Lab 1: Bayesian Decision Theory

Students:

Gurvir Parmar, 500765786

Nicholas An, 500750378

Vatsal Shreekant, 500771363

Part 1

Q-1: Using a single discriminant function $g(x_2)$, design a 2 class minimum-error-rate classifier (dichtomizer) from the given data, to classift IRIS samples into to either *Iris Setosa* or *Iris Versicolor*, according to the feature: *Sepal Width*.

The sepal width data was obtained by running the command 'runlab1' and this extracted the training set's column 2. There were 100 entries within column 2, where the entries were evenly divided into two class. The first 50 rows belonged to Setosa while the last 50 rows to Versicolour. To create a dichotomizer, the Normal distribution was computed and plotted for each class. This was achieved by calculating the mean and standard deviation for the first 50 rows (class Setosa), and last 50 rows (class Versicolor). Upon calculating the mean and standard deviations, the normal distributions were calculated and plotted.

Q-2: Using the shell program *lab1.m*, write a program that will take individual sample values as the input and will return the *posteriror probabilities* and the values of $g(x_2)$.

To calculate the posterior probabilities, the prior probabilities were determined first. For the following command 'lab1(1,trainingSet,2)', the values of the posterior probabilities were: [**0.0119 0.9881].** The value of the returned g(x2) was: **-0.9956**

Q-3: Identify the class labels for the feature values using your program, and indicat their respective posterior probabilities and discriminant values: $x_1 = [3.3, 4.4, 5.0, 5.7, 6.3]$.

Values	Posterior Probabilities	Discriminant Values	Test Features
3.3	$[0.8467, 0.1533]$	0.6935	*Setosa*
4.4	$[0.9655, 0.0345]$	0.9310	*Setosa*
5.0	$[0.8834, 0.1166]$	0.7669	*Setosa*
5.7	$[0.1897, 0.8103]$	-0.6207	*Versicolor*
6.3	$[0.0022, 0.9978]$	-0.9956	*Versicolor*

Q-4: Arrive at a optimal threshold (T_{h_1}) that seperates classes ω_1 and ω_2 (theoretically or experimentally). Justify your results.

Through the given code, the optimal threshold that separates the classes was derived. As a result the following graphs of the conditional probability for Sepal Length for Versicolour vs. Sertosa was created.

In addition it can be seen as the Distribution of Versicolour and Sertosa for Sepal Length was also shown below. The 'Distribution of Versicolour and Sertosa for Sepal Length and Width' histograms show that the Optimum threshold was found to be approximately x=5.4 and y=0.5 for Sepal Length and x=3.1 and y=0.73 for Sepal Width.

