# Exam #2

Isaac Ayala Lozano

April 27, 2020

## 1 Introduction

The results of a Bayesian multivariate classifier are presented. An explanation of its development is provided. The classifier was trained to work with three categories $C_i$ and a feature vector of size 2.

The classifier was trained using the Iris flower data set [2], available from the UCI Machine Learning Repository [1]. The dataset describes five properties of three different flower species: petal length, petal width, sepal length, sepal width and variety. Whilst the first four are quantitative traits, the last one is qualitative. This trait describes the type flower associated with the entry: Iris setosa, Iris versicolor and Iris virginica. Figure 1 presents the species in the dataset.

(a) Iris versicolor (b) Iris virginica (c) Iris setosa

Figure 1: Species in the iris dataset

[1] https://archive.ics.uci.edu/ml/datasets/Iris

## 2 Features in the data set

From the data set, two groups of features are identified: length and width for petal and sepal. To develop a classifier of size two, it is necessary to select the two entries from the feature vector that will be used in the classifier. Observing the grouping of datapoints for length and width (Figure 2) it becomes evident that the petal traits are ideal for the classifier. This is due to how the data is already grouped into three easily identifiable groups.

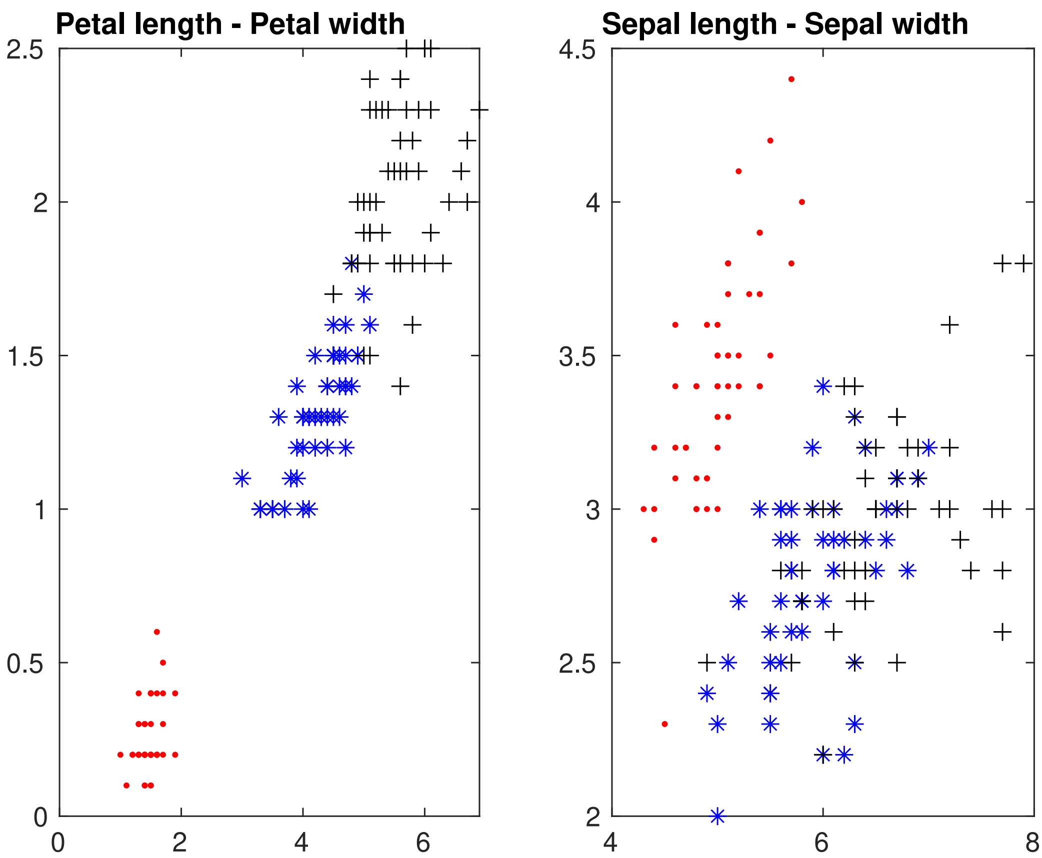


Figure 2: Data clusters

## 3 Gaussian Distributions

The behaviour of all quantitative traits of the dataset is assumed to adhere to a Gaussian Distribution. This assumption is taken due to how the values of the data set are presented. As shown in Figures 3 and 4 the distribution for each flower variety follows the general behaviour of the Gaussian Distribution.

Having confirmed that all quantitative features of the dataset adhere to Gaussian Distributions, it is possible to proceed with the development of the classifiers.

