1. Q
2. Note that .

Thus, we may observe that maximizing also maximizes , and maximizing one is equivalent to maximizing the other. Consequently, assigning the observation for which is largest is equivalent to maximizing .

1. Assume we have classes, and that if an observation belongs to the th class, then comes from a one-dimensional normal distribution, . We do not assume that the variance is equivalent across all the classes, or in other words, we may not assume that for all .Then

From this, we obtain the discriminant function , which is quadratic in and by extension not linear in .

1. Q
   1. , , with

On average, 10% of the available observations will be used.

* 1. , rest same as

On average, of the available observations will be used.

On average, of the available observations will be used.

* 1. When is very large, there exists an exponentially decaying number of observations that are considered “near” enough to any given observation in -dimensional space. The possible space for decreases relative to the constant total number of observations, making the accuracy of suffer as increases.
  2. For a -dimensional hypercube to take up of the training observations, it is necessary for the -dimensional “area” (length⇒ 1-D, area⇒ 2-D, volume⇒ 3-D, etc) to equal . As we can observe with and ,, this holds true when the hypercube’s side length multiplies with itself times in order to equal precisely , or when . Therefore, a -dimensional hypercube containing of the training observations has a side length of .

1. LDA/QDA
   1. If the Bayes decision boundary is linear, we would expect LDA to perform better on the test set, but QDA to perform better on the training set due to its flexibility.
   2. If the Bayes Decision boundary is non-linear, then QDA would outperform LDA in both cases due to its bias reduction from assuming a non-linear shape, which is true in both cases.
   3. In general, as the number of available observations *n* increases, QDA performs better relative to LDA as a result of the variance of the classifier being less and less of a concern with more data points. LDA has low variance already, but QDA’s non-trivial increase in variance is countered by having more data points.
   4. False. Even if the Bayes decision boundary for a given problem is linear, a QDA will not be able to outperform an LDA. QDA performs better in cases where the increase in variance can be offset successfully by other factors, such as a sufficiently large number of data points or a sufficient drop in bias. If the true decision boundary is not fit well by the QDA, there will be little to no reduction in bias to accommodate for the increased variance, meaning that, despite its flexibility, QDA will not perform better than LDA when the Bayes decision boundary is linear. As well, a small number of data points will lead to overfitting in the QDA, which would be problematic for its accuracy and allow it to be surpassed by LDA.

If a student studies for 40 hours and has an undergrad GPA of 3.5, they have a 37.75% chance of obtaining an A in the class.

The same student from a) would need to study for 50 hours in order to have a 50% chance of obtaining an A.

1. Let and . We represent as .

Therefore, there is a chance that a company with a profit will issue a dividend this year.

1. In general, in the given situation it would be better to use logistic regression for the classification of new observations. When , KNN has a training error rate of 0% as it perfectly follows any perceivable pattern within the training data, no matter if it represents the true relationship or is simply noise. Therefore, the average error rate between training and test sets being 18%, the test error rate can be calculated as 2\*average error-training error=2\*average error=36%. This is significantly higher than the test error rate of logistic regression, by a whole 6%, so even if the average error rate of logistic regression is higher, it is still preferable to and more accurate than KNN for classification of new observations.
2. Odds

27% of people with odds of 0.37 will default.

With 16% chance of defaulting, the odds of this individual are equivalent to 0.19.