ASSIGNMENT 8.2

Blaine Blasdell

2020-07-26
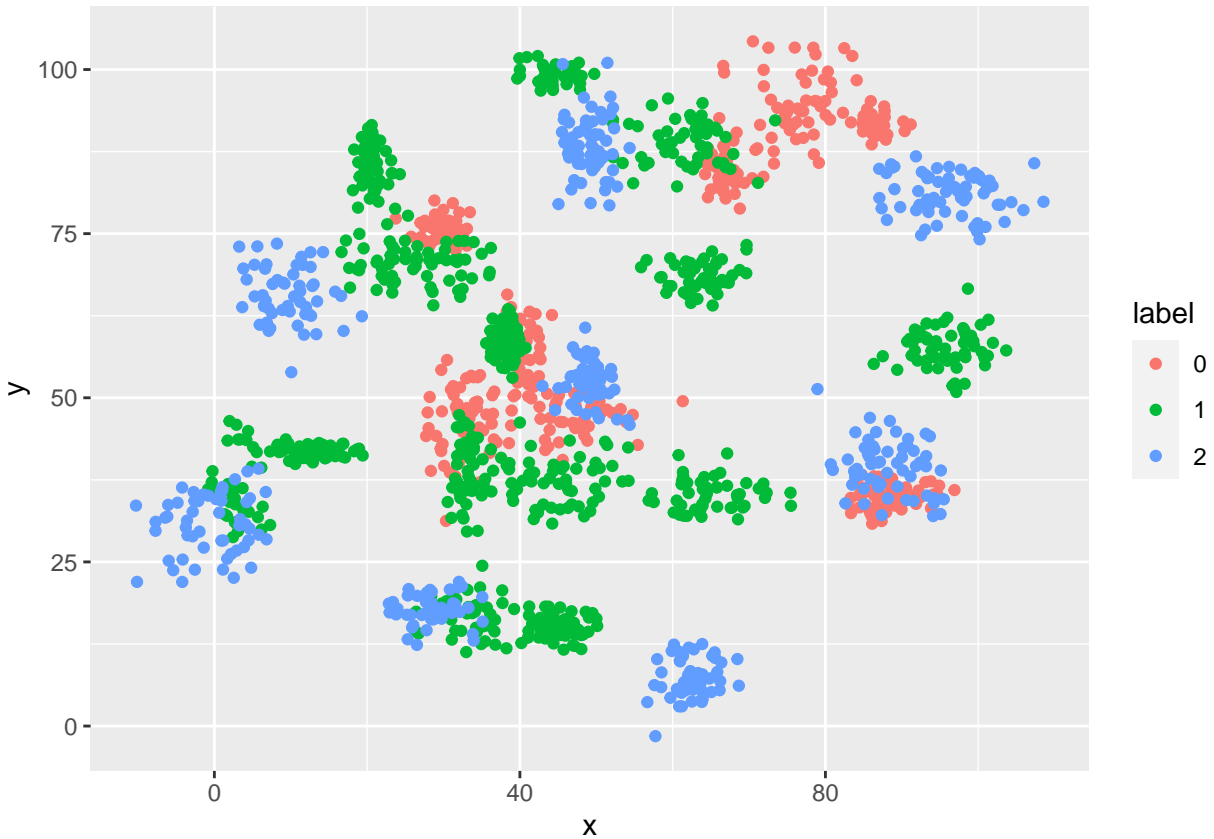

Assignment 8.2

Question 8.2a

a. Plot the data from each dataset using a scatter plot.