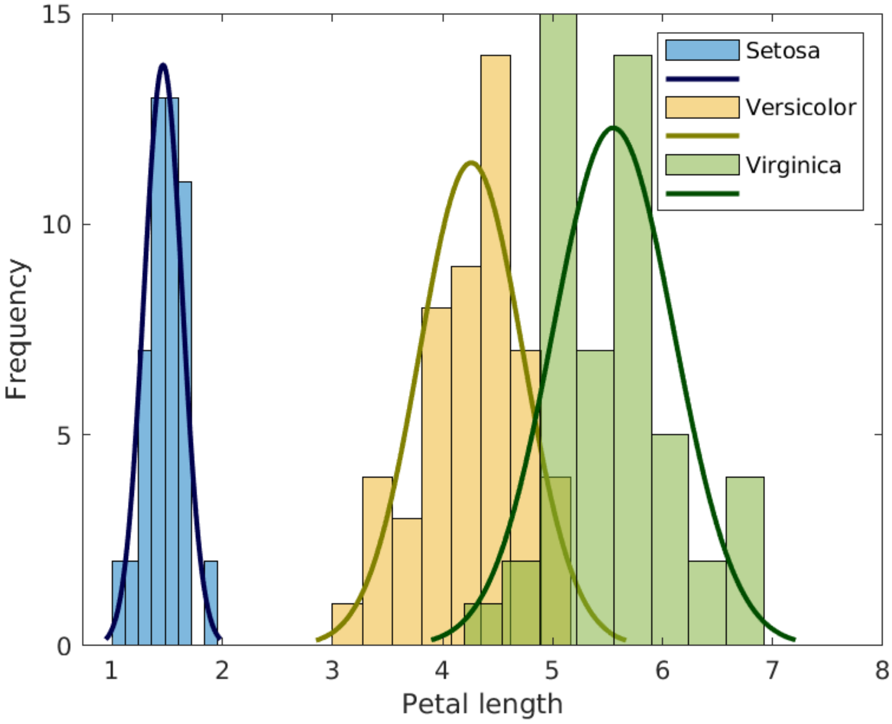


Figure 3: Petal Length

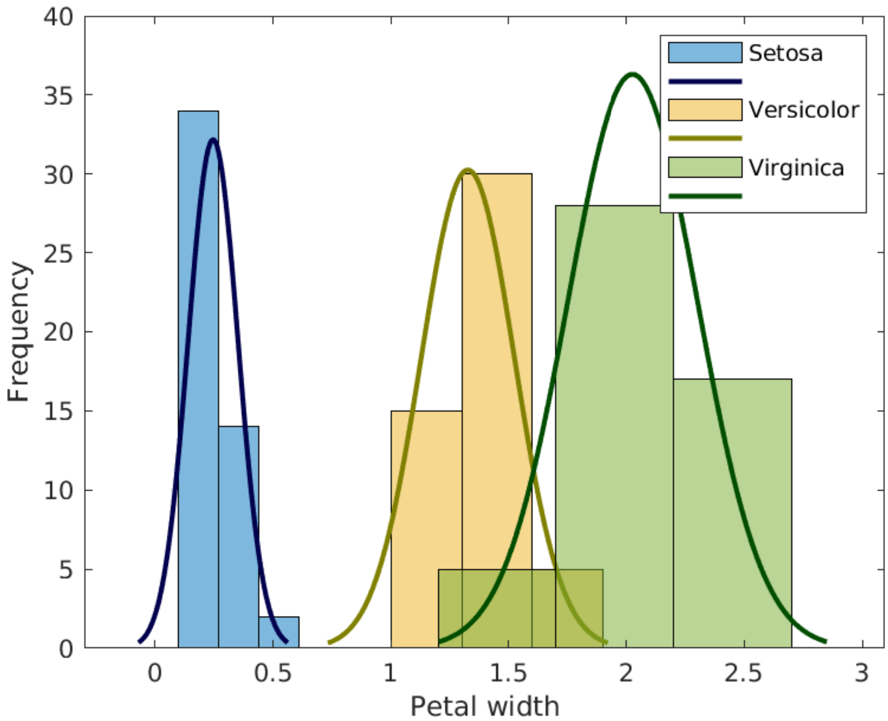


Figure 4: Petal Width

# 4 Classifier

Classifiers were developed for three cases.

- Classifier with $\Sigma = \sigma^2 \cdot \mathbf{I}$
- Classifier with $\Sigma = \Sigma_i$
- Classifier with arbitrary $\Sigma$

## 4.1 Case A

For the first case, $\sigma$ was assigned as $\sigma = 0.25$. Given this restriction, the multivariate PDF was obtained, as shown in Figure 5

