Pattern Recognition and Machine Learning

Assignment - 05

July 21, 2025

Name : Rohan Baghel
Student ID: 202116011

About

Normal Distribution. Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean.

Optimal Decision Boundary he optimal decision boundary represents the "best" solution possible for that problem. Consequently, by looking at the complexity of this boundary and at how much error it produces, we can get an idea of the inherent difficulty of the problem.

Generate the datasets A and B in R2 with each of them consisting 2000 data points from normal distribution.
The dataset A and B has been drawn from the N $(\mu_1,\Sigma_1) and N(\mu_2,\Sigma_2)$.

Let us fix the $\mu_1 = [-1, 1] and \mu_2 = [1, 1]$.

```
clc;
clearvars;
close all;
%clear function;
%-----------------------------------------------------------------------------%
% Generating dataset
m_1 = [-1;1];
m_2 = [1;1];
mat_1 = [0.6 0; 0 0.6];

A = mvnrnd(m_1,var1,2000);
B = mvnrnd(m_2,var1,2000);

%-----------------------------------------------------------------------------%
```

a)

Find the optimal decision boundary for the classification of the dataset A and B using $\Sigma_1 = \Sigma_2 = \begin{bmatrix} 0.6 & 0 \\ 0 & 0.6 \end{bmatrix}$
Plot the dataset A and B with different colors and plot the obtained optimal decision boundary.
Comment on the characteristics of obtained decision boundary.

```
%-----------------------------------------------------------------------------%
mat_1 = [0.6 0; 0 0.6];
A = mvnrnd(m_1,mat_1,2000);
B = mvnrnd(m_2,mat_1,2000);

wt_1 = (1/mat_1(1,1))*m_1;
wt_2 = (1/mat_1(2,2))*m_2;
wt_11 = (-1/(2*mat_1(1,1)))*(m_1')*m_1;
wt_12 = (-1/(2*mat_1(2,2)))*(m_2')*m_2;
x = ((wt_1')-(wt_2'))\(wt_12 - wt_11);
figure();
scatter(A(:,1),A(:,2),'.');
hold on;
scatter(B(:,1),B(:,2),'.');
hold on;
plot([x(1) x(1)],[-3.5 3.5],'black',linewidth=1.5);
title('Case - 1 ');
legend('A','B','Decision Boundary')
%-----------------------------------------------------------------------------%
```