Question 8.2b

b. Fit a k nearest neighbors model for each dataset for k=3, k=5, k=10, k=15, k=20, and k=25. Compute the accuracy of the resulting models for each value of k. Plot the results in a graph where the x-axis is the different values of k and the y-axis is the accuracy of the model.

```
## 
## Call:
## glm(formula = label ~ x + y, family = binomial(), data = bin_train)
## 
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3728  -1.1697  -0.9575   1.1646   1.3989
## 
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.424809   0.117224   3.624  0.00029 ***
## x           -0.002571   0.001823  -1.411  0.15836
## y           -0.007956   0.001869  -4.257 2.07e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 2075.8  on 1497  degrees of freedom
```

```
## Residual deviance: 2052.1  on 1495  degrees of freedom
## AIC: 2058.1
##
## Number of Fisher Scoring iterations: 4


##    x        y
## 1  3 97.17172
## 2  5 97.17172
## 3 10 96.96970
## 4 15 97.17172
## 5 20 96.56566
## 6 25 97.17172
```
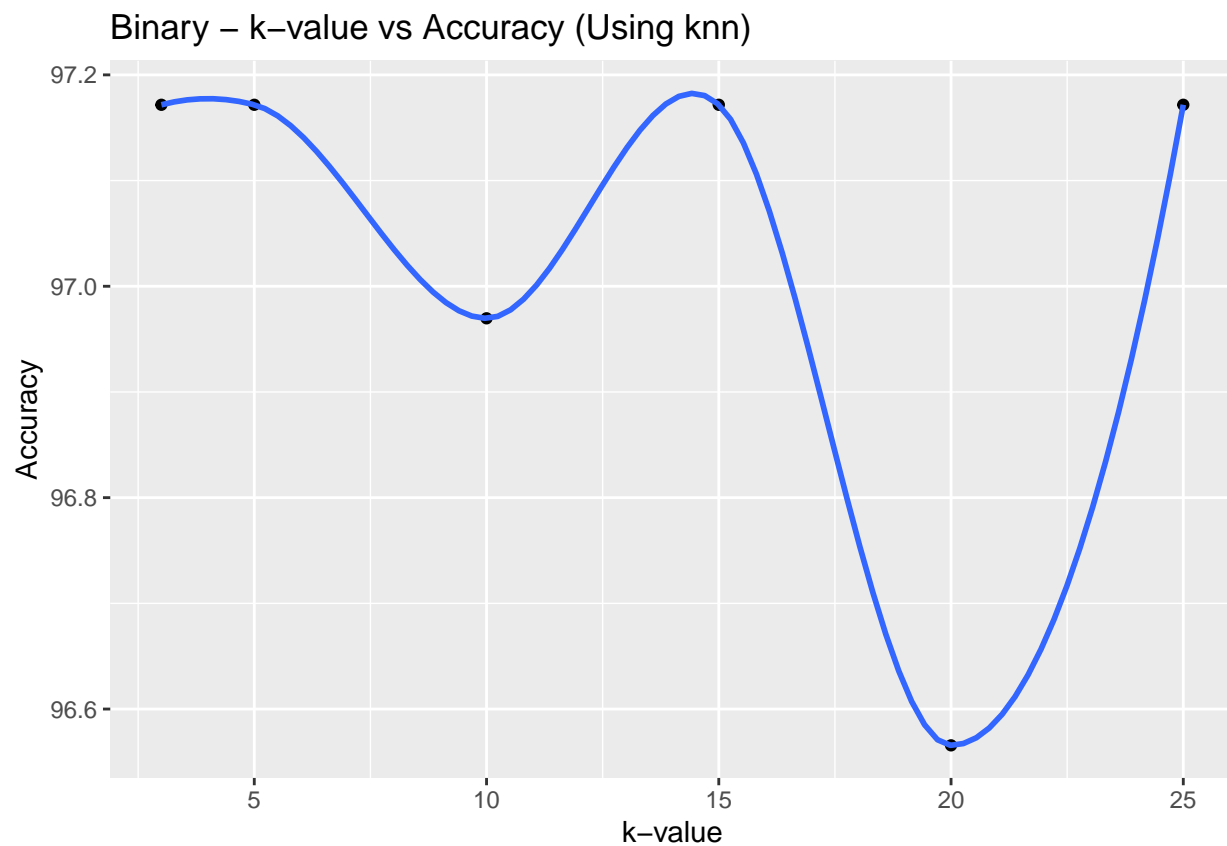


```
##
## Call:
## glm(formula = label ~ x + y, family = binomial(), data = tri_train)
##
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -1.9906  -1.1552   0.5969   0.7718   1.2460
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.771616   0.189883  14.596  < 2e-16 ***
## x           -0.014758   0.002285  -6.460 1.05e-10 ***
```

```
## y              -0.015715   0.002428  -6.473 9.60e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1767.9  on 1567  degrees of freedom
## Residual deviance: 1663.4  on 1565  degrees of freedom
## AIC: 1669.4
##
## Number of Fisher Scoring iterations: 4


##     x        y
## 1   3 63.89961
## 2   5 65.05792
## 3  10 65.44402
## 4  15 65.25097
## 5  20 63.70656
## 6  25 63.32046
```


