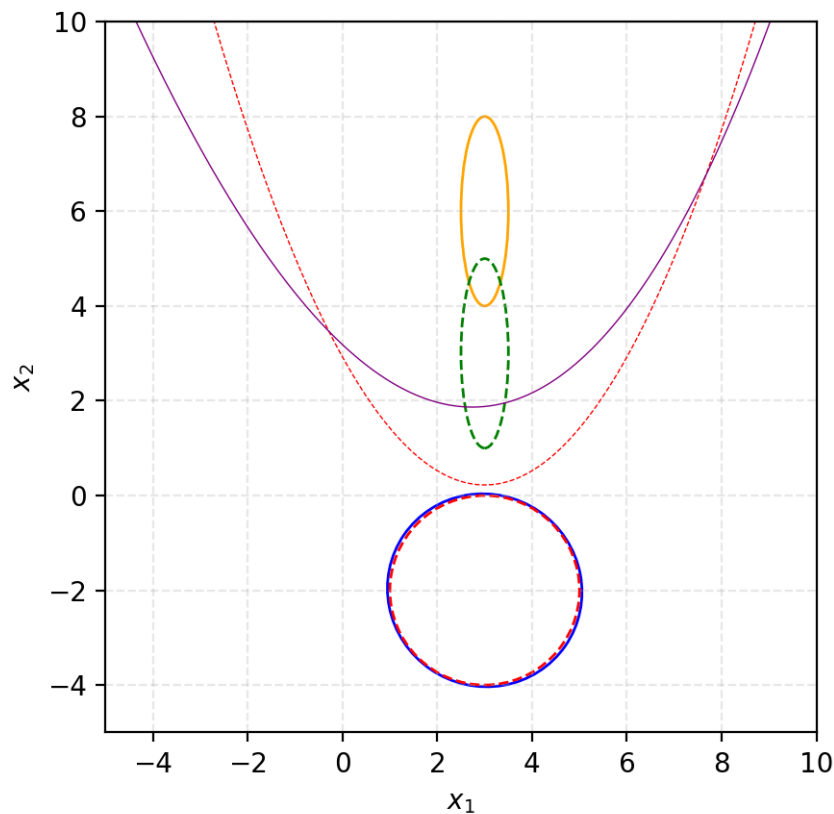


Figure 1: Bivariate gaussian distributions

In *Figure 1*, the dotted lines represents the true density functions and the true decision border found in *Exercise 2*. The solid lines represent the density functions and the decision border created from the given data points.

The density of class ω_1 , represented with an orange ellipse from the eigenvectors, is skewed. As a result, the decision boundary is therefore also skewed.

Also, note that the new decision boundary is not symmetric around the density functions anymore.

c) - How can the decision border correspond better to the true decision border.

The best way to make the new decision border correspond better to the true decision border is to collect more data points.

To better illustrate this, imagine throwing two dices and counting the number of eyes each time. If we take the average number of eyes from 4 throws, the difference between each turn will be great. On the other hand, if we were to throw the two dices 1000000 times, it would be clear that the probability density function of getting a certain amount of eyes on the dices have a Gaussian distribution. The same goes for this exercise. More data equals a better estimation of the true value/distribution.

Problem 2

Using the data set from the previous problem, classify the feature vector $\mathbf{x} = (2.5 \ 2.0)^T$. Use the following classifiers:

a) - The Bayes classifier from *Problem 1*.

$$g_1(\mathbf{x}) = -x_1^2 - \frac{1}{4}x_2^2 + 6x_1 + 3x_2 - 18.69314 \quad (2.1)$$

$$\begin{aligned} g_1\left(\begin{pmatrix} 2.5 \\ 2.0 \end{pmatrix}\right) &= -2.5^2 - \frac{1}{4}2.0^2 + 6 \cdot 2.5 + 3 \cdot 2.0 - 18.69314 \\ &= -4.94314 \end{aligned} \quad (2.2)$$

$$g_2(\mathbf{x}) = -0.25x_1^2 - 0.255625x_2^2 + 0.075x_1x_2 + 1.65x_1 - 1.2475x_2 - 5.11379 \quad (2.3)$$

$$\begin{aligned} g_2\left(\begin{pmatrix} 2.5 \\ 2.0 \end{pmatrix}\right) &= -0.25 \cdot 2.5^2 - 0.255625 \cdot 2.0^2 + 0.075 \cdot 2.5 \cdot 2.0 + 1.65 \cdot 2.5 - 1.2475 \cdot 2.0 - 5.11379 \\ &= -5.69375 \end{aligned} \quad (2.4)$$

