

# IE517 MLF F20

## Module 4 Homework (Regression)

Yaxuan Wang (662869931)

**Out[3]:** The raw code for this IPython notebook is by default hidden for easier reading. To toggle on/off the raw code, click [here](#).

**Out[2]:**

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LST.
0	0.00632	18.0	2.31	0	0.538	6.575	65.2	4.0900	1	296.0	15.3	396.90	4.
1	0.02731	0.0	7.07	0	0.469	6.421	78.9	4.9671	2	242.0	17.8	396.90	9.
2	0.02729	0.0	7.07	0	0.469	7.185	61.1	4.9671	2	242.0	17.8	392.83	4.
3	0.03237	0.0	2.18	0	0.458	6.998	45.8	6.0622	3	222.0	18.7	394.63	2.
4	0.06905	0.0	2.18	0	0.458	7.147	54.2	6.0622	3	222.0	18.7	396.90	5.

## Part 1: Exploratory Data Analysis

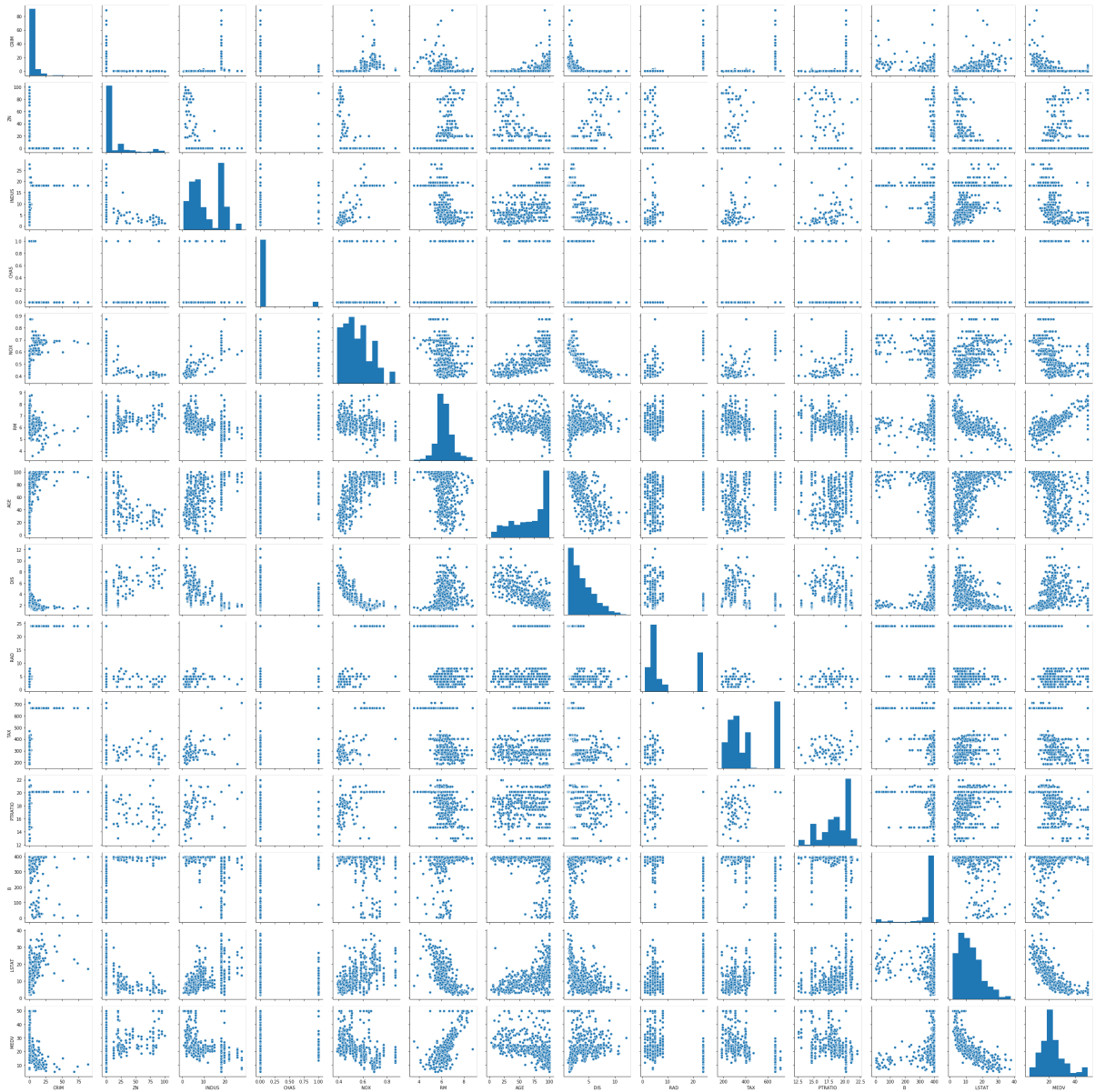
### Summary Statistic

**Out[24]:**

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	
