Lab 3: Classification

0.1 Stock Market Data

1. Load and explore (through numerical and graphical summaries) the `Smarket` data (this is in `ISLR` package).

This data contains percentage returns for the S&P 500 stock index over 1,250 days (2001 - 2005). For each date, we have the percentage returns of the five previous days (`Lag1` - `Lag5`), the number of shares traded on the previous day in billions (`Volume`), percentage return on the date in question (`Today`) and `Direction` (`Up` or `Down` on this date).

0.2 Logistic Regression

1. Fit a logistic regression model to predict `Direction` using `Lag1` through `Lag5` and `Volume`. Describe your results.

2. Create a confusion matrix for the training data.

3. What is the overall error rate of the model?

4. Create two data sets, `train` and `test` that correspond to the observations from 2001 to 2004 (`train`) and 2005 (`test`).

5. Repeat 1-3, but obtain the test confusion matrix and error rate.

6. Repeat 5, but with a model of `Direction` on `Lag1` and `Lag2` only.

0.3 LDA

1. Fit a linear discriminant analysis model to the `train` data set you created in the previous section with `Direction` as the response and `Lag1` and `Lag2` as the predictors.
2. What are the values for $\hat{\pi}_1$ and $\hat{\pi}_2$?
3. Create a confusion matrix for the `test` data.
4. What is the test error rate?

0.4 QDA

1. Fit a quadratic discriminant analysis model to the `train` data set you created in the previous section with `Direction` as the response and `Lag1` and `Lag2` as the predictors.
2. Create a confusion matrix for the `test` data.
3. What is the test error rate?

0.5 KNN

1. Fit a KNN model with $K = 1$ to the `train` data set you created in the previous section with `Direction` as the response and `Lag1` and `Lag2` as the predictors.
2. Create a confusion matrix for the `test` data.
3. What is the test error rate?
4. Repeat 1.-3. with $K = 3$ and $K = 5$.

Of all the models you fit today, which would you pick to predict values of `Direction` and why?