STT3851 Homework 9

Dr. Hasthika Rupasinghe

Due – April 14

Get data: Go to https://archive.ics.uci.edu/ml/datasets/Real+estate+valuation+data+set. You can read about the data set here. Then click on "Data Folder". Finally click on the "Real estate valuation data set.xlsx". Rename your file to "RealEstate.xlsx" and upload it to your Homework folder. Be sure to have your homework and the data set in the same folder.

Load data: Run the following to load data.

```
library(readxl)
RealEstate <- read_excel("RealEstate.xlsx")
head(RealEstate)
```

```
## # A tibble: 6 x 8
##      No `X1 transaction date` `X2 house age` X3 distance to the nearest MRT st~1
##   <dbl>                 <dbl>          <dbl>                                <dbl>
## 1     1                 2013.           32                                   84.9
## 2     2                 2013.           19.5                                307.
## 3     3                 2014.           13.3                                562.
## 4     4                 2014.           13.3                                562.
## 5     5                 2013.            5                                  391.
## 6     6                 2013.            7.1                               2175.
## # i abbreviated name: 1: `X3 distance to the nearest MRT station`
## # i 4 more variables: `X4 number of convenience stores` <dbl>,
## #   `X5 latitude` <dbl>, `X6 longitude` <dbl>,
## #   `Y house price of unit area` <dbl>
```

Rename columns: Run the following to rename columns.

```
names(RealEstate) <- c("No", "TrDate", "HouseAge", "DisMRT", "NoConve", "lat", "lon", "Price")
head(RealEstate)
```

```
## # A tibble: 6 x 8
##      No TrDate HouseAge DisMRT NoConve   lat   lon Price
##   <dbl>  <dbl>    <dbl>  <dbl>   <dbl> <dbl> <dbl> <dbl>
## 1     1  2013.     32     84.9      10  25.0  122.  37.9
## 2     2  2013.     19.5  307.        9  25.0  122.  42.2
## 3     3  2014.     13.3  562.        5  25.0  122.  47.3
## 4     4  2014.     13.3  562.        5  25.0  122.  54.8
## 5     5  2013.      5    391.        5  25.0  122.  43.1
## 6     6  2013.      7.1 2175.        3  25.0  122.  32.1
```

Goal is to predict the house price.

1) Inspect the correlations between the response and the predictors.
2) Create a multiple/simple linear regression model and find the validation set error.
3) Create another multiple linear regression model and find the validation set error.

4) Create another multiple linear regression model and find the validation set error.

5) Pick the best model using the validation set approach.

---

6) Create a multiple/simple linear regression model and find 10-fold CV error.

7) Create another multiple linear regression model and find the 10-fold CV error.

8) Create another multiple linear regression model and find the 10-fold CV error.

9) Pick the best model using the 10-fold CV error.