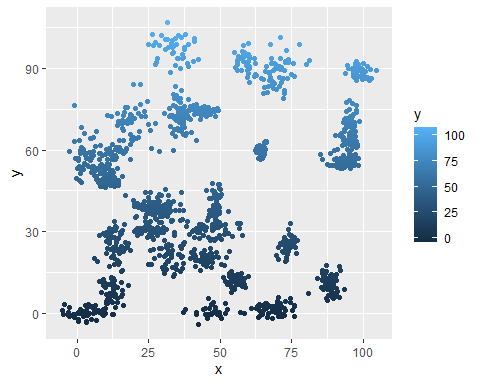
ASSIGNMENT 8 - Exercise 14: Fit a Logistic Regression Model to Previous Dataset

Ragunath Gunasekaran

2020-10-28

## Data Analysis Analysis

## Warning: package 'ggplot2' was built under R version 4.0.2



## a.What is the accuracy of the logistic regression classifier?

##   
## Call:  
## glm(formula = label ~ x + y, family = "binomial", data = binaryclass\_ds)  
##   
## 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

##   
## Call:  
## glm(formula = label ~ x, family = "binomial", data = binaryclass\_ds)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.246 -1.159 -1.064 1.184 1.293   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.137369 0.095119 1.444 0.1487   
## x -0.004119 0.001775 -2.321 0.0203 \*  
## ---  
## 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: 2070.4 on 1496 degrees of freedom  
## AIC: 2074.4  
##   
## Number of Fisher Scoring iterations: 3

## Warning: package 'ROCR' was built under R version 4.0.2

## [1] 0.5732267

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

Source - Ref 5 - In theory, nearest neighbors algorithm KNN is better than linear regression. The logistic regression classifier accuracy is 57% and KNN Accuracy is 74%.

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

Based on the Results, KNN is deterministic algorithm and logistic regression is a stochastic algorithm.

KNN Alogrithm need to verify neighbors, Hence it’s a non-parametric but Logitic Regression is requires parameter. KNN is little complex when compare to Logistic Regression, Hence the performance of operation may slow in KNN.

# References

1. Generalized Linear Models, Quick R by Datacamp - <https://www.statmethods.net/advstats/glm.html>
2. <http://ipa-tys.github.io/ROCR/>
3. <https://www.r-bloggers.com/2016/11/calculating-auc-the-area-under-a-roc-curve/>
4. <https://towardsdatascience.com/k-nearest-neighbors-algorithm-with-examples-in-r-simply-explained-knn-1f2c88da405c>
5. <https://towardsdatascience.com/comparative-study-on-classic-machine-learning-algorithms-24f9ff6ab222#>:~:text=KNN%20is%20a%20non%2Dparametric%20model%2C%20where%20LR%20is%20a,can%20only%20output%20the%20labels.