Baruch College, STA-CIS 3920, Exercise#6 Anil Poonai 37

31 July 2020

**Exercise**

6.1

Below is the code instead of in the appendix since there is only one part in this assignment and it shows that the Naïve Bayes is less than the KNN accuracy. I first did the cross validation with the lagged 1 logged ranges and the risk value associated with them. Then when I tried it out in the KlaR Naive Bayes method it came with a 52.81% accuracy. I also did the same test with the e1071 package and after cross validation and came up with a 51.26% accuracy while KNN had a 55% accuracy. I also want to make the observation since I have worked with both of these methods before this class that this is usually the case when the data usually has low parameters and has a bunch of zeros in the sets. Which is actually a common problem in classification as a whole due to data points gaining the same value as a point previously reviewed. With that said KNN also does not handle data that well as any missing data will never work unlike in Naive Bayes. Appendix shows the results from me doing the code in the lecture notes with the iris dataset.

R version 4.0.2 (2020-06-22) -- "Taking Off Again"

Copyright (C) 2020 The R Foundation for Statistical Computing

Platform: x86\_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.

You are welcome to redistribute it under certain conditions.

Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.

Type 'contributors()' for more information and

'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or

'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

[Previously saved workspace restored]

> library(class)

> library(e1071)

> library(tidyverse)

-- Attaching packages -------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------- tidyverse 1.3.0 --

v ggplot2 3.3.0 v purrr 0.3.4

v tibble 3.0.1 v dplyr 0.8.5

v tidyr 1.0.2 v stringr 1.4.0

v readr 1.3.1 v forcats 0.5.0

-- Conflicts ----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------- tidyverse\_conflicts() --

x dplyr::filter() masks stats::filter()

x dplyr::lag() masks stats::lag()

> library(MASS)

Attaching package: ‘MASS’

The following object is masked from ‘package:dplyr’:

select

> library(caret)

Loading required package: lattice

Attaching package: ‘caret’

The following object is masked from ‘package:purrr’:

lift

> library(klaR)

> AOS <- read\_csv("C:/Users/poona/Desktop/School/AOS.csv")

Parsed with column specification:

cols(

.default = col\_double(),

Date = col\_character(),

Direction = col\_character(),

HiLoRisk = col\_character()

)

See spec(...) for full column specifications.

> str(AOS)

tibble [3,649 x 20] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)

$ Date : chr [1:3649] "1/11/2006" "1/12/2006" "1/13/2006" "1/17/2006" ...

$ Open : num [1:3649] 6.68 6.6 6.47 6.48 6.53 ...

$ High : num [1:3649] 6.71 6.67 6.51 6.58 6.74 6.78 7.4 7.12 7.22 7.17 ...

$ Low : num [1:3649] 6.57 6.44 6.45 6.46 6.51 6.62 6.79 6.82 7.03 7.02 ...

$ Close : num [1:3649] 6.6 6.46 6.47 6.53 6.74 6.63 7.32 7.01 7.17 7.08 ...

$ Adj Close : num [1:3649] 2.82 2.75 2.76 2.79 2.87 2.83 3.12 2.99 3.06 3.02 ...

$ Volume : num [1:3649] 1474800 1913200 1042800 1850400 2258800 ...

$ LGFet : num [1:3649] -1.49 -2.17 0.21 0.98 3.16 ...

$ Lag1 : num [1:3649] 2.39 -1.49 -2.17 0.21 0.98 ...

$ Lag2 : num [1:3649] 3.4 2.39 -1.49 -2.17 0.21 ...

$ Lag3 : num [1:3649] 3.43 3.4 2.39 -1.49 -2.17 ...

$ Lag4 : num [1:3649] -0.94 3.43 3.4 2.39 -1.49 -2.17 0.21 0.98 3.16 -1.58 ...

$ Lag5 : num [1:3649] 0.54 -0.94 3.43 3.4 2.39 -1.49 -2.17 0.21 0.98 3.16 ...

$ Daily Range: num [1:3649] 0.14 0.23 0.06 0.12 0.23 0.16 0.61 0.3 0.2 0.15 ...