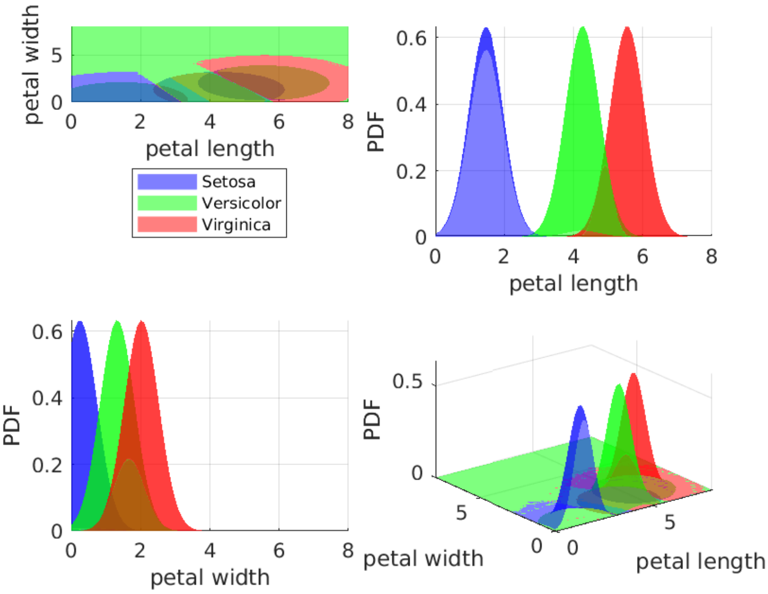


Figure 5: PDF for Case A

$$
\begin{aligned}
g_i(x) &= -\frac{\|x-\mu_i\|^2}{2\sigma^2} + \ln P(\omega_i) \\
\|x-\mu_i\|^2 &= (x-\mu_i)^T(x-\mu_i) \\
g_i(x) &= w_i^T x + \omega_{i0} \\
w_i &= \frac{1}{\sigma^2}\mu \\
w_{i0} &= \frac{-1}{2\sigma^2}\mu_i^T\mu_i + nP(\omega_i)
\end{aligned}
$$

Afterwards, the *a posteriori* probability is obtained and the boundaries of the classifier are obtained. The results are shown in Figure 6.

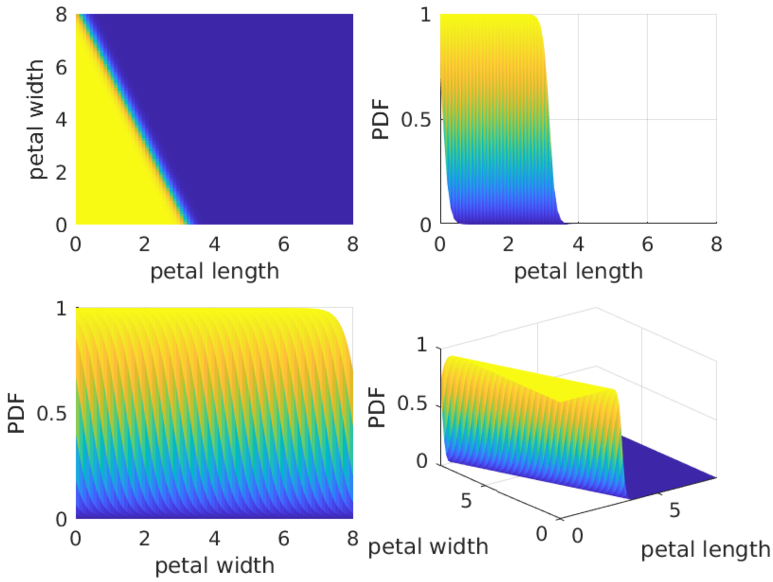


Figure 6: Case A - A posteriori probability

The classifier was tested against a sample of 75 entries from the dataset. Each entry was assigned a colour based on the assigned variety: yellow for Setosa, black for Versicolor and magenta for Virginica. This is presented in Figure 7.

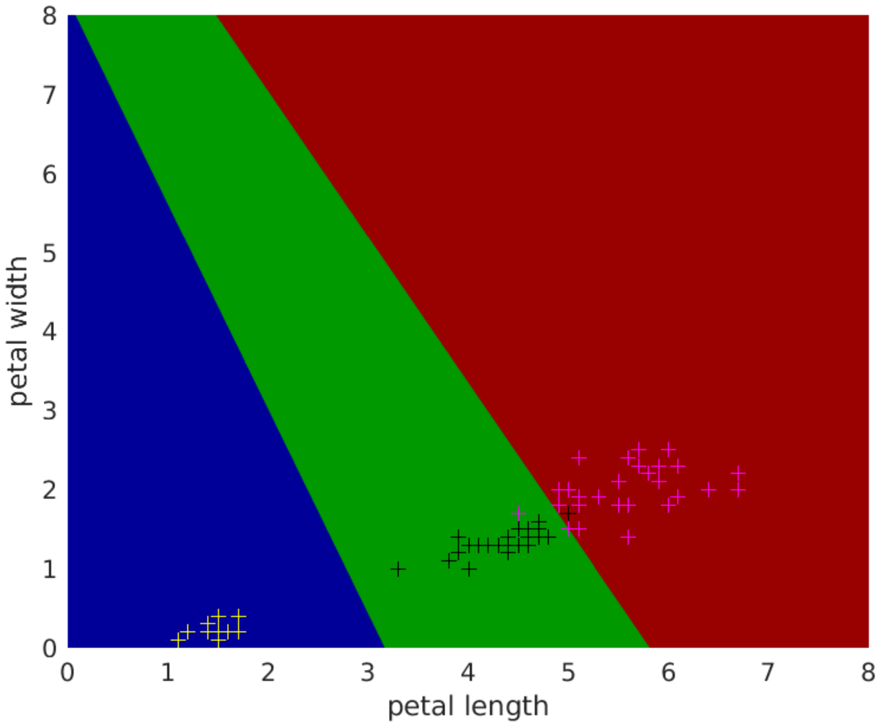


Figure 7: Case A - Boundaries

### 4.2 Case B

For the first case, $\Sigma$ was assigned as

$$\begin{pmatrix} 0.26 & 0.04 & 0.02 & 0.01 \\ 0.04 & 0.22 & 0.03 & 0.02 \\ 0.02 & 0.03 & 0.15 & 0.15 \\ 0.01 & 0.02 & 0.15 & 0.31 \end{pmatrix}$$

This new covariance matrix influenced the multivariate gaussian distribution of the system, as seen in Figure 8.

