Group Assignment

Nirav Mahnot

13/10/2022

Clean-up the environment

```
rm(list = ls())
```

##Adding library

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v tibble  3.1.6     v purrr   0.3.4
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x purrr::lift()   masks caret::lift()
```

Set working directory and load the data into workspace

```
getwd()
```

```
## [1] "C:/Users/Mahno/Desktop/Useless 5/Machine Learning"
```

```
setwd("C:/Users/Mahno/Desktop/Useless 5/Machine Learning/Data")
getwd()
```

```
## [1] "C:/Users/Mahno/Desktop/Useless 5/Machine Learning/Data"
```

```
boston_data <- read.csv("boston.csv", stringsAsFactors = TRUE)
boston_save = boston_data
```

Check and validate the type of the variables

```
str(boston_data)
```

```
## 'data.frame':    506 obs. of  14 variables:
##  $ crim   : num  0.00632 0.02731 0.02729 0.03237 0.06905 ...
##  $ zn     : num  18 0 0 0 0 0 12.5 12.5 12.5 12.5 ...
##  $ indus  : num  2.31 7.07 7.07 2.18 2.18 2.18 7.87 7.87 7.87 7.87 ...
##  $ chas   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ nox    : num  0.538 0.469 0.469 0.458 0.458 0.458 0.524 0.524 0.524 0.524 ...
##  $ rm     : num  6.58 6.42 7.18 7 7.15 ...
##  $ age    : num  65.2 78.9 61.1 45.8 54.2 58.7 66.6 96.1 100 85.9 ...
##  $ dis    : num  4.09 4.97 4.97 6.06 6.06 ...
##  $ rad    : int  1 2 2 3 3 3 5 5 5 5 ...
##  $ tax    : int  296 242 242 222 222 222 311 311 311 311 ...
##  $ ptratio: num  15.3 17.8 17.8 18.7 18.7 18.7 15.2 15.2 15.2 15.2 ...
##  $ black  : num  397 397 393 395 397 ...
##  $ lstat  : num  4.98 9.14 4.03 2.94 5.33 ...
##  $ medv   : num  24 21.6 34.7 33.4 36.2 28.7 22.9 27.1 16.5 18.9 ...
```

##Train-Test split

```
set.seed(123)
train_sample <- createDataPartition(boston_data$crim, p = 0.8, list = FALSE)
str(train_sample)
```

```
##  int [1:406, 1] 3 4 5 6 7 8 11 12 13 14 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ : NULL
##   ..$ : chr "Resample1"
```

```
boston_train <- boston_data[train_sample, ]
boston_test  <- boston_data[-train_sample, ]
```

##SLR - lm

```
model <- lm(medv~., data = boston_train)
summary(model)
```

```
##
## Call:
## lm(formula = medv ~ ., data = boston_train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.9435  -2.6157  -0.4668   1.8342  24.5696
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  35.693957   5.639032   6.330 6.72e-10 ***
## crim         -0.099886   0.041387  -2.413  0.01626 *
## zn            0.040829   0.015166   2.692  0.00740 **
## indus         0.010682   0.068508   0.156  0.87617
## chas          3.022610   1.011527   2.988  0.00298 **
## nox         -17.505818   4.298879  -4.072 5.64e-05 ***
## rm            3.746137   0.450178   8.321 1.46e-15 ***
## age           0.016893   0.015099   1.119  0.26391
## dis          -1.399007   0.223131  -6.270 9.55e-10 ***
## rad           0.315399   0.072779   4.334 1.87e-05 ***
## tax          -0.011357   0.004147  -2.738  0.00646 **
## ptratio      -0.977137   0.145266  -6.727 6.17e-11 ***
## black         0.011703   0.002933   3.990 7.88e-05 ***
## lstat        -0.621988   0.057071 -10.898  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.647 on 392 degrees of freedom
## Multiple R-squared:  0.761,  Adjusted R-squared:  0.7531
## F-statistic: 96.02 on 13 and 392 DF,  p-value: < 2.2e-16
```

```
predictions <- predict(model, boston_test)
summary(predictions)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -6.216  16.985  22.450  22.263  26.861  44.690
```

```
data.frame( R2 = R2(predictions, boston_test$medv),
            RMSE = RMSE(predictions, boston_test$medv),
            MAE = MAE(predictions, boston_test$medv))
```

```
##          R2     RMSE      MAE
## 1 0.6633282 5.286544 3.693063
```

##SLR - Train

```
model2 <- train(medv ~., data = boston_train,
                method = "lm")

predictions2 <- predict(model2, boston_test)
summary(predictions2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -6.216  16.985  22.450  22.263  26.861  44.690
```

```
data.frame( R2 = R2(predictions2, boston_test$medv),
            RMSE = RMSE(predictions2, boston_test$medv),
            MAE = MAE(predictions2, boston_test$medv))
```

```
##          R2     RMSE      MAE
## 1 0.6633282 5.286544 3.693063
```

##SLR - Train + CV