$ RangeL1 : num [1:3649] 1 3.89 0.68 2.38 0.77 ...

$ RangeL2 : num [1:3649] 0.03 1 3.89 0.68 2.38 ...

$ LogR1 : num [1:3649] 0 0.59 -0.17 0.38 -0.11 0.34 0.68 1.08 1.16 0.81 ...

$ LogR2 : num [1:3649] -1.46 0 0.59 -0.17 0.38 -0.11 0.34 0.68 1.08 1.16 ...

$ Direction : chr [1:3649] "DOWN" "DOWN" "UP" "UP" ...

$ HiLoRisk : chr [1:3649] "LoRisk" "LoRisk" "LoRisk" "LoRisk" ...

- attr(\*, "spec")=

.. cols(

.. Date = col\_character(),

.. Open = col\_double(),

.. High = col\_double(),

.. Low = col\_double(),

.. Close = col\_double(),

.. `Adj Close` = col\_double(),

.. Volume = col\_double(),

.. LGFet = col\_double(),

.. Lag1 = col\_double(),

.. Lag2 = col\_double(),

.. Lag3 = col\_double(),

.. Lag4 = col\_double(),

.. Lag5 = col\_double(),

.. `Daily Range` = col\_double(),

.. RangeL1 = col\_double(),

.. RangeL2 = col\_double(),

.. LogR1 = col\_double(),

.. LogR2 = col\_double(),

.. Direction = col\_character(),

.. HiLoRisk = col\_character()

.. )

> train=sample(3649,2100)

> x.train=AOS[train,17]

> y.train=AOS[train,20]

> x.test=AOS[-train,17]

> y.test=AOS[-train,20]

> y.train=factor(unlist((y.train)))

> y.test=factor(unlist((y.test)))

> model = train(x.train,y.train,'nb',trControl=trainControl(method='cv',number=10))

There were 25 warnings (use warnings() to see them)

> warnings()

Warning messages:

1: Setting row names on a tibble is deprecated.

2: Setting row names on a tibble is deprecated.

3: Setting row names on a tibble is deprecated.

4: Setting row names on a tibble is deprecated.

5: In FUN(X[[i]], ...) :

Numerical 0 probability for all classes with observation 151

6: Setting row names on a tibble is deprecated.

7: Setting row names on a tibble is deprecated.

8: In FUN(X[[i]], ...) :

Numerical 0 probability for all classes with observation 201

9: Setting row names on a tibble is deprecated.

10: Setting row names on a tibble is deprecated.

11: In FUN(X[[i]], ...) :

Numerical 0 probability for all classes with observation 152

12: Setting row names on a tibble is deprecated.

13: Setting row names on a tibble is deprecated.

14: Setting row names on a tibble is deprecated.

15: Setting row names on a tibble is deprecated.

16: In FUN(X[[i]], ...) :

Numerical 0 probability for all classes with observation 133

17: Setting row names on a tibble is deprecated.

18: Setting row names on a tibble is deprecated.

19: Setting row names on a tibble is deprecated.

20: Setting row names on a tibble is deprecated.

21: Setting row names on a tibble is deprecated.

22: Setting row names on a tibble is deprecated.

23: Setting row names on a tibble is deprecated.

24: Setting row names on a tibble is deprecated.

25: Setting row names on a tibble is deprecated.

> table(predict(model$finalModel, x.test)$class, y.test)

y.test

HiRisk LoRisk

HiRisk 472 409

LoRisk 322 346

> str(x.train)

tibble [2,100 x 1] (S3: tbl\_df/tbl/data.frame)

$ LogR1: num [1:2100] 0.15 0.68 0.27 0.18 0.02 0.06 -0.34 0.07 0.63 0.17 ...

> str(y.train)

Factor w/ 2 levels "HiRisk","LoRisk": 1 2 1 2 2 1 2 2 1 2 ...

- attr(\*, "names")= chr [1:2100] "HiLoRisk1" "HiLoRisk2" "HiLoRisk3" "HiLoRisk4" ...

> str(x.test)

tibble [1,549 x 1] (S3: tbl\_df/tbl/data.frame)

$ LogR1: num [1:1549] 0 0.59 -0.17 0.38 0.34 0.68 1.08 1.16 0.33 0.05 ...

