Experiment with Gaussian Discriminators

Implemented Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA):

1. Created two functions in Python: ldaLearn and qdaLearn which take a training data set (a feature matrix and labels) and return the means and covariance matrix.

2. Implemented two functions ldaTest and qdaTest which return the true labels for a given test data set and the accuracy using the true labels for the test data.

3. Trained both methods using the sample training data.

4. Reported the accuracy of LDA and QDA on the provided test data set and plotted the discriminating boundary for linear and quadratic discriminators.

Report

After training the model on the sample training set, we got that the accuracy for Linear Discriminant Analysis is 0.97, while using Quadratic Discriminant Analysis the accuracy is 0.96

Below we have plotted the boundary for both models:
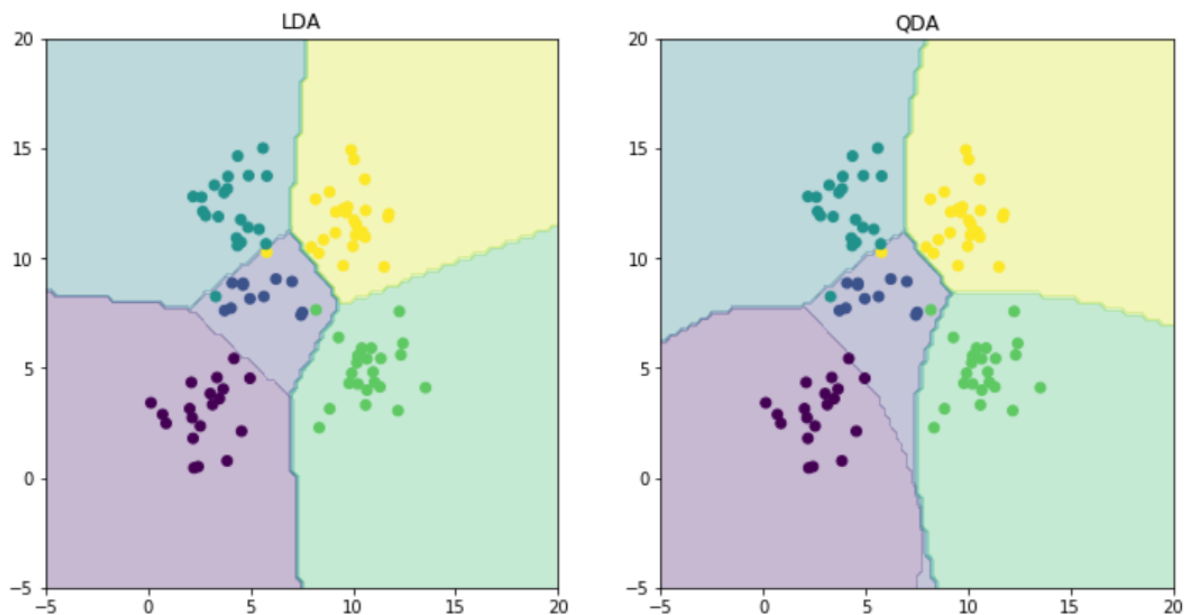
As we can see from the plot, LDA can only learn linear boundaries, while QDA can learn quadratic boundaries and is therefore more flexible and can be more accurate, especially for large datasets.

Generally QDA is tend to be more accurate but it requires enough amount of data to estimate μ and $\sum$ and thus tends to fit the data better than LDA, but then it has more parameters to estimate. Because, with QDA, we have a separate covariance matrix for every class, thus it might be computationally inefficient to use the model for datasets with large number of classes. Thus the main difference between LDA and QDA is that for LDA we use the entire training data and compute MLE for covariance matrix $\sum$.

To summarize, the accuracy for our training model using Linear Discriminant Analysis is better than using Discriminant Linear Analysis, 0.97 comparing to 0.96 accordingly. This is caused hence the dataset has a small number of inputs and we do not compute quadratic term, thus for small datasets this gives a better variance estimation, then QDA. Also if the dataset has the assumption of Gaussian distribution - LDA is a perfect model for calculating MSE and provide better results with less computations.