Bayesian Decision Theory

# Objective

The objective for this laboratory experiment is to understand and implement the Bayes Decision rule for performing classifications.

# Laboratory Exercises Part 1:

For the first portion of this lab, we were tasked with designing a 2-class minimum-error-rate (dichotomize) from the given data to classify the data into two distinct classes; Iris Sentosa or Iris Versicolor according to the classes feature of Sepal width. Using the provided shell program, we are able to finish and fill the code, such that a individual sample can be used as an input along with the training data and the function will return the posterior probabilities and the value of the discriminant function.

With the MATLAB function completed we now run a few values through the code and interpret the output. **Table 1**below showcases the result of running a few variables through the system.

Table : Results of different Width Lengths

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Input Value (x) | Posterior of Sentosa  (w1) | Posterior of Versicolor  (w2) | Single Discriminant Function (g(x2)) | Sentosa or Versicolor |
| 3.3 | 0.7657 | 0.2343 | 0.5314 | Sentosa |
| 4.4 | 1.0000 | 0.0000 | 0.9999 | Sentosa |
| 5.0 | 1.0000 | 0.0000 | 1.0000 | Sentosa |
| 5.7 | 1.0000 | 0.0000 | 1.0000 | Sentosa |
| 6.3 | 1.0000 | 0.0000 | 1.0000 | Sentosa |

From this table we can clearly see that the longer width lengths are all classified under the Sentosa. This result is not surprising since a quick look at the histogram comparing the Iris Sentosa and the Iris Versicolor showcases that the Sentosa has a slightly higher density at longer width lengths which results in the Bayesian Decision Theory to conclude that the larger widths are a part of the Sentosa classification.

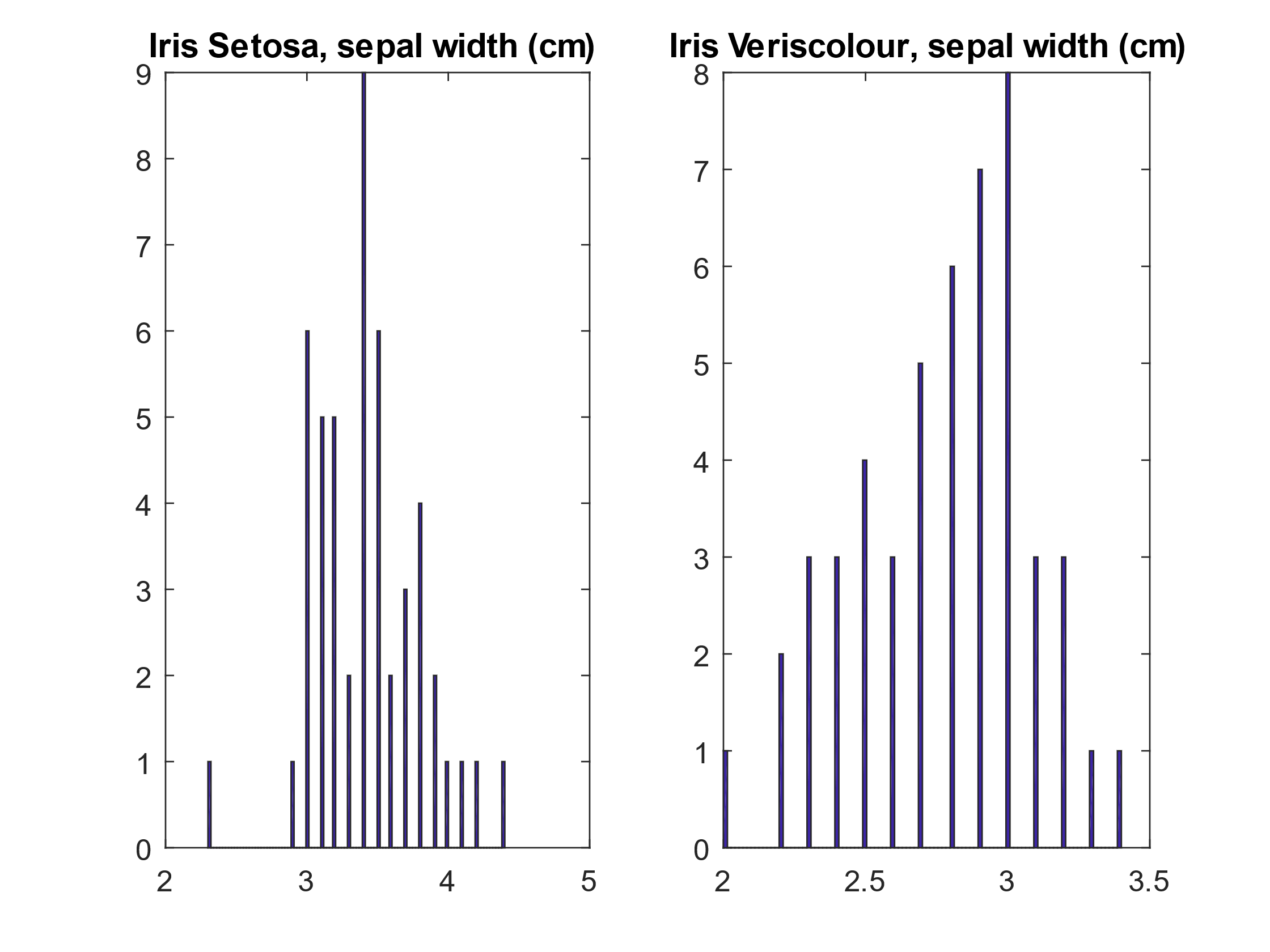


Figure : Histogram of Sepal width

Using the program, we are also able to show the optimal threshold that can separate the two classifications. From the histogram in **Figure 1** we can estimate an appropriate threshold to be around the sepal width of 3.1 cm. We can then test values around 3.1 to verify our result. The results of the test can be shown below in **Table 2**.

Table : Threshold Verification

|  |  |  |  |
| --- | --- | --- | --- |
| Sepal Width | (w1) Sentosa | (w2) Veriscolor | G(x) |
| 2.8 | 0.1817 | 0.8183 | -0.6366 |
| 2.9 | 0.2626 | 0.7374 | -0.4748 |
| 3.0 | 0.3712 | 0.6288 | -0.2577 |
| 3.1 | 0.5026 | 0.4974 | 0.0053 |
| 3.2 | 0.6413 | 0.3587 | 0.2826 |
| 3.3 | 0.7657 | 0.2343 | 0.5314 |

From the Table above, we can see that the posterior probability between the Sentosa and Versicolor becomes almost 50% near the Sepal Width of 3.1cm. This would make an excellent threshold because since the probability of a given Sepal width is effectively 50%, there is an equal chance for it to be classified correctly or incorrectly. This will also result in minimal error between the two classifications. For example, if the sepal length is just 0.1cm shorter than 3.1 cm, it already has a 60% chance of being correct, which is more ideal than a 50% chance. Thus, the best threshold can be 3.1cm.