count	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.
mean	3.613524	11.363636	11.136779	0.069170	0.554695	6.284634	68.574901	3.
std	8.601545	23.322453	6.860353	0.253994	0.115878	0.702617	28.148861	2.
min	0.006320	0.000000	0.460000	0.000000	0.385000	3.561000	2.900000	1.
25%	0.082045	0.000000	5.190000	0.000000	0.449000	5.885500	45.025000	2.
50%	0.256510	0.000000	9.690000	0.000000	0.538000	6.208500	77.500000	3.
75%	3.677082	12.500000	18.100000	0.000000	0.624000	6.623500	94.075000	5.
max	88.976200	100.000000	27.740000	1.000000	0.871000	8.780000	100.000000	12.

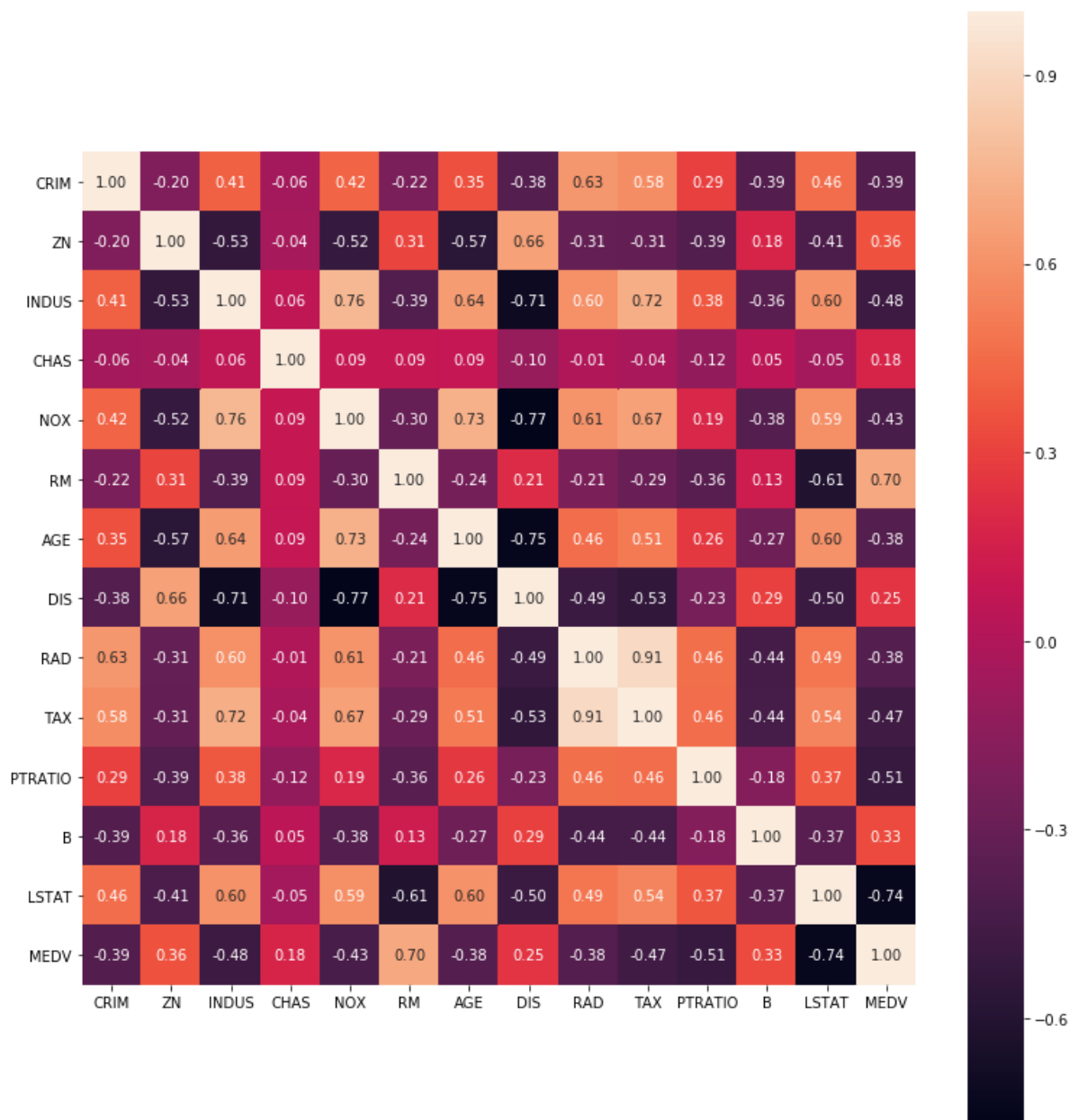
As we can see from the summary statistic, this dataset contains 506 samples, 13 features, and one target variable. All the variables are continuous.