> str(y.test)

Factor w/ 2 levels "HiRisk","LoRisk": 2 2 2 2 2 1 2 2 2 2 ...

- attr(\*, "names")= chr [1:1549] "HiLoRisk1" "HiLoRisk2" "HiLoRisk3" "HiLoRisk4" ...

> classifier <- naiveBayes(x.train, y.train)

> table(predict(classifier, x.test), y.test)

y.test

HiRisk LoRisk

HiRisk 232 193

LoRisk 562 562

**Appendix**

R version 4.0.2 (2020-06-22) -- "Taking Off Again"

Copyright (C) 2020 The R Foundation for Statistical Computing

Platform: x86\_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.

You are welcome to redistribute it under certain conditions.

Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.

Type 'contributors()' for more information and

'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or

'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

Warning: namespace ‘caret’ is not available and has been replaced

by .GlobalEnv when processing object ‘model’

[Previously saved workspace restored]

> pairs(iris[,1:4], main = "Iris Data (red=setosa,green=versicolor,blue=virginica)",

+ pch = 21, bg = c("red", "green3", "blue")[unclass(iris$Species)])

A screenshot of a cell phone

Description automatically generated

> install.packages('e1071', dependencies = TRUE)\

Error: unexpected input in "install.packages('e1071', dependencies = TRUE)\"

> install.packages('e1071', dependencies = TRUE)

Installing package into ‘C:/Users/poona/Documents/R/win-library/4.0’

(as ‘lib’ is unspecified)

--- Please select a CRAN mirror for use in this session ---

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/e1071\_1.7-3.zip'

Content type 'application/zip' length 1011333 bytes (987 KB)

downloaded 987 KB

package ‘e1071’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\poona\AppData\Local\Temp\RtmpAFIXQq\downloaded\_packages

> C

function (object, contr, how.many, ...)

{

if (isFALSE(as.logical(Sys.getenv("\_R\_OPTIONS\_STRINGS\_AS\_FACTORS\_"))))

object <- as.factor(object)

if (!nlevels(object))

stop("object not interpretable as a factor")

if (!missing(contr) && is.name(Xcontr <- substitute(contr)))

contr <- switch(as.character(Xcontr), poly = "contr.poly",

helmert = "contr.helmert", sum = "contr.sum",

treatment = "contr.treatment", SAS = "contr.SAS",

contr)

if (missing(contr)) {

oc <- getOption("contrasts")

contr <- if (length(oc) < 2L)

if (is.ordered(object))

contr.poly

else contr.treatment

else oc[1 + is.ordered(object)]

}

if (missing(how.many) && missing(...))

contrasts(object) <- contr

else {

if (is.character(contr))

contr <- get(contr, mode = "function")

if (is.function(contr))

contr <- contr(nlevels(object), ...)

contrasts(object, how.many) <- contr

}

object

}

<bytecode: 0x00000000155a10c0>

<environment: namespace:stats>

> library(class)

> library(e1071)

> classifier <- naiveBayes(iris[,1:4], iris[,5])

> table(predict(classifier, iris[,-5]), iris[,5])

setosa versicolor virginica

setosa 50 0 0

versicolor 0 47 3

virginica 0 3 47

> install.packages("caret")

Installing package into ‘C:/Users/poona/Documents/R/win-library/4.0’

(as ‘lib’ is unspecified)

also installing the dependencies ‘numDeriv’, ‘SQUAREM’, ‘lava’, ‘prodlim’, ‘iterators’, ‘gower’, ‘ipred’, ‘timeDate’, ‘foreach’, ‘ModelMetrics’, ‘recipes’, ‘pROC’

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/numDeriv\_2016.8-1.1.zip'

Content type 'application/zip' length 116026 bytes (113 KB)

downloaded 113 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/SQUAREM\_2020.3.zip'

Content type 'application/zip' length 180324 bytes (176 KB)

downloaded 176 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/lava\_1.6.7.zip'

Content type 'application/zip' length 2195067 bytes (2.1 MB)

downloaded 2.1 MB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/prodlim\_2019.11.13.zip'

