**Linear vs. Quadratic Discriminant Analysis**

In this blog post we will be looking at the differences between Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA). Both statistical learning methods are used for classifying observations to a class or category. So that means that our response variable is categorical.

What we will be covering:

* The Bayes Classifier
* A Quick Example of the Bayes Classifier
* Assumptions for Linear Discriminant Analysis
* Assumptions for Quadratic Discriminant Analysis
* LDA vs. QDA. What Algorithm Performs Better and in What Situations?

Let's get started!

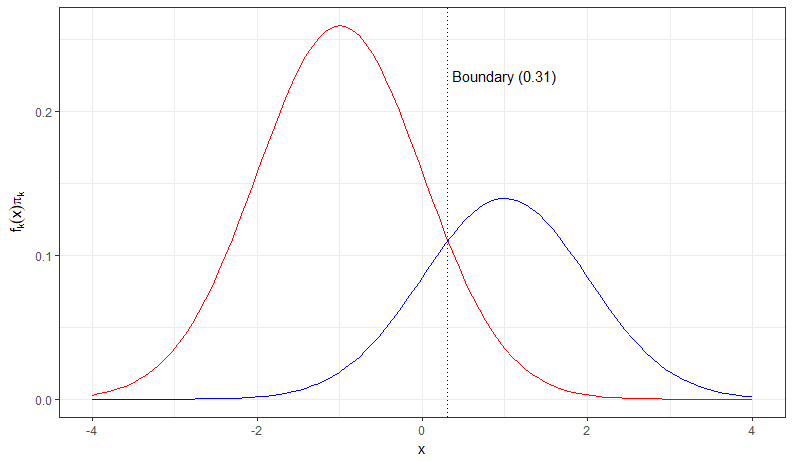


**The Bayes Classifier**

In theory, we would always like to predict a qualitative response with the Bayes classifier because this classifier gives us the lowest test error rate out of all classifiers. So why don’t we do that? Unfortunately for using the Bayes classifier, we need to know the true conditional population distribution of Y given X and the we have to know the true population parameters $latex \sigma$ and $latex \mu$.

**An Example of the Bayes Classifier**

In the plot below, we show two normal density functions which are representing two distinct classes. The variance parameters are $latex \sigma\_1 = \sigma\_2$ = 1 and the mean parameters are $latex \mu\_1$ = -1 and $latex \mu\_2$ = 1. There is some uncertainty to which class an observation belongs where the densities overlap. This means that the Bayes error rate will be greater than zero. We have specified $latex \Pi$ = 0.35and $latex \Pi$ = 0.65. By inspection of the decision boundary at 0.31 we see that the Bayes Classifier assigns the observation to class 1 if x < 0.31 and to class 2 otherwise.



In this case, the decision boundary is the Bayes decision boundary and we can compute the Bayes classifier because we know that X is drawn from a normal distribution within each class. In addition to that we know all our parameters ($latex \sigma\_1 = \sigma\_2$ = 1, $latex \mu\_1$ = -1 and $latex \mu\_2$ = 1, and $latex \Pi\_1$ = 0.35 and $latex \Pi\_2$ = 0.65).

In practice, we unfortunately cannot apply the Bayes classifier because we do not know if X is drawn from a normal distribution and we do not know our true population parameters. Therefore, we are tying to approximate the Bayes classifier with LDA and QDA. These two methods use the estimated $latex \hat{\mu}\_k$ (average of all training observations from the kth class), $latex \hat{\sigma}^2$ (weighted average of sample variances for each of the K classes), and $latex \hat{\Pi}\_k$ in the discriminant functions in order to assign an observation to a class.

**Linear Discriminant Analysis**

Assumptions:

* LDA assumes normal distributed data and a class-specific mean vector.
* LDA assumes a common covariance matrix. So, a covariance matrix that is common to all classes in a data set.

When these assumptions hold, then LDA approximates the Bayes classifier very closely and the discriminant function produces a linear decision boundary. However, LDA also achieves good performances when these assumptions do not hold and a common covariance matrix among groups and normality are often violated.

**Quadratic Discriminant Analysis**

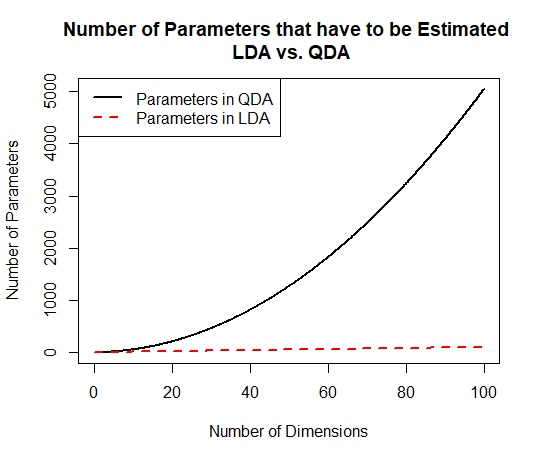
Assumptions:

* Observation of each class are drawn from a normal distribution (same as LDA).
* QDA assumes that each class has its own covariance matrix (different from LDA).

When these assumptions hold, QDA approximates the Bayes classifier very closely and the discriminant function produces a quadratic decision boundary.

**LDA versus QDA**

When the number of predictors is large the number of parameters we have to estimate with QDA becomes very large because we have to estimate a separate covariance matrix for each class. In the plot below, we show how many parameters we have to estimate for LDA versus QDA.



The higher the dimension of the data set (the more predictors in a data set) the more parameters we have to estimate. This can lead to a high variance and so we have to be careful when using QDA.

In conclusion, LDA is less flexible than QDA because we have to estimate less parameters. This can be good when we have only a few observations in our training data set so we lower the variance. On the other hand, when the K classes have very different covariance matrices then LDA suffers from high bias and QDA might be a better choice. So, what is comes down to is the bias-variance trade-off. Therefore, it is crucial to test the underlying assumptions of LDA and QDA on the data set and then use both methods to decide which one is more appropriate.