## Scatterplot matrix



If we focus on the target variable, MEDV, we can observe that LSTAT, and RM, and CRIM has some positive or negative linear relationship with MEDV, although there are some noise.

## Heatmap/Correlation Matrix



The heatmap shows the Pearson's correlation between each pair of variables. Let's still focus on our target variable, MEDV . We can see that the largest correlation with MEDV is LSTAT , whose correlation is -0.74. And the second large correlated explanatory variable is RM , whose correlation is 0.70. This observation actually have the same conclusion we draw from scatter plot. Thus, in the next part, we will test the linear regression model on those high correlated variables.

## Split Data into Training and Testing Sets

Now in order to test and compare our model performance in next part, we are splitting data into training and test sets. I use 80% of the data for the training set and the rest of 20% as our testing set.

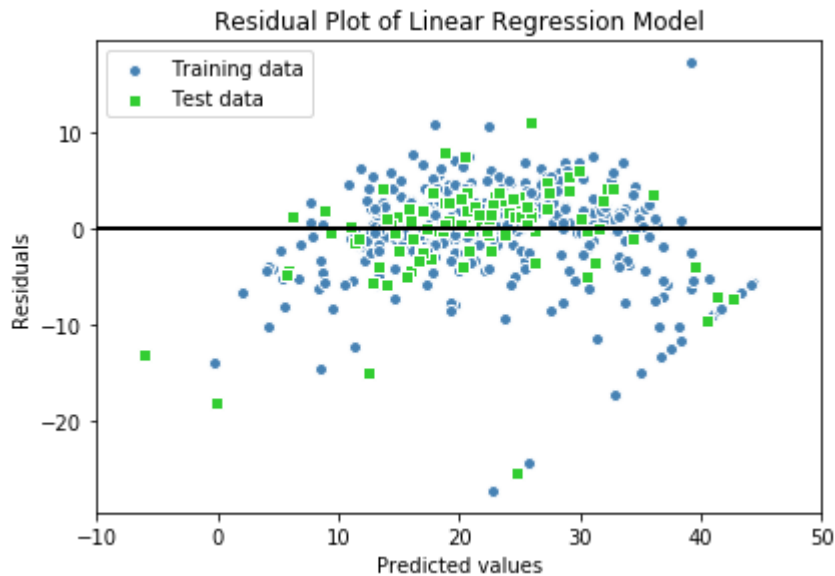
## Part 2: Linear Regression

### The Fitted Model

```
[ -1.13055924e-01  3.01104641e-02  4.03807204e-02  2.78443820e+00  
 -1.72026334e+01  4.43883520e+00 -6.29636221e-03 -1.44786537e+00  
  2.62429736e-01 -1.06467863e-02 -9.15456240e-01  1.23513347e-02  
 -5.08571424e-01]
```

The above are the coefficients for this linear regression model. The first number  $-1.13055924 \times 10^{-1}$  is our y-intercept when all x's are zero. The second number  $3.01104641 \times 10^{-2}$  are the coefficient for the explanatory variable CRIM. And the rest numbers are coefficients for each features.

### Residual Plot of Our Fitted Linear Regression Model



From the residual plot, we can see that most of the errors are randomly distributed and they are scattered around the center line, which means most of explanatory information are captured in this linear regression model and have small residuals. However, there is a pattern in the right bottom corner. Some residuals form a straight line, which means some explanatory info did not been captured. And those points have relatively large residuals.

### Performance Metrics by MSE

The MSE train: 21.641, test: 24.291

We can see that the Mean Squared Error (MSE) for the training set is 21.641, while the one for the testing set is 24.291, which is relatively larger. It can indicate that our model is overfitting the training set. Thus, we would need to use regularized regression in the next part to prevent overfitting.