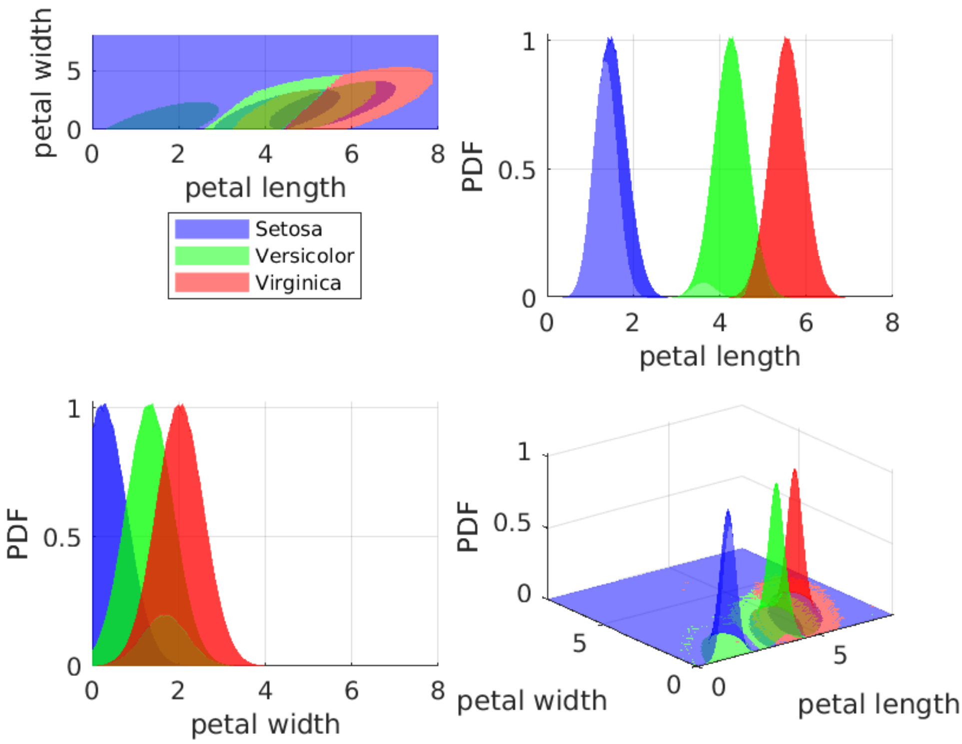


Figure 8: PDF for Case B

Afterwards, the *a posteriori* probability is obtained and the boundaries of the classifier are obtained. The results are shown in Figure 9.

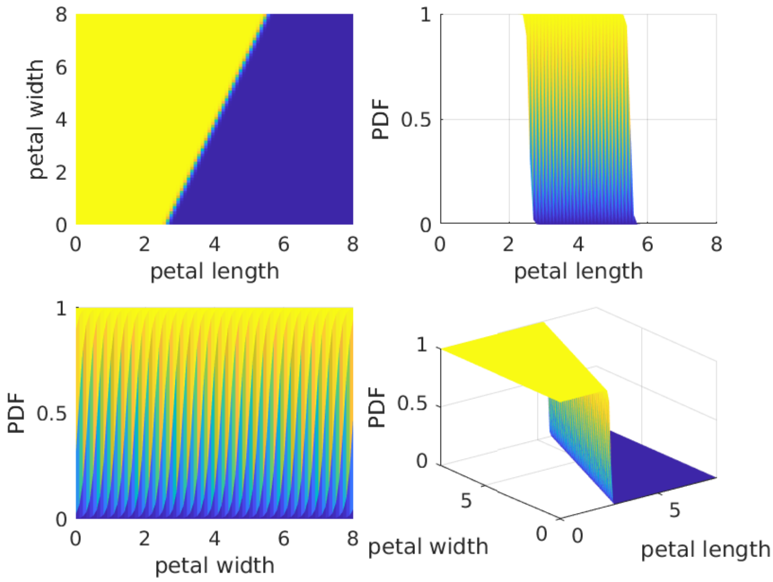


Figure 9: Case B - A posteriori probability

$$
\begin{aligned}
g_i(x) &= x^T W_i x + w_i^T x + \omega_{i0} \\
W_i &= -\frac{1}{2}\Sigma_i^{-1} \\
w_i &= \Sigma_i^{-1}\mu_i \\
\omega_{i0} &= -\frac{1}{2}\mu_i^T \Sigma_i^{-1}\mu_i - \frac{1}{2}\ln\det(\Sigma_i) + \ln P(\omega_i)
\end{aligned}
$$

The classifier was tested against a sample of 75 entries from the dataset. Each entry was assigned a colour based on the assigned variety: yellow for Setosa, black for Versicolor and magenta for Virginica. This is presented in Figure 10.

