

Exercise 14: Fit a Logistic Regression Model to Previous Dataset

Daniel Angel

2/7/2021

Logistic Regression

First I must load my data in then split it into two subsets of data according to a 4 to 1 ratio for training to testing, then I use that trained model to make predictions which I will test for accuracy against the original training data.

```
##               Predicted_Value
## Actual_Value FALSE TRUE
##           0    282   229
##           1    186   302
```

```
## [1] 0.5845846
```

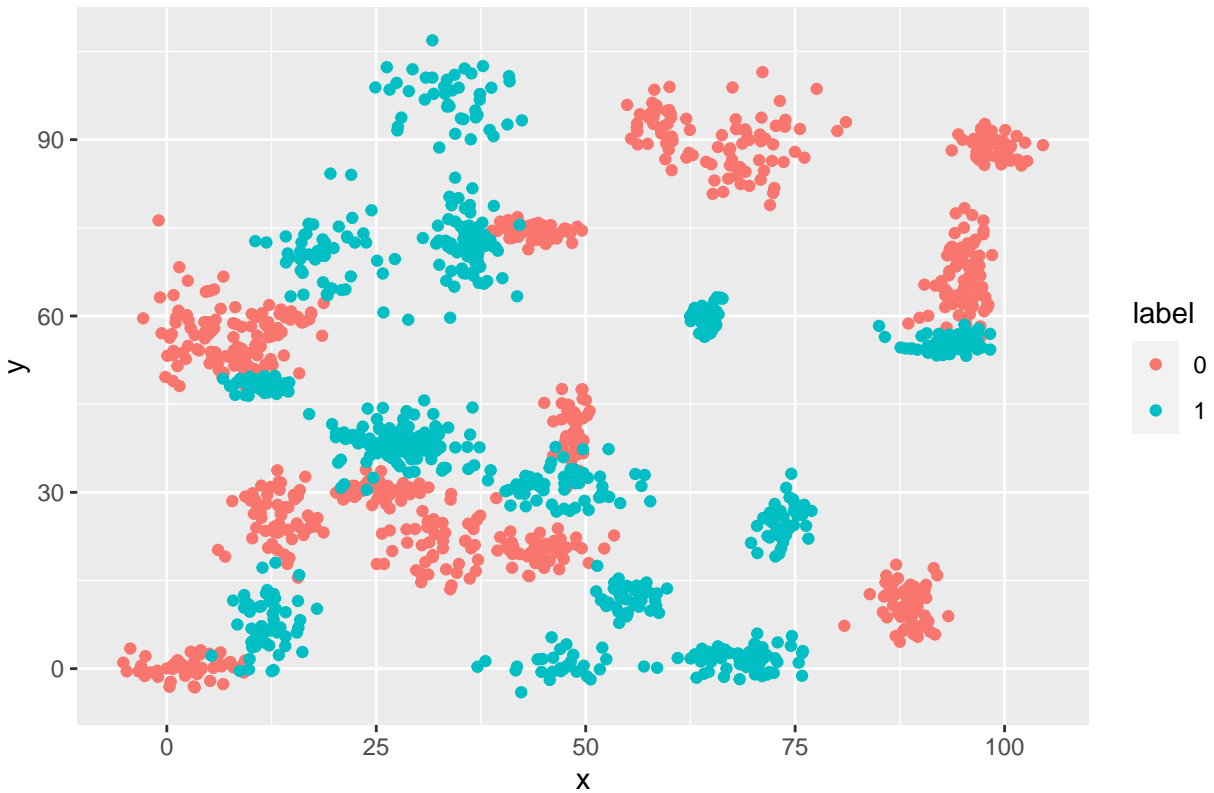
A. What is the accuracy of the logistic regression classifier?

The accuracy is 58.4584585 percent.

Nearest Neighbor

B. How does the accuracy of the logistic regression classifier compare to the nearest neighbors algorithm?

K Nearest Neighbors



Above we can see that there is natural clustering in this data set.

I will try $k=10$ for my number of clusters and also $k = 38$ as the square root of the number of observations.

```
##  
## knn10  0  1  
##      0 248  8  
##      1  8 235
```

```
## [1] 0.9679359
```

```
##  
## knn38  0  1  
##      0 244 10  
##      1 12 233
```

```
## [1] 0.9559118
```

Despite less clusters the KNN Model for $k = 10$ had the highest accuracy at a whopping 96.7935872 percent. This is significantly better than the approximately 58% achieved by the logistic regression.

Comparing the Models

C. Why is the accuracy of the logistic regression classifier different from that of the nearest neighbors?

Logistic Regression



Comparing this to the plot above you can see that the logistic regression classifier is not well suited to classifying which is so scattered and further it is obvious to see why it would be beaten by a clustering algorithm when the data is already somewhat naturally clustered.