## Performance Metrics by Coefficient of Determination

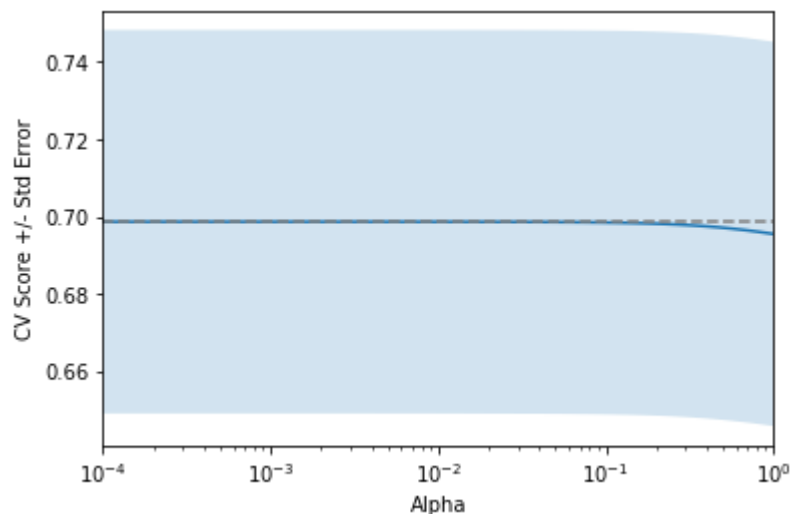
The coefficient of determination train: 0.751, test: 0.669

We can see that the coefficient of determination,  $R^2$ , for the training set is 0.751, which is pretty high. So the linear regression model fits the training set good. The  $R^2$  for the testing set is 0.669, which is only a little bit lower. Still not bad.

## Part 3 Regularized Linear Regression

### Part 3.1: Ridge Regression

#### Best Alpha for Ridge by Comparing CV Score



The above plot shows the average scores of Ridge Regression by 10-cross-validation tested by different alpha. As the alpha increase, the CV scores decreased after  $\alpha = 10^{-2}$ . Thus, I will pick  $\alpha = 10^{-2}$  as our best alpha for this data set.

#### The Fitted Model

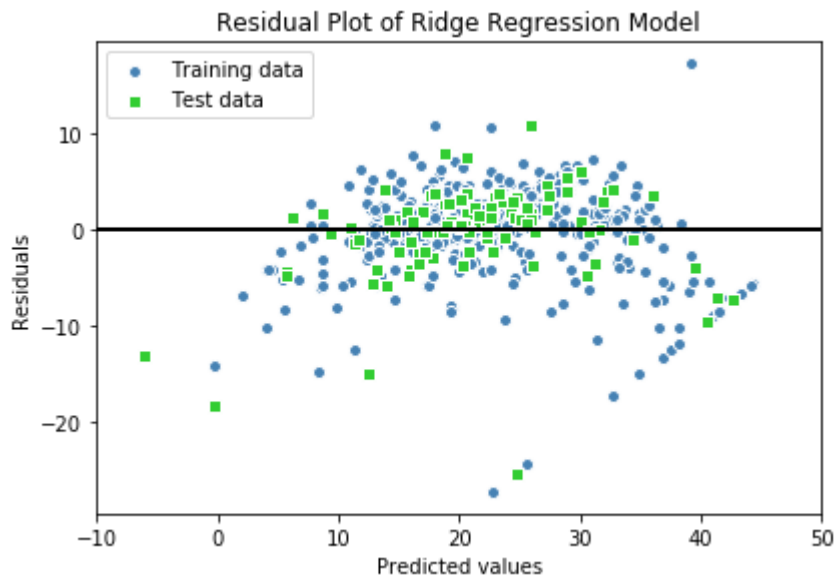
```

[-1.12985398e-01  3.01477290e-02  3.97929679e-02  2.78084775e+00
 -1.70656728e+01  4.43960807e+00 -6.40468889e-03 -1.44590837e+00
 2.62172574e-01 -1.06610344e-02 -9.13880463e-01  1.23566074e-02
 -5.08818863e-01]

```

The above are the coefficients for this ridge regression model. The first number  $-1.12985398 \times 10^{-1}$  is our y-intercept when all x's are zero. The second number  $3.01477290 \times 10^{-2}$  are the coefficient for the explanatory variable CRIM. And the rest numbers are coefficients for each features.

### Residual Plot of Our Fitted Ridge Regression Model



The residual plot of ridge regression actually is the same as our linear regression. This is because our optimal alpha, 0.01, is too small. When alpha is too small, the ridge regression is very similar to linear regression.

### Performance Metrics by MSE

The MSE train: 21.646, test: 24.301

The Mean Squared Error (MSE) for the training set is 21.641, while the one for the testing set is 24.291, which is relatively larger. Note that the MSE are still the same as the linear regression due to the small alpha.