Output

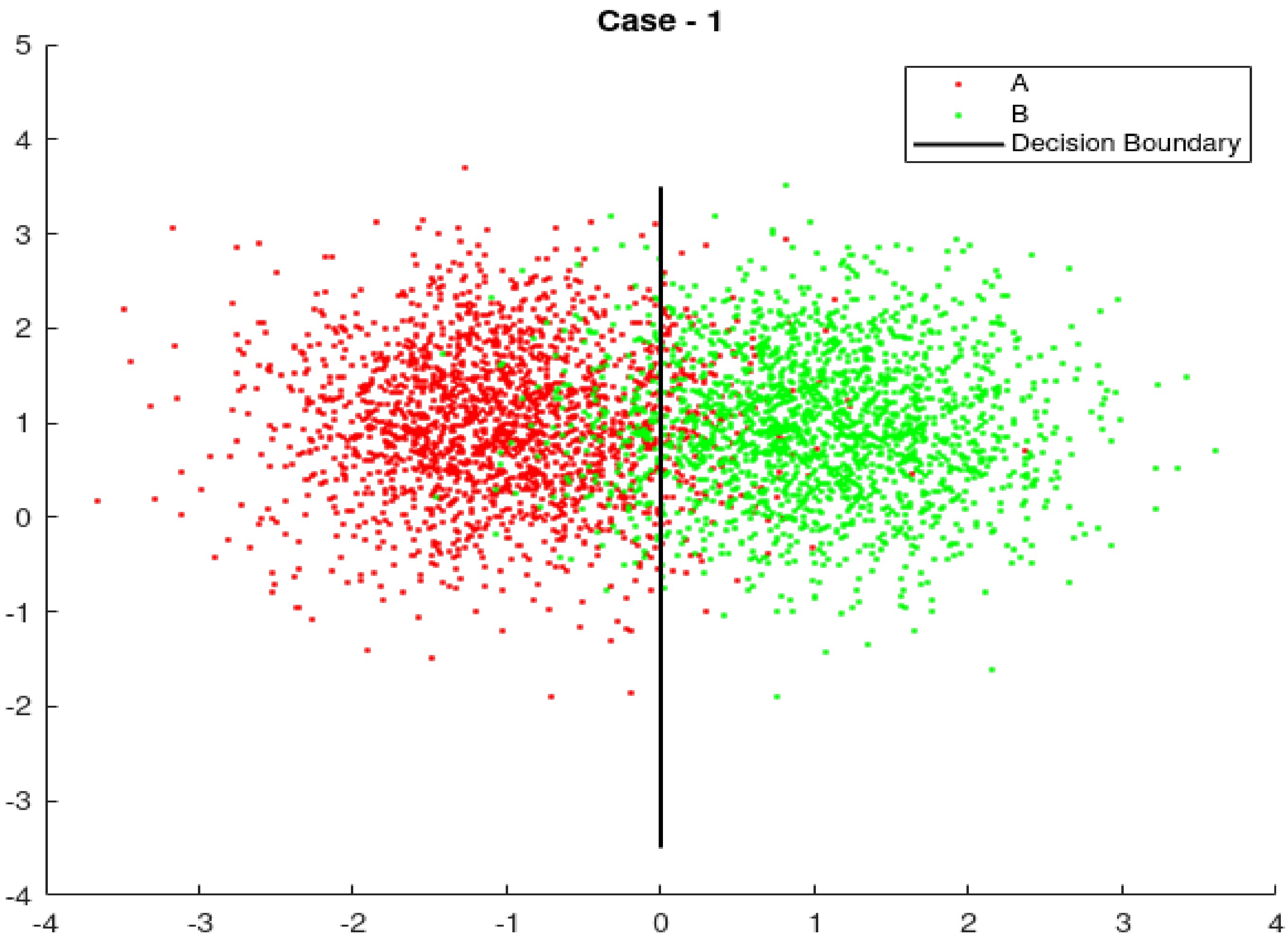

Figure 1: Case 1

For Σ_1 and $\Sigma_2 = \begin{bmatrix} 0.6 & 0 \\ 0 & 0.6 \end{bmatrix}$

- Main diagonal: variances for each individual variable.
- Off-diagonal: covariances of each variable pairing i j (note: values are repeated, as matrix is symmetric)

Off-diagonal entries with a value of 0 indicate uncorrelated variables, that are statistically independent (variables likely do not influence one another)

- Decision boundary goes through x0 along line between means, orthogonal to this line

- If priors equal, x0 between means (minimum distance classifier), otherwise x0 shifted
- If variance small relative to distance between means, priors have limited effect on boundary location

We can look that a vertical line is formed as diagonal values are same.

b)

Find the optimal decision boundary for the classification of the dataset A and B using $\Sigma_1 = \Sigma_2 = \begin{bmatrix} 0.7 & 0 \\ 0 & 0.3 \end{bmatrix}$

Plot the dataset A and B with different colors and plot the obtained optimal decision boundary.

Comment on the characteristics of obtained decision boundary.

```

%-----------------------------------------------------------------------%

mat_2=[0.7 0; 0 0.3];

Ab = mvnrnd(m_1,mat_2,2000);
Bb = mvnrnd(m_2,mat_2,2000);

w=(mat_2)\(m_1-m_2);

x0=1/2*(m_1+m_2);


wt_1 = (mat_2)\m_1;
wt_2 = (mat_2)\m_2;

wt_11 = -0.5*(m_1')*((mat_2)\m_1);
wt_12 = -0.5*(m_2')*((mat_2)\m_2);

x2 = ((wt_1')-(wt_2'))\(wt_12 - wt_11);

figure();
scatter(Ab(:,1),Ab(:,2),'.');
hold on;
scatter(Bb(:,1),Bb(:,2),'.');
hold on;
plot([x2(1) x2(1)],[-3.5 3.5],'black',linewidth=1.5);
title('Case - 2');
legend('A','B','Decision Boundary')
%-----------------------------------------------------------------------%
```