Content type 'application/zip' length 421259 bytes (411 KB)

downloaded 411 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/iterators\_1.0.12.zip'

Content type 'application/zip' length 343828 bytes (335 KB)

downloaded 335 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/gower\_0.2.2.zip'

Content type 'application/zip' length 298201 bytes (291 KB)

downloaded 291 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/ipred\_0.9-9.zip'

Content type 'application/zip' length 399049 bytes (389 KB)

downloaded 389 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/timeDate\_3043.102.zip'

Content type 'application/zip' length 1551588 bytes (1.5 MB)

downloaded 1.5 MB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/foreach\_1.5.0.zip'

Content type 'application/zip' length 145823 bytes (142 KB)

downloaded 142 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/ModelMetrics\_1.2.2.2.zip'

Content type 'application/zip' length 849604 bytes (829 KB)

downloaded 829 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/recipes\_0.1.13.zip'

Content type 'application/zip' length 1137965 bytes (1.1 MB)

downloaded 1.1 MB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/pROC\_1.16.2.zip'

Content type 'application/zip' length 1509434 bytes (1.4 MB)

downloaded 1.4 MB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/caret\_6.0-86.zip'

Content type 'application/zip' length 6253812 bytes (6.0 MB)

downloaded 6.0 MB

package ‘numDeriv’ successfully unpacked and MD5 sums checked

package ‘SQUAREM’ successfully unpacked and MD5 sums checked

package ‘lava’ successfully unpacked and MD5 sums checked

package ‘prodlim’ successfully unpacked and MD5 sums checked

package ‘iterators’ successfully unpacked and MD5 sums checked

package ‘gower’ successfully unpacked and MD5 sums checked

package ‘ipred’ successfully unpacked and MD5 sums checked

package ‘timeDate’ successfully unpacked and MD5 sums checked

package ‘foreach’ successfully unpacked and MD5 sums checked

package ‘ModelMetrics’ successfully unpacked and MD5 sums checked

package ‘recipes’ successfully unpacked and MD5 sums checked

package ‘pROC’ successfully unpacked and MD5 sums checked

package ‘caret’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\poona\AppData\Local\Temp\RtmpAFIXQq\downloaded\_packages

> install.packages("klaR")

Installing package into ‘C:/Users/poona/Documents/R/win-library/4.0’

(as ‘lib’ is unspecified)

also installing the dependencies ‘R.methodsS3’, ‘R.oo’, ‘R.utils’, ‘rlang’, ‘httpuv’, ‘htmltools’, ‘sourcetools’, ‘later’, ‘promises’, ‘fastmap’, ‘commonmark’, ‘R.cache’, ‘rematch2’, ‘vctrs’, ‘shiny’, ‘miniUI’, ‘styler’, ‘classInt’, ‘labelled’, ‘haven’, ‘combinat’, ‘questionr’

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/R.methodsS3\_1.8.0.zip'

Content type 'application/zip' length 79954 bytes (78 KB)

downloaded 78 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/R.oo\_1.23.0.zip'

Content type 'application/zip' length 957777 bytes (935 KB)

downloaded 935 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/R.utils\_2.9.2.zip'

Content type 'application/zip' length 1404878 bytes (1.3 MB)

downloaded 1.3 MB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/rlang\_0.4.7.zip'

Content type 'application/zip' length 1129581 bytes (1.1 MB)

downloaded 1.1 MB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/httpuv\_1.5.4.zip'

Content type 'application/zip' length 1725904 bytes (1.6 MB)

downloaded 1.6 MB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/htmltools\_0.5.0.zip'

Content type 'application/zip' length 231227 bytes (225 KB)

downloaded 225 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/sourcetools\_0.1.7.zip'

Content type 'application/zip' length 691355 bytes (675 KB)

downloaded 675 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/later\_1.1.0.1.zip'

Content type 'application/zip' length 867483 bytes (847 KB)

downloaded 847 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/promises\_1.1.1.zip'

Content type 'application/zip' length 1970000 bytes (1.9 MB)

downloaded 1.9 MB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/fastmap\_1.0.1.zip'

Content type 'application/zip' length 196578 bytes (191 KB)

downloaded 191 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/commonmark\_1.7.zip'