### Performance Metrics by Coefficient of Determination

The coefficient of determination train: 0.751, test: 0.669

The  $R^2$  for the training set is 0.751, which is pretty high. And the  $R^2$  for the testing set is 0.669, which is only a little bit lower. Note that the MSE are still the same as the linear regression due to the small alpha.

### **Best Alpha by Comparing MSE and $R^2$**

When alpha = 0.0001 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.00012067926406393288 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.00014563484775012445 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.00017575106248547912 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.00021209508879201905 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.0002559547922699536 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.00030888435964774815 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.0003727593720314938 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.0004498432668969444 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.0005428675439323859 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.0006551285568595509 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.0007906043210907702 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.0009540954763499944 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.0011513953993264468 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.0013894954943731374 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.0016768329368110067 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.0020235896477251557 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.0024420530945486497 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.0029470517025518097 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.0035564803062231283 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.004291934260128779 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.005179474679231213 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.0062505519252739694 , The MSE train: 21.641, test: 24.292, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.007543120063354615 , The MSE train: 21.641, test: 24.292, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.009102981779915217 , The MSE train: 21.641, test: 24.292, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.010985411419875584 , The MSE train: 21.641, test: 24.292, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.013257113655901081 , The MSE train: 21.642, test: 24.292, and R<sup>2</sup> train: 0.751, test: 0.669  
When alpha = 0.015998587196060572 , The MSE train: 21.642, test: 24.292, and R<sup>2</sup> train: 0.751, test: 0.669



When alpha = 0.019306977288832496 , The MSE train: 21.642, test: 24.292,  
and R^2 train: 0.751, test: 0.669

When alpha = 0.023299518105153717 , The MSE train: 21.642, test: 24.293,  
and R^2 train: 0.751, test: 0.669

When alpha = 0.02811768697974228 , The MSE train: 21.642, test: 24.293, a  
nd R^2 train: 0.751, test: 0.669

When alpha = 0.033932217718953266 , The MSE train: 21.642, test: 24.294,  
and R^2 train: 0.751, test: 0.669

When alpha = 0.040949150623804234 , The MSE train: 21.642, test: 24.294,  
and R^2 train: 0.751, test: 0.669

When alpha = 0.04941713361323833 , The MSE train: 21.643, test: 24.295, a  
nd R^2 train: 0.751, test: 0.669

When alpha = 0.05963623316594643 , The MSE train: 21.643, test: 24.296, a  
nd R^2 train: 0.751, test: 0.669

When alpha = 0.07196856730011514 , The MSE train: 21.644, test: 24.298, a  
nd R^2 train: 0.751, test: 0.669

When alpha = 0.08685113737513521 , The MSE train: 21.645, test: 24.299, a  
nd R^2 train: 0.751, test: 0.669

When alpha = 0.10481131341546852 , The MSE train: 21.647, test: 24.302, a  
nd R^2 train: 0.751, test: 0.669

When alpha = 0.12648552168552957 , The MSE train: 21.649, test: 24.305, a  
nd R^2 train: 0.751, test: 0.669

When alpha = 0.15264179671752318 , The MSE train: 21.652, test: 24.309, a  
nd R^2 train: 0.751, test: 0.669

When alpha = 0.18420699693267145 , The MSE train: 21.657, test: 24.314, a  
nd R^2 train: 0.751, test: 0.668

When alpha = 0.22229964825261933 , The MSE train: 21.662, test: 24.321, a  
nd R^2 train: 0.751, test: 0.668

When alpha = 0.2682695795279725 , The MSE train: 21.670, test: 24.329, an  
d R^2 train: 0.751, test: 0.668

When alpha = 0.32374575428176433 , The MSE train: 21.680, test: 24.341, a  
nd R^2 train: 0.750, test: 0.668

When alpha = 0.3906939937054613 , The MSE train: 21.694, test: 24.354, an  
d R^2 train: 0.750, test: 0.668

When alpha = 0.47148663634573895 , The MSE train: 21.711, test: 24.372, a  
nd R^2 train: 0.750, test: 0.668

When alpha = 0.5689866029018293 , The MSE train: 21.732, test: 24.392, an  
d R^2 train: 0.750, test: 0.667

When alpha = 0.6866488450042998 , The MSE train: 21.758, test: 24.417, an  
d R^2 train: 0.750, test: 0.667

When alpha = 0.8286427728546842 , The MSE train: 21.788, test: 24.445, an  
d R^2 train: 0.749, test: 0.667