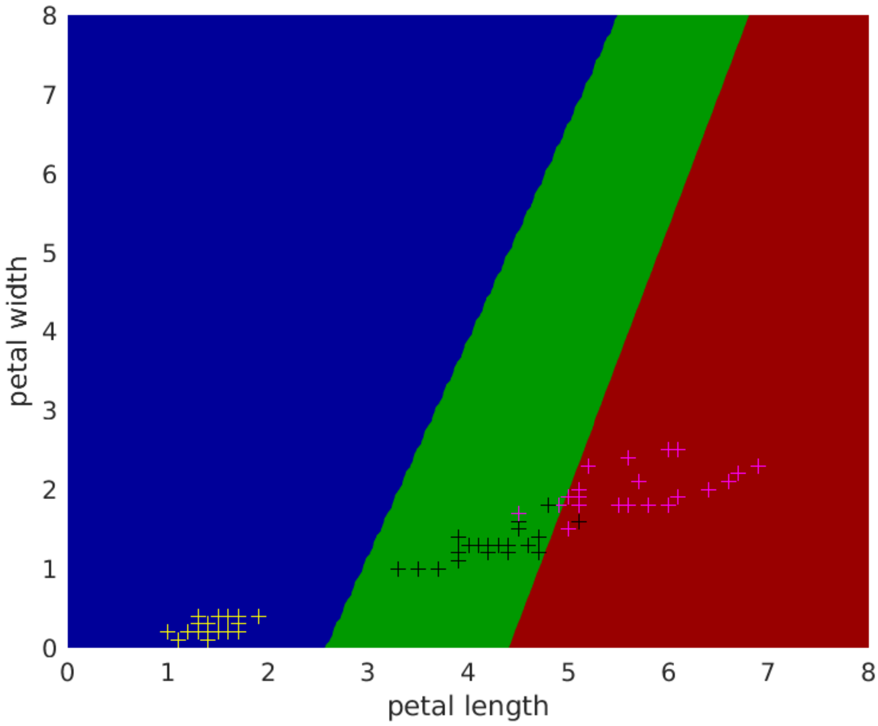


Figure 10: Case B - Boundaries

## 4.3 Case C

For the third case, $\Sigma$ was left untouched, allowing the covariance matrix to reflect the nature of the data available. This is observed in Figure 11.

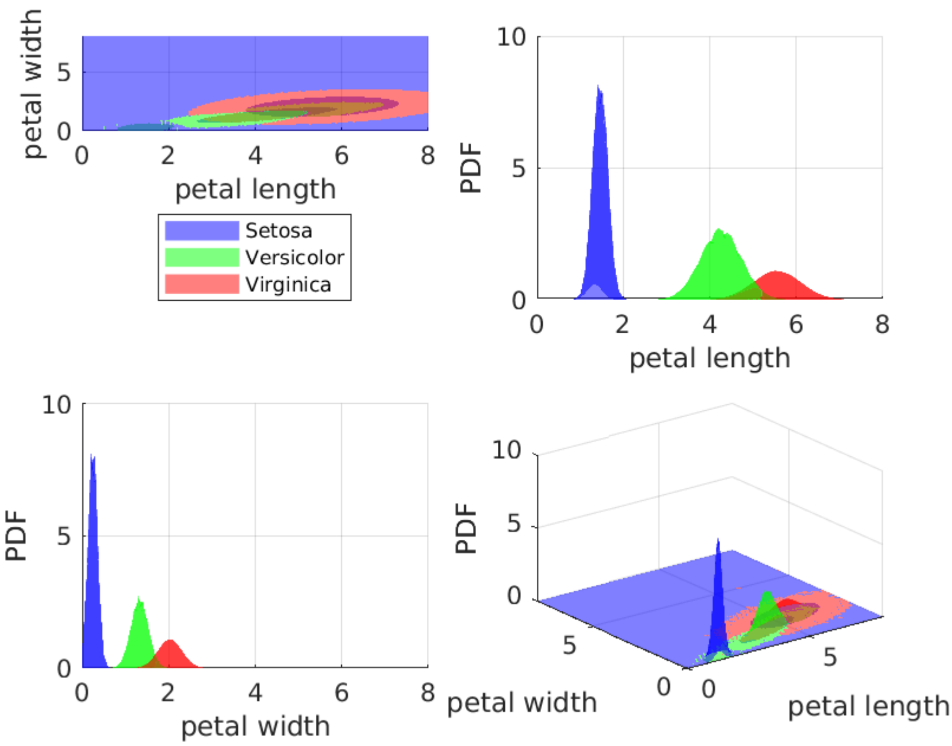


Figure 11: PDF for Case C

Afterwards, the *a posteriori* probability is obtained and the boundaries of the classifier are obtained. The results are shown in Figure 12.

