Lab 4

Angeline Chen

The purpose of the lab is to implement LDA and QDA. When the model is trained, the training data is used to calculate the mean and the covariance for each class. When the model is tested, the normal distribution probability function is used to calculate the probability that the data point is in each class, and the data point is classified to the class with the highest probability.

LDA Train Error Rate	QDA Train Error Rate	LDA Test Error Rate	QDA Test Error Rate
2.5%	1.67%	0%	0%

The train error rate is greater than the test error rate. Because the training data is more than the testing data, the model fits the testing data better than the training data. The LDA error rate is greater than the QDA error rate. Because LDA learns linear boundaries, and QDA learns quadratic boundaries, the data does not have a fixed covariance, and the decision boundary is not linear.

Classes	Train Error Rate
setosa and versicolor	0%
setosa and virginica	0%
versicolor and virginica	3.75%

If the train error rate is equal to 0, the values are linearly separable. Setosa and versicolor are linearly separable. Setosa and virginica are linearly separable. Versicolor and virginica are not linearly separable.

Diagonal Matrix

LDA Train Error Rate	QDA Train Error Rate	LDA Test Error Rate	QDA Test Error Rate
5%	4.17%	0%	0%

When the diagonal matrix is used, the train error rates are greater. Because the diagonal matrix has fewer values, the result is less accurate.

Training Times

	LDA	QDA

Non-Diagonal Matrix	0.0011	0.00093
Diagonal Matrix	0.00056	0.00056

The training time of using a diagonal matrix is faster than the training time of using a non-diagonal matrix. The time complexity of calculating the non-diagonal matrix is quadratic. The time complexity of calculating the diagonal matrix is linear.