```
train_control <- trainControl(method = "cv",
                              number = 10)
model3 <- train(medv ~., data = boston_train,
                method = "lm",
                trControl = train_control)

predictions3 <- predict(model3, boston_test)
summary(predictions3)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -6.216  16.985  22.450  22.263  26.861  44.690
```

```
data.frame( R2 = R2(predictions3, boston_test$medv),
            RMSE = RMSE(predictions3, boston_test$medv),
            MAE = MAE(predictions3, boston_test$medv))
```

```
##          R2     RMSE      MAE
## 1 0.6633282 5.286544 3.693063
```

Ridge with lambda = 100

```
lambda = 100
ridge <- train(
  medv ~., data = boston_train, method = "glmnet",
  trControl = trainControl("cv", number = 10),
  tuneGrid = expand.grid(alpha = 0, lambda = lambda)
)
coef(ridge$finalModel, ridge$bestTune$lambda)
```

```
## 14 x 1 sparse Matrix of class "dgCMatrix"
##                       s1
## (Intercept) 23.794844456
## crim        -0.028304153
## zn           0.008211273
## indus       -0.038037022
## chas         0.543963743
## nox         -1.825531045
## rm           0.683358627
## age         -0.006286560
## dis          0.038185661
## rad         -0.022072024
## tax         -0.001528266
## ptratio     -0.160938534
## black        0.002212812
## lstat       -0.069721855
```

```
predictions4 <- predict(ridge, boston_test)
data.frame( R2 = R2(predictions4, boston_test$medv),
            RMSE = RMSE(predictions4, boston_test$medv),
            MAE = MAE(predictions4, boston_test$medv))
```

```
##          R2     RMSE      MAE
## 1 0.5275031 7.077353 5.026608
```

##Ridge with lambda = 10

```
lambda = 10
ridge <- train(
  medv ~., data = boston_train, method = "glmnet",
  trControl = trainControl("cv", number = 10),
  tuneGrid = expand.grid(alpha = 0, lambda = lambda)
)
coef(ridge$finalModel, ridge$bestTune$lambda)
```

```
## 14 x 1 sparse Matrix of class "dgCMatrix"
##                       s1
## (Intercept) 21.222408202
## crim        -0.061146842
## zn           0.016487504
## indus       -0.075949886
## chas         2.395891325
## nox         -3.430104321
## rm           2.835503904
## age         -0.006173919
## dis         -0.235153596
## rad         -0.006480343
## tax         -0.002789962
## ptratio     -0.565658863
## black        0.007083729
## lstat       -0.277150993
```

```
predictions5 <- predict(ridge, boston_test)
data.frame( R2 = R2(predictions5, boston_test$medv),
            RMSE = RMSE(predictions5, boston_test$medv),
            MAE = MAE(predictions5, boston_test$medv))
```

```
##          R2     RMSE      MAE
## 1 0.6365168 5.227731 3.721245
```

##Lasso with lambda = 0.8

```
lambda = 0.8
lasso <- train(
  medv ~., data = boston_train, method = "glmnet",
  trControl = trainControl("cv", number = 10),
  tuneGrid = expand.grid(alpha = 1, lambda = lambda)
)
coef(lasso$finalModel, lasso$bestTune$lambda)
```

```
## 14 x 1 sparse Matrix of class "dgCMatrix"
##                       s1
## (Intercept) 13.983408784
## crim         .
## zn           .
## indus        .
## chas         0.668305818
## nox          .
## rm           4.035643042
## age          .
## dis          .
## rad          .
## tax          .
## ptratio     -0.638058179
## black        0.005027811
## lstat       -0.551484953
```

```
predictions6 <- predict(lasso, boston_test)
data.frame( R2 = R2(predictions6, boston_test$medv),
            RMSE = RMSE(predictions6, boston_test$medv),
            MAE = MAE(predictions6, boston_test$medv))
```

```
##          R2     RMSE      MAE
## 1 0.6051045 5.429482 3.705982
```

Lasso with lambda = 0.01

```
lambda = 0.01
lasso <- train(
  medv ~., data = boston_train, method = "glmnet",
  trControl = trainControl("cv", number = 10),
  tuneGrid = expand.grid(alpha = 1, lambda = lambda)
)
coef(lasso$finalModel, lasso$bestTune$lambda)
```

```
## 14 x 1 sparse Matrix of class "dgCMatrix"
##                       s1
## (Intercept)  34.82146946
## crim         -0.09517025
## zn            0.03864489
## indus         .
## chas          3.02153741
## nox         -16.74515048
## rm            3.77652310
## age           0.01536689
## dis          -1.37398954
## rad           0.29374599
## tax          -0.01034104
## ptratio      -0.96754527
## black         0.01164556
## lstat        -0.61950269
```

```
predictions7 <- predict(lasso, boston_test)
data.frame( R2 = R2(predictions7, boston_test$medv),
            RMSE = RMSE(predictions7, boston_test$medv),
            MAE = MAE(predictions7, boston_test$medv))
```

```
##          R2     RMSE      MAE
## 1 0.6629065 5.279182 3.678083
```