When alpha = 1.0 , The MSE train: 21.824, test: 24.477, and R^2 train: 0.  
749, test: 0.666

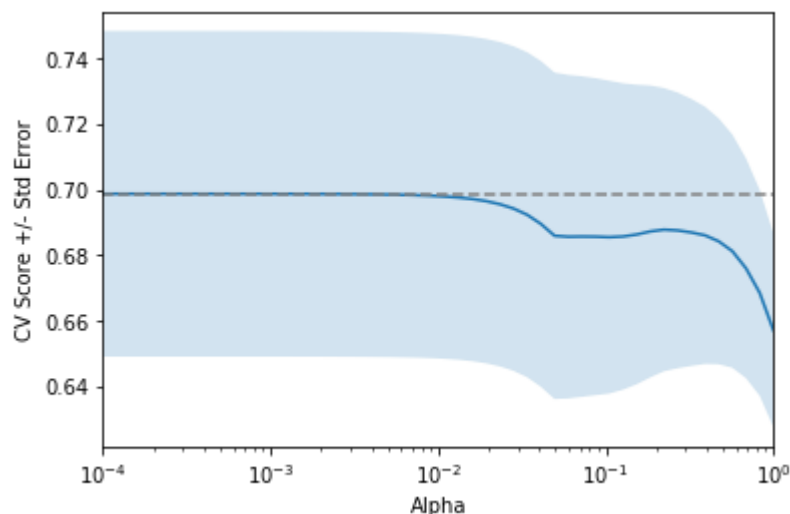
Let's mainly focus on the testing set's MSE and  $R^2$ .

For MSE, as alpha increase, the MSE first keeps constant of 24.291 from alpha=0.0001 to alpha = 0.005179474679231213, then keeps increasing. Thus, the MSE for testing test of 24.291 is the smallest. The best alpha will be any one of the alpha from 0.0001 to 0.005179474679231213 for the testing set since its MSE is the smallest.

For  $R^2$ , as alpha increase,  $R^2$  first keeps the same of 0.669, then starts decreasing. When alpha = 1.0, the  $R^2$  reaches the smallest value, 0.666. Thus, the best alpha will be any alpha from 0.001 to 0.15264179671752318 since its  $R^2$  is the largest, if choosing  $R^2$  as our performance metric.

## Part 3.2: LASSO Regression

### Best Alpha for LASSO by Comparing CV Score



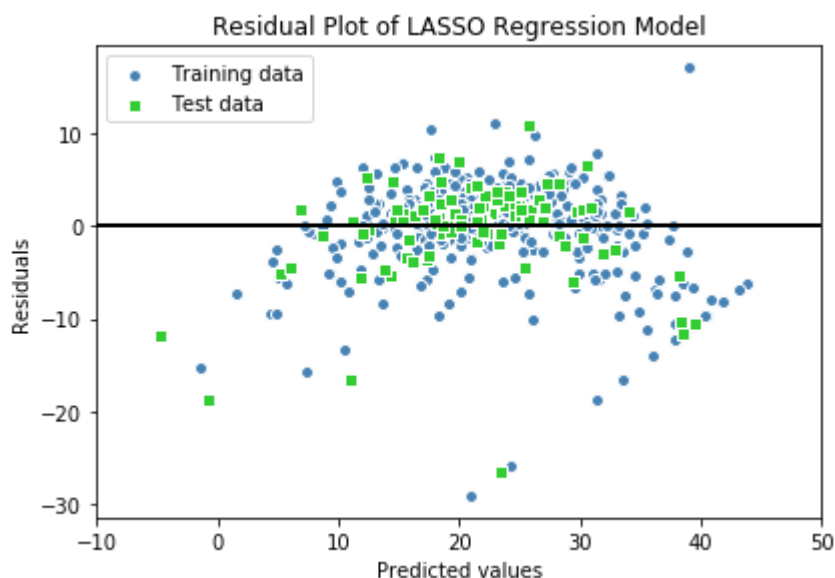
The above plot shows the average scores of LASSO Regression by 10-cross-validation tested by different alpha. As the alpha increase, the CV scores decreased after alpha =  $8 \times 10^{-3}$ . Thus, I will pick alpha =  $8 \times 10^{-3}$  as our best alpha for this data set. Notice that our alpha is really small, so we expect the performance metric of LASSO regression will be very similar to our linear regression.