Because $g_1(\mathbf{x}) > g_2(\mathbf{x})$, \mathbf{x} is classified to belong to class ω_1 .

b) - A Parzen-window classifier. Use a gaussian window function so that:

$$p_n(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \frac{1}{V_N} \phi(\mathbf{u}) \quad (2.5)$$

where $V_N = h_N^l$,

$$\phi(\mathbf{u}) = \frac{1}{(2\pi)^{\frac{d}{2}} |\mathbf{I}|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(\mathbf{u})^T \mathbf{I}^{-1}(\mathbf{u})} \quad (2.6)$$

with $\mathbf{u} = \frac{\mathbf{x} - \mathbf{x}_i}{h_N}$, $h_N = \frac{h_1}{\sqrt{N}}$, and $h_1 = 0.5$

The gaussian window function corresponds to $\sim \mathcal{N}(\mathbf{x}, h_N^2 \mathbf{I})$.

$N = 4$ for both ω_1 and ω_2 , which results in $h_N = \frac{0.5}{\sqrt{4}} = 0.25$, and $\mathbf{u} = \frac{\mathbf{x} - \mathbf{x}_i}{0.25}$.

As we are working with two dimensions, $l = d = 2$.

As in the last subtask, the discriminant function $g_i(\mathbf{x}) = P(\omega_i)p_i(\mathbf{x}|\omega_i)$.

$$\begin{aligned}
p_n(\mathbf{x}) &= \frac{1}{N} \sum_{i=1}^N \frac{1}{V_N} \phi(\mathbf{u}) \\
&= \frac{1}{4} \sum_{i=1}^N \frac{1}{h_N^2} \frac{1}{(2\pi)^{\frac{2}{2}} |\mathbf{I}|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(\mathbf{u})^T \mathbf{I}^{-1}(\mathbf{u})} \\
&= \frac{1}{8\pi h_N^2} \sum_{i=1}^N e^{-\frac{1}{2}(\frac{\mathbf{x}-\mathbf{x}_i}{0.25})^T \mathbf{I}^{-1}(\frac{\mathbf{x}-\mathbf{x}_i}{0.25})} \\
&= \frac{1}{8\pi h_N^2} \sum_{i=1}^N e^{-8(\mathbf{x}-\mathbf{x}_i)^T (\mathbf{x}-\mathbf{x}_i)} \\
&= \frac{2}{\pi} \sum_{i=1}^N e^{-8(\mathbf{x}-\mathbf{x}_i)^T (\mathbf{x}-\mathbf{x}_i)}
\end{aligned} \tag{2.7}$$

Solving for ω_1 :

$$\begin{aligned}
g_1(\mathbf{x}) &= P(\omega_1) p_n(\mathbf{x}|\omega_1) \\
&= \frac{1}{2} \frac{2}{\pi} \sum_{i=1}^N e^{-8(\mathbf{x}-\mathbf{x}_i)^T (\mathbf{x}-\mathbf{x}_i)} \\
&= \frac{1}{\pi} \left[e^{-8((2.5-2.0)^2 + (2.0-6.0)^2)} + e^{-8((2.5-3.0)^2 + (2.0-4.0)^2)} + \dots + e^{-8((2.5-4.0)^2 + (2.0-6.0)^2)} \right] \\
&= \frac{1}{\pi} \left[e^{-130} + e^{-34} + e^{-290} + e^{-146} \right] \\
&= 5.45554E - 16
\end{aligned} \tag{2.8}$$