Output

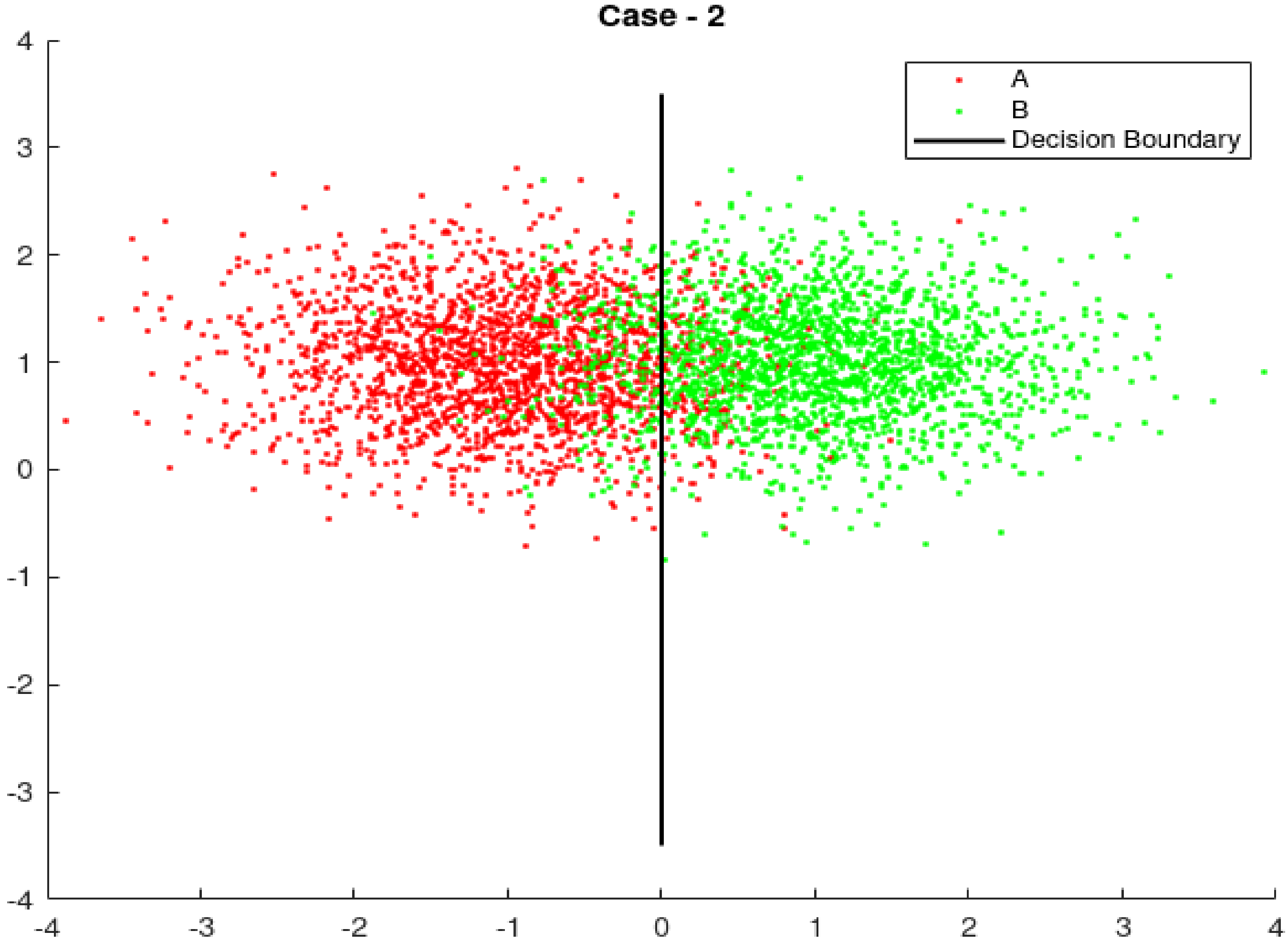

Figure 2: Case 2

As for Case I, passes through point x0 lying on the line between the two class means.Again, x0 in the middle if priors identical

In this graph we can see that optimal decision boundary is same as in the case 1 graph
here
Σ_1 and $\Sigma_2 = \begin{bmatrix} 0.7 & 0 \\ 0 & 0.3 \end{bmatrix}$

c)

Find the optimal decision boundary for the classification of the dataset A and B using $\Sigma_1 = \Sigma_2 = \begin{bmatrix} 0.6 & 0.25 \\ 0.25 & 0.4 \end{bmatrix}$

Plot the dataset A and B with different colors and plot the obtained optimal decision boundary.

Comment on the characteristics of obtained decision boundary.

```

%------------------------------------------------------------------------%

mat_3=[0.6 0.25; 0.25 0.4];

Ac = mvnrnd(m_1,mat_3,2000);
Bc = mvnrnd(m_2,mat_3,2000);

wt_1 = (mat_3)\m_1;
wt_2 = (mat_3)\m_2;

w = (wt_1'-wt_2');

wi1=w(1);
wi2=w(2);

wt_11 = -0.5*(m_1')*((mat_3)\m_1);
wt_12 = -0.5*(m_2')*((mat_3)\m_2);

w0=wt_12-wt_11;

x3_1 = linspace(-5,5,100);
x3_2 = (w0 - wi1*x3_1)/wi2;

figure();
scatter(Ac(:,1),Ac(:,2),'.');
hold on;
scatter(Bc(:,1),Bc(:,2),'.');
hold on;
plot(x3_1,x3_2,'black',linewidth=1.5);
title('Case - 3');
legend('A','B','Decision Boundary')
%------------------------------------------------------------------------%
```