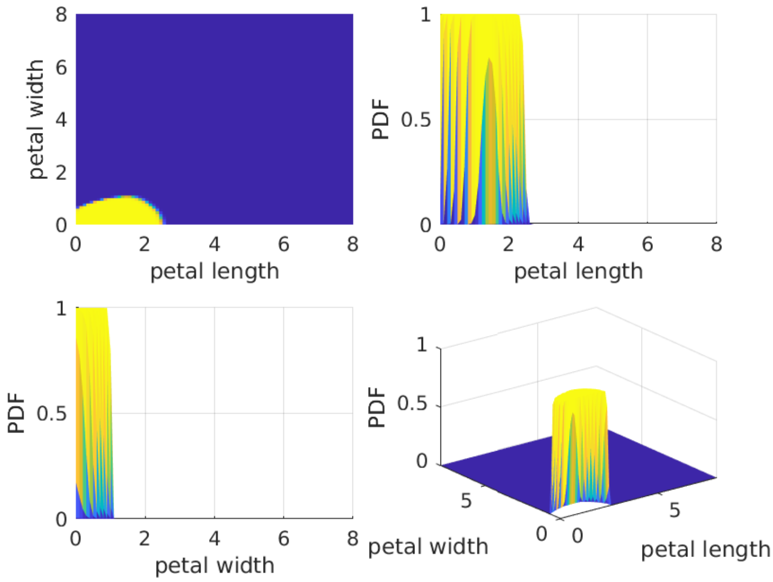


Figure 12: Case C - A posteriori probability

$$
\begin{aligned}
g_i(x) &= -\frac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu) + \ln P(\omega_i) \\
g_i(x) &= w_i^T x + \omega_{i0} \\
w_i &= \Sigma^{-1}\mu_i \\
w_{i0} &= \frac{-1}{2}\mu_i^T\Sigma^{-1}\mu_i + nP(\omega_i)
\end{aligned}
$$

The classifier was tested against a sample of 75 entries from the dataset. Each entry was assigned a colour based on the assigned variety: yellow for Setosa, black for Versicolor and magenta for Virginica. This is presented in Figure 13.

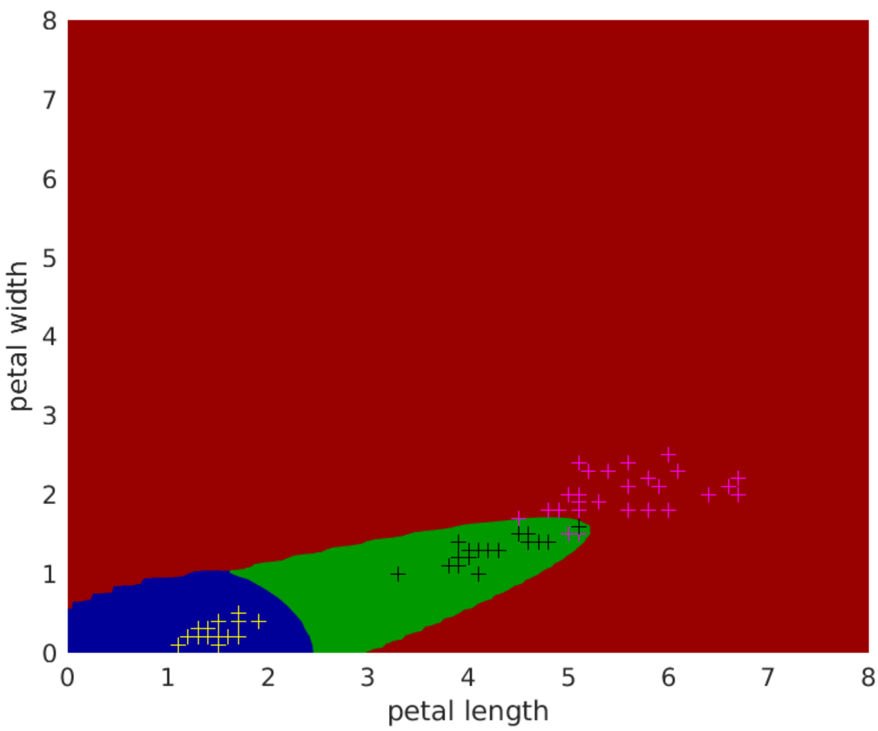


Figure 13: Case C - Boundaries

# 5 Errors

The error boundaries for the classifiers were obtained for both the Chernoff and Bhattacharyya Bounds [1].

For the Chernoff Bounds, the following equations were employed.

$$\begin{aligned}
&\min[a, b] \leq a^{\beta} b^{1-\beta} \quad \text{for } a, b \geq 0 \text{ and } 0 \leq \beta \leq 1 \\
&P(\text{error}) \leq P^{\beta}(\omega_1) P^{1-\beta}(\omega_2) \int p^{\beta}(x|\omega_1) p^{1-\beta}(x|\omega_2) dx \text{ for } 0 \leq \beta \leq 1 \\
&\quad \int p^{\beta}(x|\omega_1) p^{1-\beta}(x|\omega_2) dx = \exp(-k(\beta)) \\
&\quad k(\beta) = \frac{\beta(1-\beta)}{2} (\mu_2 - \mu_1)^T \det(\beta\Sigma_1 + (1-\beta)\Sigma_2)^{-1} (\mu_2 - \mu_1) \\
&\qquad\quad + \frac{1}{2} \ln \frac{\det(\beta\Sigma_1 + (1-\beta)\Sigma_2)}{\det(\Sigma_1)^{\beta} \det(\Sigma_2)^{1-\beta}}
\end{aligned}$$

Likewise, for the Bhattacharyya Bounds another set of equations was utilized

$$\begin{aligned}
P(\text{error}) &\leq \sqrt{P\omega_1 P(\omega_2)} \int \sqrt{p(x|\omega_1) p(x|\omega_2)} dx \\
&= \sqrt{P\omega_1 P(\omega_2)} \exp -k(1/2) \\
k(1/2) &= \frac{1}{8} (\mu_2 - \mu_1)^T \det(\frac{\Sigma_1 + Sigma_2}{2})^{-1} (\mu_2 + mu_1) \\
&+ \frac{1}{2} \ln \frac{\det(\frac{\Sigma_1 + \Sigma_2}{2})}{\sqrt{\det(\Sigma_1) \det(\Sigma_2)}}
\end{aligned}$$

## 5.1 Case A

For this case, the Chernoff bounds were determined to be

| | |
|---|---|
| $P_{12}$ | 0.3712% |
| $P_{23}$ | 11.3240% |
| $P_{31}$ | 0.0016% |

Table 1: Case A - Chernoff Bound

Likewise the Bhattacharyya bounds were determined

The experimental error was also examined for the data set.