Solving for ω_2 :

$$\begin{aligned}
g_2(\mathbf{x}) &= P(\omega_2) p_n(\mathbf{x}|\omega_2) \\
&= \frac{1}{2} \frac{2}{\pi} \sum_{i=1}^N e^{-8(\mathbf{x}-\mathbf{x}_i)^T (\mathbf{x}-\mathbf{x}_i)} \\
&= \frac{1}{\pi} \left[e^{-8((2.5-1.0)^2 + (2.0+2.0)^2)} + e^{-8((2.5-2.7)^2 + (2.0+4.0)^2)} + \dots + e^{-8((2.5-5.0)^2 + (2.0+2.0)^2)} \right] \\
&= \frac{1}{\pi} \left[e^{-146} + e^{-288.32} + e^{-37.12} + e^{-178} \right] \\
&= 2.40901E - 17
\end{aligned} \tag{2.9}$$

Because $g_1(\mathbf{x}) > g_2(\mathbf{x})$, \mathbf{x} is classified to belong to class ω_1 .

c) - A Parzen-window classifier as in the previous subtask, but with $h_1 = 5$. Compare the results and explain the change.

Changing h_1 from 0.5 to 5 results in the following changes:

$$h_N = \frac{5}{\sqrt{4}} = 2.5, \text{ and } \mathbf{u} = \frac{\mathbf{x} - \mathbf{x}_i}{2.5}.$$

The computation is done the same way as the previous subtask.

$$\begin{aligned}
 p_n(\mathbf{x}) &= \frac{1}{N} \sum_{i=1}^N \frac{1}{V_N} \phi(\mathbf{u}) \\
 &= \frac{1}{4} \sum_{i=1}^N \frac{1}{h_N^2} \frac{1}{(2\pi)^{\frac{2}{2}} |\mathbf{I}|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(\mathbf{u})^T \mathbf{I}^{-1}(\mathbf{u})} \\
 &= \frac{1}{8\pi h_N^2} \sum_{i=1}^N e^{-\frac{1}{2}(\frac{\mathbf{x} - \mathbf{x}_i}{2.5})^T \mathbf{I}^{-1}(\frac{\mathbf{x} - \mathbf{x}_i}{2.5})} \\
 &= \frac{1}{8\pi h_N^2} \sum_{i=1}^N e^{-0.08(\mathbf{x} - \mathbf{x}_i)^T (\mathbf{x} - \mathbf{x}_i)} \\
 &= \frac{0.02}{\pi} \sum_{i=1}^N e^{-0.08(\mathbf{x} - \mathbf{x}_i)^T (\mathbf{x} - \mathbf{x}_i)}
 \end{aligned} \tag{2.10}$$

Solving for ω_1 :

$$\begin{aligned}
 g_1(\mathbf{x}) &= P(\omega_1) p_n(\mathbf{x}|\omega_1) \\
 &= \frac{1}{2} \frac{0.02}{\pi} \sum_{i=1}^N e^{-0.08(\mathbf{x} - \mathbf{x}_i)^T (\mathbf{x} - \mathbf{x}_i)} \\
 &= \frac{0.01}{\pi} [e^{-1.30} + e^{-0.34} + e^{-2.90} + e^{-1.46}] \\
 &= 4.04750E - 3
 \end{aligned} \tag{2.11}$$

Solving for ω_2 :

$$\begin{aligned}
 g_2(\mathbf{x}) &= P(\omega_2) p_n(\mathbf{x}|\omega_2) \\
 &= \frac{1}{2} \frac{0.02}{\pi} \sum_{i=1}^N e^{-0.08(\mathbf{x} - \mathbf{x}_i)^T (\mathbf{x} - \mathbf{x}_i)} \\
 &= \frac{0.01}{\pi} [e^{-1.46} + e^{-2.8832} + e^{-0.3712} + e^{-1.78}] \\
 &= 3.65016E - 3
 \end{aligned} \tag{2.12}$$

Because $g_1(\mathbf{x}) > g_2(\mathbf{x})$, \mathbf{x} is classified to belong to class ω_1 .