### The Fitted Model

```
[-6.14780468e-02  2.54179758e-03 -0.00000000e+00  2.58754176e+00
-1.05582205e+01  4.73040570e+00 -0.00000000e+00 -8.66694594e-01
 7.31922216e-03 -0.00000000e+00 -8.01300261e-01  1.06024400e-02
-5.05074359e-01]
```

The above are the coefficients for this LASSO regression model. The first number  $-6.14780468 \times 10^{-2}$  is our y-intercept when all x's are zero. The second number  $3.2.54179758 \times 10^{-3}$  are the coefficient for the explanatory variable CRIM . And the rest numbers are coefficients for each features.

### Residual Plot of Our Fitted LASSO Regression Model



This is exactly the same issue with Ridge regression. The residual plot is the same as our linear regression. This is because our optimal alpha, 0.008, is too small. When alpha is too small, the LASSO regression is very similar to linear regression.

### Performance Metrics by MSE

The MSE train: 22.717, test: 26.724

The Mean Squared Error (MSE) for the training set is 22.717, while the one for the testing set is 26.724, which is relatively larger. The MSE's of LASSO regression are both a bit bigger than the linear regression and ridge regression.

### Performance Metrics by Coefficient of Determination

The coefficient of determination train: 0.739, test: 0.636

The  $R^2$  for the training set is 0.739, which is pretty high. And the  $R^2$  for the testing set is 0.636, which is only a little bit lower. The  $R^2$ 's of LASSO regression are both a smaller than the linear regression and ridge regression.

### **Best Alpha by Compareing MSE and $R^2$**

When alpha = 0.0001 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.00012067926406393288 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.00014563484775012445 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.00017575106248547912 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.00021209508879201905 , The MSE train: 21.641, test: 24.291, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.0002559547922699536 , The MSE train: 21.641, test: 24.290, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.00030888435964774815 , The MSE train: 21.641, test: 24.290, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.0003727593720314938 , The MSE train: 21.641, test: 24.290, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.0004498432668969444 , The MSE train: 21.641, test: 24.290, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.0005428675439323859 , The MSE train: 21.642, test: 24.290, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.0006551285568595509 , The MSE train: 21.642, test: 24.290, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.0007906043210907702 , The MSE train: 21.642, test: 24.289, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.0009540954763499944 , The MSE train: 21.642, test: 24.289, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.0011513953993264468 , The MSE train: 21.642, test: 24.289, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.0013894954943731374 , The MSE train: 21.642, test: 24.288, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.0016768329368110067 , The MSE train: 21.642, test: 24.288, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.0020235896477251557 , The MSE train: 21.643, test: 24.287, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.0024420530945486497 , The MSE train: 21.644, test: 24.287, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.0029470517025518097 , The MSE train: 21.645, test: 24.287, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.0035564803062231283 , The MSE train: 21.646, test: 24.286, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.004291934260128779 , The MSE train: 21.648, test: 24.287, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.005179474679231213 , The MSE train: 21.651, test: 24.288, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.0062505519252739694 , The MSE train: 21.656, test: 24.290, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.007543120063354615 , The MSE train: 21.662, test: 24.293, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.009102981779915217 , The MSE train: 21.672, test: 24.299, and R<sup>2</sup> train: 0.751, test: 0.669

When alpha = 0.010985411419875584 , The MSE train: 21.686, test: 24.309, and R<sup>2</sup> train: 0.750, test: 0.669

When alpha = 0.013257113655901081 , The MSE train: 21.706, test: 24.324, and R<sup>2</sup> train: 0.750, test: 0.668

When alpha = 0.015998587196060572 , The MSE train: 21.736, test: 24.349, and R<sup>2</sup> train: 0.750, test: 0.668

When alpha = 0.019306977288832496 , The MSE train: 21.779, test: 24.386,  
and R^2 train: 0.749, test: 0.667

When alpha = 0.023299518105153717 , The MSE train: 21.841, test: 24.442,  
and R^2 train: 0.749, test: 0.667

When alpha = 0.02811768697974228 , The MSE train: 21.932, test: 24.527, a  
nd R^2 train: 0.748, test: 0.666

When alpha = 0.033932217718953266 , The MSE train: 22.059, test: 24.663,  
and R^2 train: 0.746, test: 0.664

When alpha = 0.040949150623804234 , The MSE train: 22.249, test: 24.852,  
and R^2 train: 0.744, test: 0.661

When alpha = 0.04941713361323833 , The MSE train: 22.526, test: 25.130, a  
nd R^2 train: 0.741, test: 0.657

When alpha = 0.05963623316594643 , The MSE train: 22.622, test: 25.196, a  
nd R^2 train: 0.740, test: 0.656

When alpha = 0.07196856730011514 , The MSE train: 22.651, test: 25.179, a  
nd R^2 train: 0.739, test: 0.657