Trinary – k–value vs Accuracy (knn)
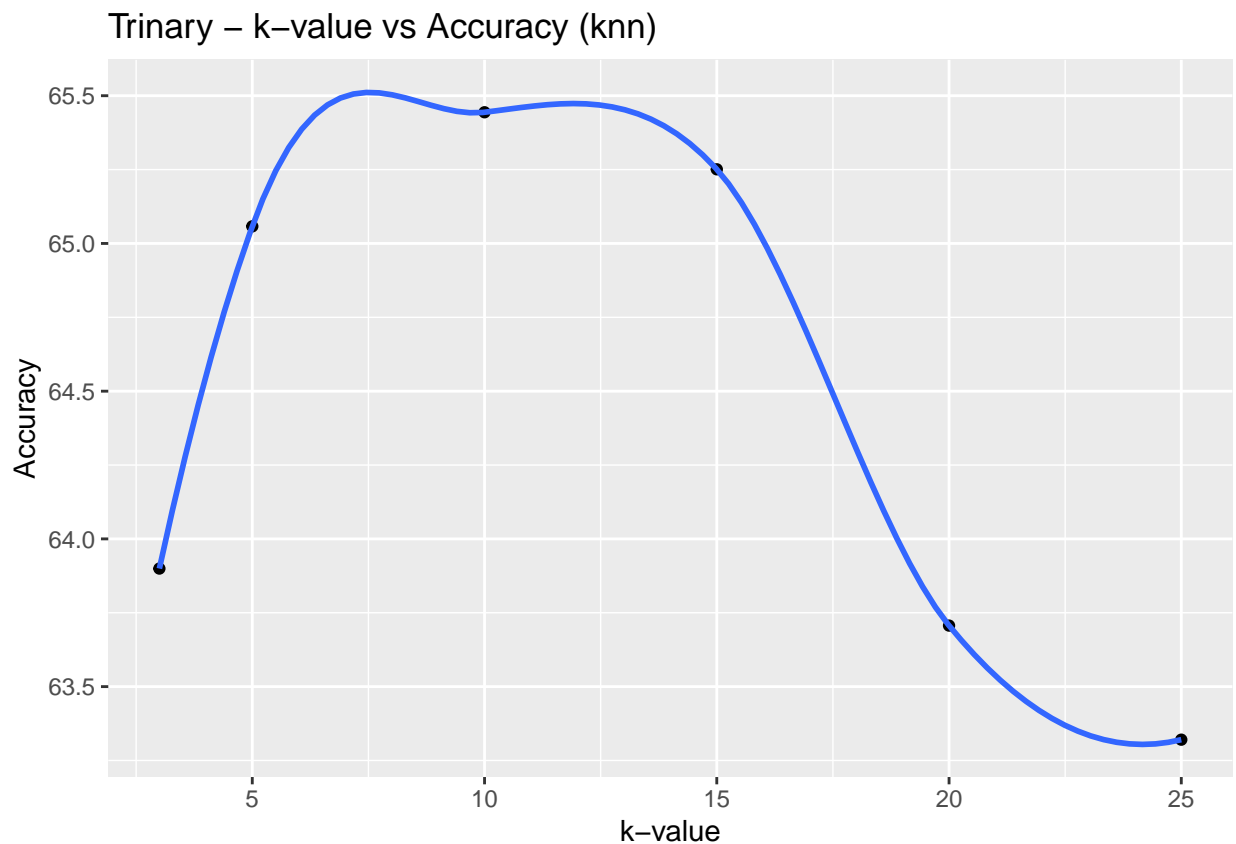
Question 8.2c

c. In later lessons, you will learn about linear classifiers. These algorithms work by defining a decision
boundary that separates the different categories. Looking back at the plots of the data, do you think
a linear classifier would work well on these datasets?

Answer - Based on what I see (visual observation) a linear classifier would not work a the data points do not look to easily fit a straght line. Clustering seems like a better model, just looking at the visual.

References