Increasing the value of h_1 from 0.5 to 5 results in a "smoother" probability density function. This also means that the variance is increased. This is reflected in the answer, as the discriminant functions almost have the same value. The h parameter (also called the smoothing parameter) should most of the time be as small as the data will allow. Setting $h = 50$ would result in $g_1(\mathbf{x}) < g_2(\mathbf{x})$, which justifies the claim.

d) - A kN -nearest neighbourhood classifier where $k_N = 1$.

From the lecture notes, we have that:

$$p_n(\mathbf{x}) = \frac{k_n/n}{V_n} = \frac{k_n}{nV_n} \quad (2.13)$$

with $V_n = \pi r^2$

Obtaining the feature vector for each class:

$$\begin{aligned} R_1 &= \left\{ \left\| \begin{pmatrix} 2 \\ 6 \end{pmatrix} - \begin{pmatrix} 2.5 \\ 2 \end{pmatrix} \right\|, \left\| \begin{pmatrix} 3 \\ 4 \end{pmatrix} - \begin{pmatrix} 2.5 \\ 2 \end{pmatrix} \right\|, \left\| \begin{pmatrix} 3 \\ 8 \end{pmatrix} - \begin{pmatrix} 2.5 \\ 2 \end{pmatrix} \right\|, \left\| \begin{pmatrix} 4 \\ 6 \end{pmatrix} - \begin{pmatrix} 2.5 \\ 2 \end{pmatrix} \right\| \right\} \\ &= \left\{ \sqrt{(-0.5)^2 + (4)^2}, \sqrt{(0.5)^2 + (2)^2}, \sqrt{(0.5)^2 + (6)^2}, \sqrt{(1.5)^2 + (4)^2}, \right\} \\ &= \{4.031, 2.061, 6.021, 4.272\} \end{aligned} \quad (2.14)$$

$$\begin{aligned} R_2 &= \left\{ \left\| \begin{pmatrix} 1 \\ -2 \end{pmatrix} - \begin{pmatrix} 2.5 \\ 2 \end{pmatrix} \right\|, \left\| \begin{pmatrix} 2.7 \\ -4 \end{pmatrix} - \begin{pmatrix} 2.5 \\ 2 \end{pmatrix} \right\|, \left\| \begin{pmatrix} 3.3 \\ 0 \end{pmatrix} - \begin{pmatrix} 2.5 \\ 2 \end{pmatrix} \right\|, \left\| \begin{pmatrix} 5 \\ -2 \end{pmatrix} - \begin{pmatrix} 2.5 \\ 2 \end{pmatrix} \right\| \right\} \\ &= \left\{ \sqrt{(-1.5)^2 + (-4)^2}, \sqrt{(0.2)^2 + (-6)^2}, \sqrt{(0.8)^2 + (-2)^2}, \sqrt{(2.5)^2 + (-4)^2}, \right\} \\ &= \{4.272, 6.003, 2.154, 4.717\} \end{aligned} \quad (2.15)$$

Solving for ω_1 :

$$\begin{aligned} g_1(\mathbf{x}) &= P(\omega_1)p_n(\mathbf{x}|\omega_1) \\ &= \frac{1}{2} \frac{k_n}{nV_n} = \frac{k_n}{2n\pi r^2} \\ &= \frac{1}{8\pi 2.061^2} \\ &= 9.367E - 3 \end{aligned} \quad (2.16)$$

Solving for ω_2 :

$$\begin{aligned} g_2(\mathbf{x}) &= P(\omega_2)p_n(\mathbf{x}|\omega_2) \\ &= \frac{1}{2} \frac{k_n}{nV_n} = \frac{k_n}{2n\pi r^2} \\ &= \frac{1}{8\pi 2.154^2} \\ &= 8.576E - 3 \end{aligned} \quad (2.17)$$

Because $g_1(\mathbf{x}) > g_2(\mathbf{x})$, \mathbf{x} is classified to belong to class ω_1 .

e) - A k_N -nearest neighbourhood classifier where $k_N = 3$.