Content type 'application/zip' length 265079 bytes (258 KB)

downloaded 258 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/R.cache\_0.14.0.zip'

Content type 'application/zip' length 104088 bytes (101 KB)

downloaded 101 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/rematch2\_2.1.2.zip'

Content type 'application/zip' length 47445 bytes (46 KB)

downloaded 46 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/vctrs\_0.3.2.zip'

Content type 'application/zip' length 1162041 bytes (1.1 MB)

downloaded 1.1 MB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/shiny\_1.5.0.zip'

Content type 'application/zip' length 5359829 bytes (5.1 MB)

downloaded 5.1 MB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/miniUI\_0.1.1.1.zip'

Content type 'application/zip' length 36626 bytes (35 KB)

downloaded 35 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/styler\_1.3.2.zip'

Content type 'application/zip' length 626343 bytes (611 KB)

downloaded 611 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/classInt\_0.4-3.zip'

Content type 'application/zip' length 466238 bytes (455 KB)

downloaded 455 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/labelled\_2.5.0.zip'

Content type 'application/zip' length 210416 bytes (205 KB)

downloaded 205 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/haven\_2.3.1.zip'

Content type 'application/zip' length 1248806 bytes (1.2 MB)

downloaded 1.2 MB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/combinat\_0.0-8.zip'

Content type 'application/zip' length 42745 bytes (41 KB)

downloaded 41 KB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/questionr\_0.7.1.zip'

Content type 'application/zip' length 1956803 bytes (1.9 MB)

downloaded 1.9 MB

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.0/klaR\_0.6-15.zip'

Content type 'application/zip' length 566920 bytes (553 KB)

downloaded 553 KB

package ‘R.methodsS3’ successfully unpacked and MD5 sums checked

package ‘R.oo’ successfully unpacked and MD5 sums checked

package ‘R.utils’ successfully unpacked and MD5 sums checked

package ‘rlang’ successfully unpacked and MD5 sums checked

package ‘httpuv’ successfully unpacked and MD5 sums checked

package ‘htmltools’ successfully unpacked and MD5 sums checked

package ‘sourcetools’ successfully unpacked and MD5 sums checked

package ‘later’ successfully unpacked and MD5 sums checked

package ‘promises’ successfully unpacked and MD5 sums checked

package ‘fastmap’ successfully unpacked and MD5 sums checked

package ‘commonmark’ successfully unpacked and MD5 sums checked

package ‘R.cache’ successfully unpacked and MD5 sums checked

package ‘rematch2’ successfully unpacked and MD5 sums checked

package ‘vctrs’ successfully unpacked and MD5 sums checked

package ‘shiny’ successfully unpacked and MD5 sums checked

package ‘miniUI’ successfully unpacked and MD5 sums checked

package ‘styler’ successfully unpacked and MD5 sums checked

package ‘classInt’ successfully unpacked and MD5 sums checked

package ‘labelled’ successfully unpacked and MD5 sums checked

package ‘haven’ successfully unpacked and MD5 sums checked

package ‘combinat’ successfully unpacked and MD5 sums checked

package ‘questionr’ successfully unpacked and MD5 sums checked

package ‘klaR’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\poona\AppData\Local\Temp\RtmpAFIXQq\downloaded\_packages

> library(caret)

Loading required package: lattice

Loading required package: ggplot2

> library(klaR)

Loading required package: MASS

> x.train = iris[ train, -5]

> y.train = iris[ train, 5]

> x.test=iris[ -train, -5]

> y.test=iris[ -train, 5]

> model = train(x.train,y.train,'nb',trControl=trainControl(method='cv',number=10))

> table(predict(model$finalModel, x.test)$class, y.test)

y.test

setosa versicolor virginica

setosa 19 0 0

versicolor 0 17 0

virginica 0 2 12

> model =train(TrainStLnX.IBM,TrainY.IBM,'nb',trControl=trainControl(method='cv',number=10))

Error in train(TrainStLnX.IBM, TrainY.IBM, "nb", trControl = trainControl(method = "cv", :

object 'TrainStLnX.IBM' not found

>