Ethan Paek

TA: Juliana Shihadeh

COEN 140L - Wednesday Group

22 April 2020

# Lab 4 Report

## Means for Flower Attributes

| **Flower Class** | **Sepal Length (cm)** | **Sepal Width (cm)** | **Petal Length (cm)** | **Petal Width (cm)** |
|---|---|---|---|---|
| Iris-Setosa | 5.0375 | 3.44 | 1.4625 | 0.2325 |
| Iris-Versicolor | 6.01 | 2.78 | 4.3175 | 1.35 |
| Iris-Virginica | 6.6225 | 2.96 | 5.6075 | 1.99 |

## Covariance Matrix for Iris-Setosa

| | | | |
|---|---|---|---|
| 0.13112179 | 0.09897436 | 0.01298077 | 0.01362179 |
| 0.09897436 | 0.13271795 | 0.00205128 | 0.0145641 |
| 0.01298077 | 0.00205128 | 0.02958333 | 0.00458333 |
| 0.01362179 | 0.0145641 | 0.00458333 | 0.00994231 |

## Covariance Matrix for Iris-Versicolor

| | | | |
|---|---|---|---|
| 0.27374359 | 0.08661538 | 0.17212821 | 0.05230769 |
| 0.08661538 | 0.11087179 | 0.08087179 | 0.04538462 |
| 0.17212821 | 0.08087179 | 0.20353205 | 0.07371795 |
| 0.05230769 | 0.04538462 | 0.07371795 | 0.04307692 |

Covariance Matrix for Iris-Virginica

| | | | |
|---|---|---|---|
| 0.46794231 | 0.11041026 | 0.35777564 | 0.05125641 |
| 0.11041026 | 0.11323077 | 0.08107692 | 0.04625641 |
| 0.35777564 | 0.08107692 | 0.34532692 | 0.05930769 |
| 0.05125641 | 0.04625641 | 0.05930769 | 0.07425641 |

Error Rates From Training and Testing Data

| **Data Subset** | **LDA** | **QDA** |
|---|---|---|
| Training | 2.5% | 1.67% |
| Testing | 0% | 0% |

Steps 1-3. Based on the results from last week's lab, it's clear that I have correctly implemented LDA and QDA since the error rates match.

Testing for Linear Separability on each Flower Type

| **Flower Class** | **Error Rate with LDA** |
|---|---|
| Iris-Setosa | 0% |
| Iris-Versicolor | 4% |
| Iris-Virginica | 2% |

Step 4. From these results, it's clear that 'Iris-Setosa' is linearly separable from the other classes since it achieved perfect classification from LDA whereas the other classes did not. To achieve these results, I ran LDA on each of the flower classes with all of their data, rather than separating the attributes into training and testing subsets. The reason why I did this is to reach more accurate results since we are able to test LDA on more data points than before.

Original QDA Results vs QDA with Independent Features

| **Data Subset** | **Error Rate for Original QDA** | **Error Rate for QDA with Independent Features** | **Time to Execute** |
|---|---|---|---|
| Training | 1.67% | 1.67% | 68.2 ms |
| Testing | 0% | 0% | 59.8 ms |

Step 5. From these results, it's clear that by assuming the features are independent, the computation time is much shorter compared to the original QDA function. Although we achieved the same results, it’s important to note that we saved nearly 10 ms. If our number of attributes (K) was much larger, we would expect a much greater difference in execution time. This is because, with original QDA, we have $O(K^2)$ whereas assuming independent features gives us O(K).