Using the same feature vectors as in the previous subtask, with the same approach:

Solving for ω_1 :

$$\begin{aligned}
 g_1(\mathbf{x}) &= P(\omega_1)p_n(\mathbf{x}|\omega_1) \\
 &= \frac{1}{2} \frac{k_n}{nV_n} = \frac{k_n}{2n\pi r^2} \\
 &= \frac{3}{8\pi 4.272^2} \\
 &= 6.540E - 3
 \end{aligned} \tag{2.18}$$

Solving for ω_2 :

$$\begin{aligned}
 g_2(\mathbf{x}) &= P(\omega_2)p_n(\mathbf{x}|\omega_2) \\
 &= \frac{1}{2} \frac{k_n}{nV_n} = \frac{k_n}{2n\pi r^2} \\
 &= \frac{3}{8\pi 4.717^2} \\
 &= 5.365E - 3
 \end{aligned} \tag{2.19}$$

Because $g_1(\mathbf{x}) > g_2(\mathbf{x})$, \mathbf{x} is classified to belong to class ω_1 .

Problem 3

Derive the maximum-likelihood-estimate for:

$$\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \quad (3.1)$$

for the case where both $\boldsymbol{\mu}$ and Σ in the multivariate probability density function:

$$p(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{\frac{l}{2}} |\Sigma|^{\frac{1}{2}}} \cdot e^{\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})} \quad (3.2)$$

are unknown.

Defining the likelihood function $L(\boldsymbol{\mu})$

$$L(\boldsymbol{\mu}) = \prod p(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = \prod \frac{1}{(2\pi)^{\frac{l}{2}} |\Sigma|^{\frac{1}{2}}} \cdot e^{\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})} \quad (3.3)$$

Defining the log-likelihood function $l(\boldsymbol{\mu})$

$$\begin{aligned} l(\boldsymbol{\mu}) &= \ln L(\boldsymbol{\mu}) = \ln \prod p(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = \ln \prod \frac{1}{(2\pi)^{\frac{l}{2}} |\Sigma|^{\frac{1}{2}}} \cdot e^{\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})} \\ &= \ln \frac{1}{(2\pi)^{\frac{nl}{2}} |\Sigma|^{\frac{n}{2}}} \cdot e^{\sum_{i=1}^n \frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}_i - \boldsymbol{\mu})} \\ &= -\frac{nl}{2} \ln(2\pi) - \frac{n}{2} \ln|\Sigma| - \frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}_i - \boldsymbol{\mu}) \end{aligned} \quad (3.4)$$

Calculating the derivative of the log-likelihood function:

$$\frac{\partial}{\partial \boldsymbol{\mu}} l(\boldsymbol{\mu}) = -\frac{1}{2} \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\mu}} (\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}_i - \boldsymbol{\mu}) \quad (3.5)$$

Using the following matrix calculus identity:

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \mathbf{A} \mathbf{x} = 2\mathbf{A} \mathbf{x}$$

Which stands true when \mathbf{x} does not depend on \mathbf{A} , and \mathbf{A} is symmetric.

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\mu}} l(\boldsymbol{\mu}) &= -\frac{1}{2} \sum_{i=1}^n 2\Sigma^{-1}(\mathbf{x}_i - \boldsymbol{\mu}) \\ &= -\sum_{i=1}^n (\Sigma^{-1} \mathbf{x}_i - \Sigma^{-1} \boldsymbol{\mu}) \\ &= n\Sigma^{-1} \boldsymbol{\mu} - \sum_{i=1}^n \Sigma^{-1} \mathbf{x}_i \end{aligned} \quad (3.6)$$

Finding the maxima by setting $\frac{\partial}{\partial \boldsymbol{\mu}} l(\boldsymbol{\mu}) = 0$:

$$\begin{aligned} n\Sigma^{-1}\hat{\boldsymbol{\mu}} - \sum_{i=1}^n \Sigma^{-1}\mathbf{x}_i &= 0 \\ n\Sigma^{-1}\hat{\boldsymbol{\mu}} &= \sum_{i=1}^n \Sigma^{-1}\mathbf{x}_i \\ \hat{\boldsymbol{\mu}} &= \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \end{aligned} \tag{3.7}$$

Q.E.D