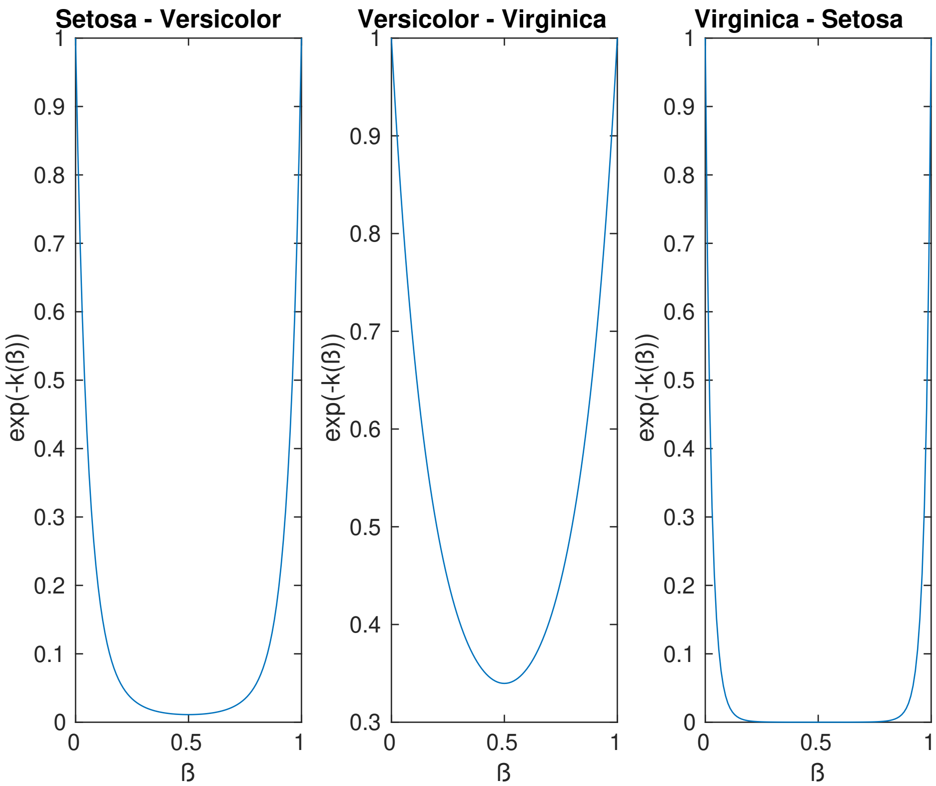

Figure 14: Case A - Error bound

| | |
|---|---|
| $P_{12}$ | 0.3712% |
| $P_{23}$ | 11.3240% |
| $P_{31}$ | 0.0016% |

Table 2: Case A - Bhattacharyya Bound

| | |
|---|---|
| $P_{12}$ | 0.000% |
| $P_{23}$ | 4.000% |
| $P_{31}$ | 0.000% |

Table 3: Case A - Experimental error

## 5.2 Case B

For this case, the Chernoff bounds were determined to be

| | |
|---|---|
| $P_{12}$ | 0.0049% |
| $P_{23}$ | 6.3073% |
| $P_{31}$ | 0.0000% |

Table 4: Case B - Chernoff Bound

Likewise the Bhattacharyya bounds were determined

| | |
|---|---|
| $P_{12}$ | 0.0049% |
| $P_{23}$ | 6.3073% |
| $P_{31}$ | 0.0000 |

Table 5: Case B - Bhattacharyya Bound

The experimental error was also examined for the data set.