Output

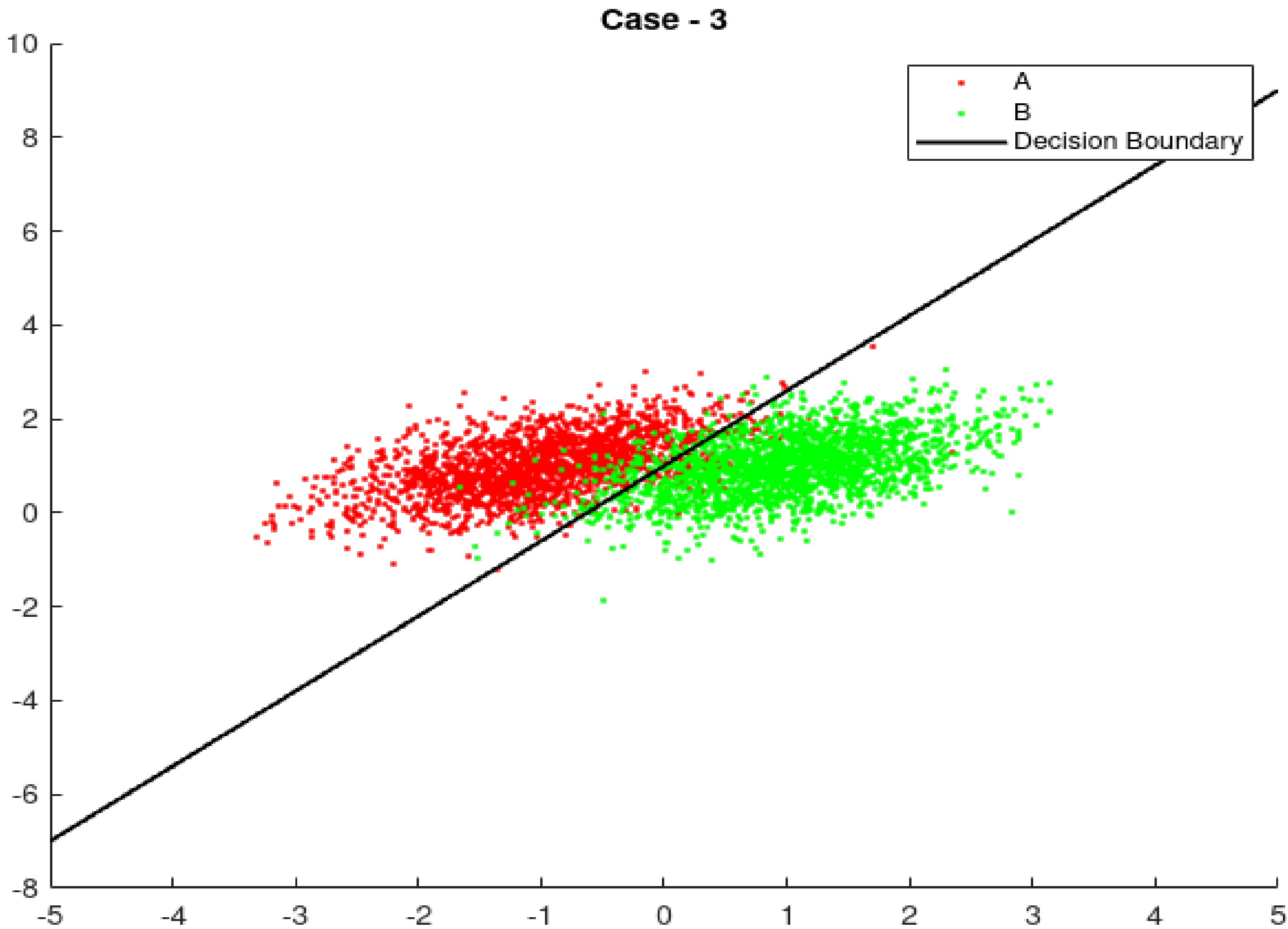

Figure 3: Case 3

For Σ_1 and $\Sigma_2 = \begin{bmatrix} 0.6 & 0.25 \\ 0.25 & 0.4 \end{bmatrix}$
In this graph we can observe that a tilled boundary is formed between the two classifications
as in this Covariance is more than the 0 'zero' all though they are same.

It should be apparent that because of the overlap in these distributions, any decision rule will necessarily misclassify some observations fairly often.

boundary runs through the points where the contours of the two conditional distributions intersect. These points of intersection are where the classes have equal posterior probability.