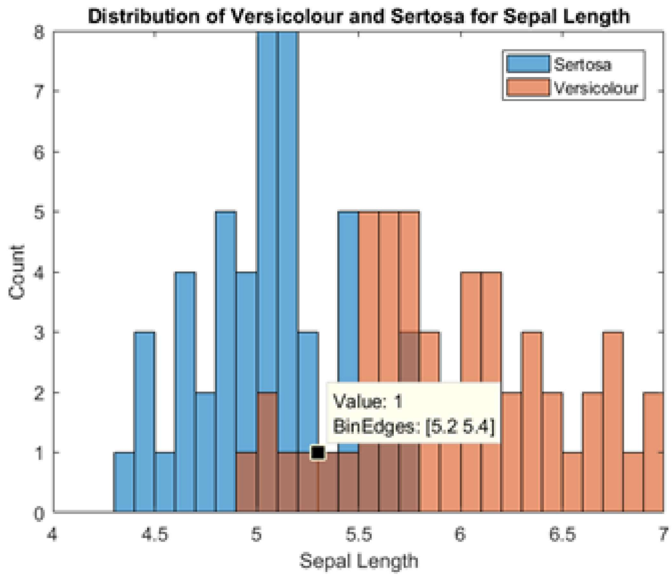

Figure 1: Distribution of classes for Sepal Length

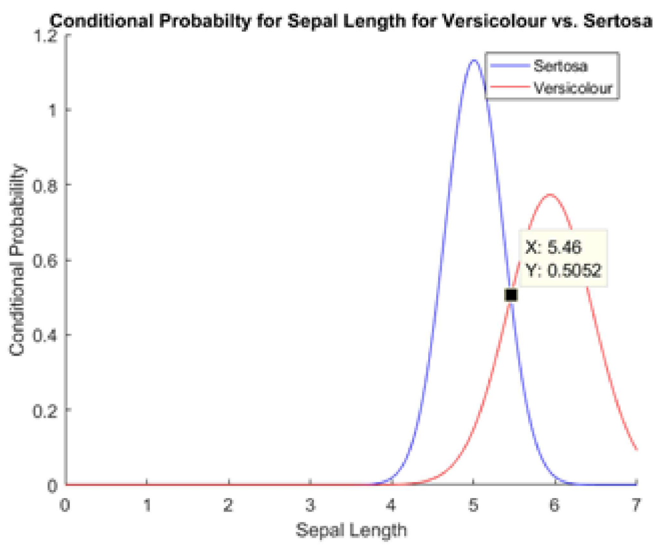

Figure 2: Conditional Probability for Sepal Length

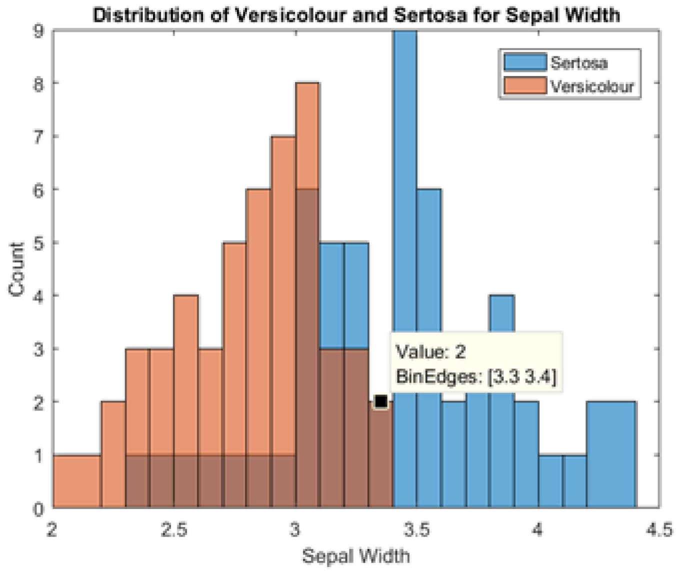

Figure 3: Distribution of classes for Sepal Width

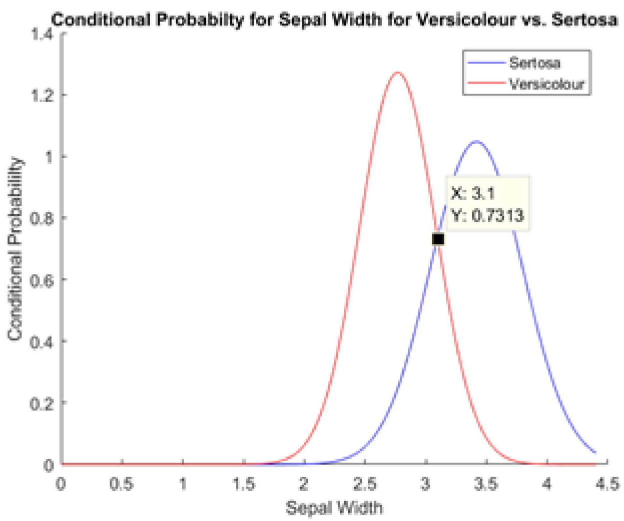

Figure 4: Conditional Probability for Sepal Width

Q-5: Suggest how T_{h1} would be affected if a higher penalty is associated with classifying class ω_2 as class ω_1 - show with experiment.

Using the 'Conditional Probabilty for Sepal Width for Versicolour vs. Sertosa' figure 4 graph as a reference, the function g(x) would need to be shifted to the left to neglect the high penalty associated with classifying versicolor as sertosa. Translating the function g(x) to the left from the original position of x=3.1 will balance off the high penalty of classifying class w2 as class w1. This process however, also leads to mislabel sertosa as versicolor.

Q-6: Adjust your program to accept *Sepal Length* as the discriminating feature $g(x_1)$. Suggest which of the two features (x_1, x_2) might be a better choice for sperating the two classes ω_1 and ω_2. Justify.

The code snippet from question 3 was used to find the posterior probabilities and the discriminant function values using feature sepal length (x1). Sepal Length as well as Sepal Width discriminant features perform well individually but not together; refer to the figure below, they do not intersect. It would be a better choice to use Sepal Width if the penalty was higher with classifying Versicolour as Sertosa, thus making it the more reliable feature.

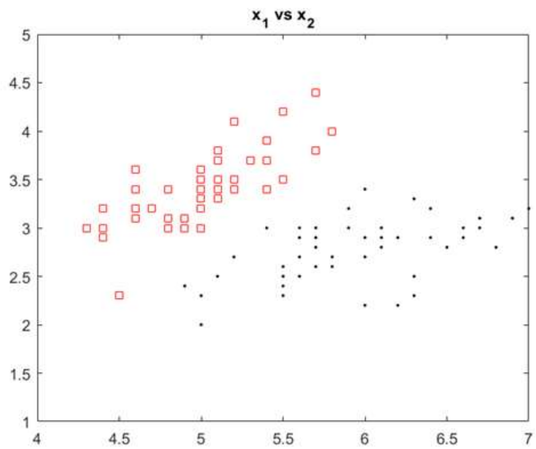

Figure 5: Scatter plot for the distribution between x1 and x2