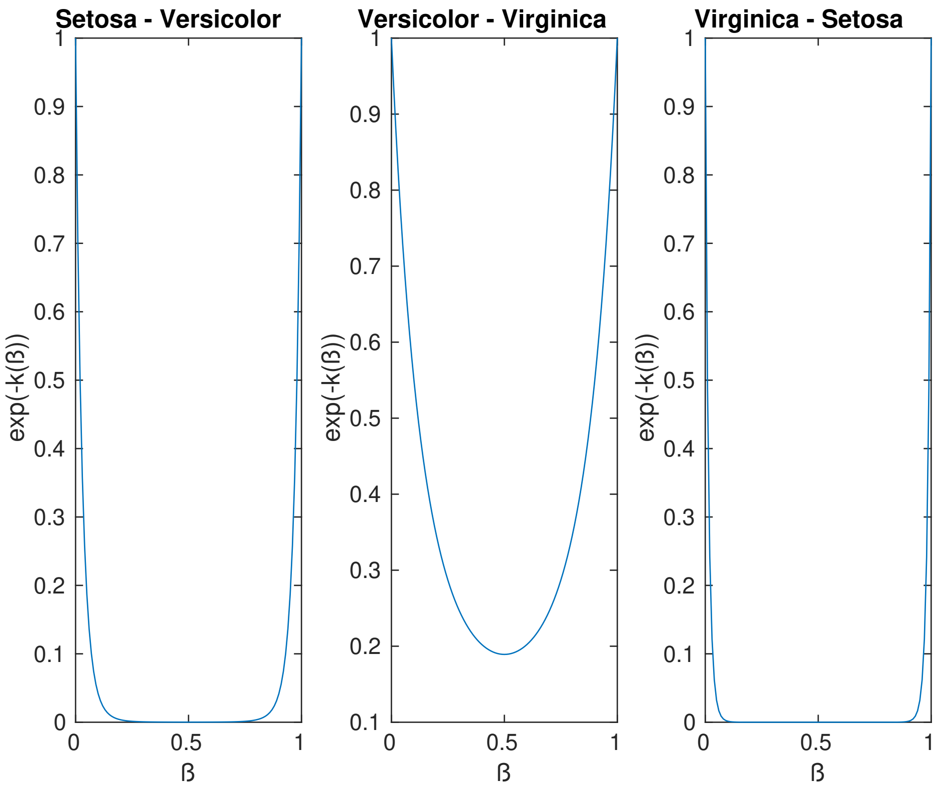


Figure 15: Case B - Error bound

| | |
|---|---|
| $P_{12}$ | 0.000% |
| $P_{23}$ | 8.000% |
| $P_{31}$ | 0.000% |

Table 6: Case B - Experimental error

### 5.3 Case C

For this case, the Chernoff bounds were determined to be

| | |
|---|---|
| $P_{12}$ | 0.0012% |
| $P_{23}$ | 7.9909% |
| $P_{31}$ | 0.0000% |

Table 7: Case C - Chernoff Bound

Likewise the Bhattacharyya bounds were determined

| | |
|---|---|
| $P_{12}$ | 0.0074% |
| $P_{23}$ | 8.0726% |
| $P_{31}$ | 0.0000% |

Table 8: Case C - Bhattacharyya Bound

The experimental error was also examined for the data set.

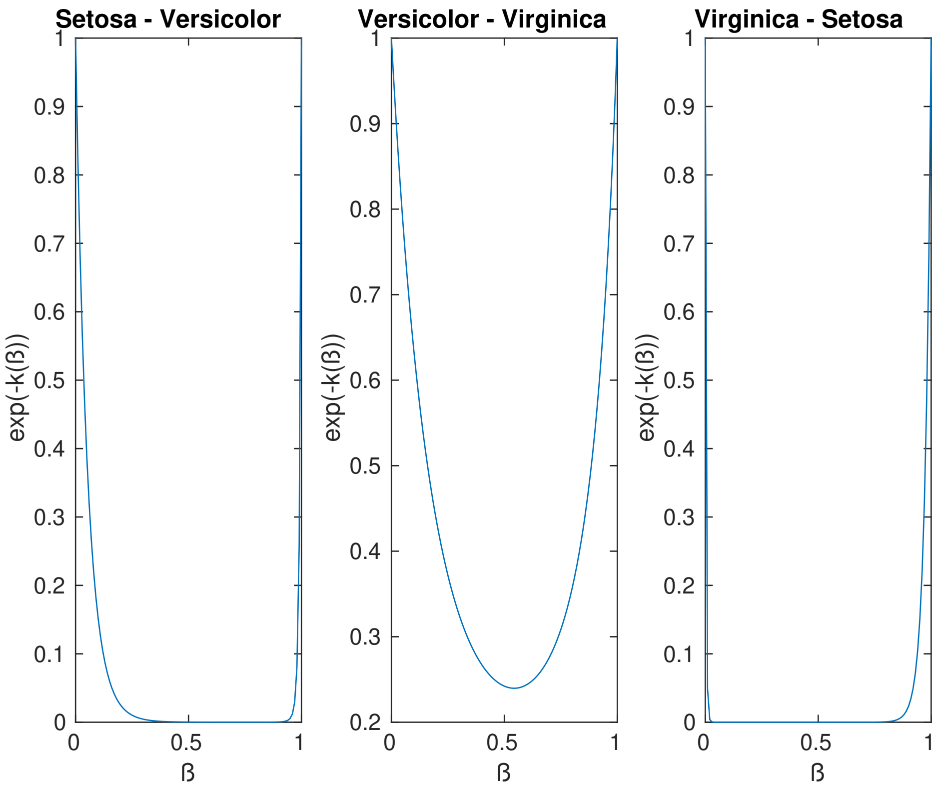

Figure 16: Case C - Error bound

| | |
|---|---|
| $P_{12}$ | 0.000% |
| $P_{23}$ | 3.333% |
| $P_{31}$ | 0.000% |

Table 9: Case C - Experimental error

# References


[1] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. Wiley, 2012. ISBN: 9781118586006. URL: `https://books.google.com.mx/books?id=Br33IRC3PkQC`.

[2] Ronald Fisher. "The Use of Multiple Measurements in Taxonomic Problems". In: *Annals Eugen.* 7 (Jan. 1936), pp. 179–188. DOI: `10.1111/j.1469-1809.1936.tb02137.x`.