Exercise 14

Bilal Kudaimi

October 23, 2020

```
## Warning: package 'tidyverse' was built under R version 4.0.3

## -- Attaching packages ---------------------------------------------------------------------------

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## Warning: package 'tidyr' was built under R version 4.0.3

## Warning: package 'dplyr' was built under R version 4.0.3

## -- Conflicts ------------------------------------------------------------------------------------ ti
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

##
## Call:
## glm(formula = label ~ x + y, data = binary)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.6108  -0.4956  -0.3664   0.4925   0.6253
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.6049208  0.0287423  21.046  < 2e-16 ***
## x           -0.0006309  0.0004481  -1.408    0.159
## y           -0.0019662  0.0004578  -4.295 1.86e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2464139)
##
##     Null deviance: 374.28  on 1497  degrees of freedom
## Residual deviance: 368.39  on 1495  degrees of freedom
## AIC: 2157.8
##
## Number of Fisher Scoring iterations: 2
```

```
## [1] 61.28171
```

#A

The accuracy of the model was found to be 61.28%, meaning this model can accurately predict the label of a data point 61.28% of the time.

#B

The accuracy of KNN tends to be greater than the accuracy of the logistic regression classifier.

#C

The accuracies are different due to the way each method is run. Logistic regression is used for predicting a binary value from data, while KNN used for classifying data based on a data point's spatial neighbors.

Logistic regression is used for linear solutions while KNN can be used for nonlinear solutions. KNN is relatively slower than logistic regression. Logistic regression is a parametric model and can derive probabilities and confidence levels, while KNN is non-parametric, and can only output labels.