When alpha = 0.08685113737513521 , The MSE train: 22.695, test: 25.163, a  
nd R^2 train: 0.739, test: 0.657

When alpha = 0.10481131341546852 , The MSE train: 22.758, test: 25.154, a  
nd R^2 train: 0.738, test: 0.657

When alpha = 0.12648552168552957 , The MSE train: 22.850, test: 25.156, a  
nd R^2 train: 0.737, test: 0.657

When alpha = 0.15264179671752318 , The MSE train: 22.984, test: 25.178, a  
nd R^2 train: 0.735, test: 0.657

When alpha = 0.18420699693267145 , The MSE train: 23.053, test: 25.097, a  
nd R^2 train: 0.735, test: 0.658

When alpha = 0.22229964825261933 , The MSE train: 23.110, test: 24.967, a  
nd R^2 train: 0.734, test: 0.660

When alpha = 0.2682695795279725 , The MSE train: 23.192, test: 24.823, an  
d R^2 train: 0.733, test: 0.662

When alpha = 0.32374575428176433 , The MSE train: 23.312, test: 24.669, a  
nd R^2 train: 0.732, test: 0.664

When alpha = 0.3906939937054613 , The MSE train: 23.487, test: 24.509, an  
d R^2 train: 0.730, test: 0.666

When alpha = 0.47148663634573895 , The MSE train: 23.729, test: 24.357, a  
nd R^2 train: 0.727, test: 0.668

When alpha = 0.5689866029018293 , The MSE train: 24.060, test: 24.233, an  
d R^2 train: 0.723, test: 0.670

When alpha = 0.6866488450042998 , The MSE train: 24.575, test: 24.157, an  
d R^2 train: 0.717, test: 0.671

When alpha = 0.8286427728546842 , The MSE train: 25.325, test: 24.190, an  
d R^2 train: 0.708, test: 0.670

When alpha = 1.0 , The MSE train: 26.417, test: 24.409, and R^2 train: 0.  
696, test: 0.667

Let's mainly focus on the testing set's MSE and  $R^2$ .

For MSE, as alpha increase, the MSE first decreases until alpha = 0.0035564803062231283, whose MSE is 24.286. Then the MSE increases until alpha = 0.15264179671752318, then keeps decreasing. If we choose MSE as our performance metric, the best alpha will be 0.0035564803062231283 for the testing set, whose MSE 24.286 is the minimum.

For  $R^2$ , as alpha increase,  $R^2$  first keeps the same of 0.669, then decreases until the minimum of 0.656 when alpha = 0.05963623316594643, then starts increasing to its maximum, 0.671 when alpha = 0.6866488450042998. Thus, the best alpha will be 0.6866488450042998 for the testing set since its  $R^2$  reaches the maximum, if choosing  $R^2$  as our performance metric.

## Part 4: Conclusions

In this report, I first conducted EDA on the housing.csv dataset and found out there existed linear relationship between some features and the target variable, MEDV. I split data into 80% training and 20% test sets. Then, I run linear regression models on all the features and target variable. I found there were overfitting problem. To test the performance of the fitted model, I plotted the residual plot of this model, and computed the performance metrics, MSE and  $R^2$ . After that, to prevent overfitting problems, I used Ridge regression (L2 Regularization) and LASSO regression (L1 Regularization) to fit our training data, then test the model in our testing data by residual plot, MSE, and  $R^2$ . Additionally, in each of the regularization regression model, I used two different methods to find the best alpha. The first one is using 10-cross-validation to compute the averaged CV scores, and the second one is testing different alpha on ridge/LASSO models to find which alpha gives the lowest MSE and highest  $R^2$ .

In conclusion, if choosing MSE as our performance metrics, the linear regression model gives the smallest MSE, 21.641. If choosing  $R^2$  as our performance metrics, the LASSO regression gives the largest  $R^2$ , 0.671 when alpha is around 0.6866.

Also, when comparing the averaged 10-fold-cv-scores to choose the best alpha, the best alpha for both ridge and LASSO regressions are very small. Thus, they would be very similar to linear regression model, then the performance metrics would be very similar as well.

## Part 5: Appendix

Link to github repo:

[https://github.com/yaxuanw3/IE517\\_F20\\_HW4](https://github.com/yaxuanw3/IE517_F20_HW4) ([https://github.com/yaxuanw3/IE517\\_F20\\_HW4](https://github.com/yaxuanw3/IE517_F20_HW4))

My name is {Yaxuan Wang}

My NetID is: {662869931}

I hereby certify that I have read the University policy on Academic Integrity and that I am not in violation.