In [6]:
```
# -*- coding: utf-8 -*-
"""
Created on Sun Oct  7 15:47:50 2018

@author: amitp
"""

import numpy as np
import pandas as pd
import scipy.stats as stats
import matplotlib.pyplot as plt
import sklearn
from sklearn.datasets import load_boston
from sklearn.metrics import r2_score

boston = load_boston()
bos = pd.DataFrame(boston.data)

X = bos.iloc[:, [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 11]].values
Y = bos.iloc[:, 12].values

from sklearn.cross_validation import train_test_split
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size = 0.1, random_state = 0)

from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train, Y_train)

Y_pred = regressor.predict(X_test)
print("Rsquare %f.10f\n"%(r2_score(Y_test,Y_pred)))

df2 = pd.DataFrame(np.random.randint(low = 0, high = 10, size=(51,2)), columns = ["Test", "Predicted"])
df2 ["Test"] = Y_test
df2["Predicted"] = Y_pred
print(df2)

X = bos.iloc[:, :-1].values
Y = bos.iloc[:, 12].values


# Building the optimal model using Backward Elimination
import statsmodels.formula.api as sm
X = np.append(arr = np.ones((506, 1)).astype(int), values = X, axis = 1)
#X_opt = X[:, [0, 1, 2, 3, 4, 5]]
X_opt = X[:, :-1]
```

```
regressor_OLS = sm.OLS(endog = Y, exog = X_opt).fit()
print(regressor_OLS.summary())

#X_opt = X[:, [0, 1, 2, 3, 4, 5]]
X_opt = X[:, [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 11]]
regressor_OLS = sm.OLS(endog = Y, exog = X_opt).fit()
print(regressor_OLS.summary())
```

```
Rsquare 0.472757.10f

     Test  Predicted
0    7.34   5.959792
1    9.53  17.564101
2   10.50  13.449262
3   19.77  22.878492
4   12.34  10.563367
5    8.47  12.495776
6   11.45  14.111716
7    9.29  11.820954
8   12.64  15.424514
9   10.63  15.498500
10  20.62  29.524109
11  21.22  19.341722
12  17.79  17.263366
13  28.28  26.457282
14   1.92   7.920136
15   4.74   3.029344
16  10.74  13.489045
17   4.08   3.684551
18   5.33   6.436183
19   8.77  10.668035
20   7.20   9.154575
21  12.03  17.664826
22  14.09  13.462900
23   7.19   7.192765
24  10.21  10.760460
25  13.33  31.480650
26  13.83  14.887065
27  15.79  20.151852
28   4.21   2.779322
29  17.09  14.759709
30  18.03  17.447196
31  14.70  16.623954
32  10.15  13.219908
33  12.40  12.096116
34   9.04  11.633471
35  14.64  18.278034
36  26.77  20.272443
37   8.88  18.080766
38  17.64  18.439250
39  24.10  17.902636
40   8.58   7.918184
41  14.15  10.229469
```

```
42   9.09   8.371196
43  18.14  20.857925
44   8.65   8.754513
45   9.14  11.427659
46  15.84  13.246092
47  15.12  16.976412
48  29.29  22.516690
49   9.42  10.177163
50  15.02  18.243472
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.651
Model:                            OLS   Adj. R-squared:                  0.643
Method:                 Least Squares   F-statistic:                     83.78
Date:                Sat, 13 Oct 2018   Prob (F-statistic):          2.00e-105
Time:                        18:12:17   Log-Likelihood:                -1445.9
No. Observations:                 506   AIC:                             2916.
Df Residuals:                     494   BIC:                             2966.
Df Model:                          11
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         26.1867      4.227      6.194      0.000      17.881      34.493
x1             0.1099      0.029      3.815      0.000       0.053       0.166
x2             0.0144      0.012      1.169      0.243      -0.010       0.039
x3             0.1008      0.055      1.831      0.068      -0.007       0.209
x4            -1.1829      0.772     -1.533      0.126      -2.699       0.333
x5             6.2343      3.413      1.827      0.068      -0.471      12.939
x6            -4.3814      0.317    -13.806      0.000      -5.005      -3.758
x7             0.0885      0.011      7.927      0.000       0.067       0.110
x8             0.1671      0.179      0.933      0.351      -0.185       0.519
x9             0.0652      0.059      1.098      0.273      -0.051       0.182
x10           -0.0009      0.003     -0.254      0.800      -0.007       0.006
x11            0.1068      0.117      0.909      0.364      -0.124       0.337
==============================================================================
Omnibus:                       49.986   Durbin-Watson:                   1.185
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              295.703
Skew:                           0.016   Prob(JB):                     6.15e-65
Kurtosis:                       6.745   Cond. No.                     1.16e+04
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.16e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.651
Model:                            OLS   Adj. R-squared:                  0.644
Method:                 Least Squares   F-statistic:                     92.33
Date:                Sat, 13 Oct 2018   Prob (F-statistic):          2.09e-106
Time:                        18:12:17   Log-Likelihood:                -1445.9
No. Observations:                 506   AIC:                             2914.
Df Residuals:                     495   BIC:                             2960.
Df Model:                          10
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         26.0718      4.199      6.209      0.000      17.821      34.322
x1             0.1100      0.029      3.821      0.000       0.053       0.166
x2             0.0137      0.012      1.142      0.254      -0.010       0.037
x3             0.0947      0.049      1.917      0.056      -0.002       0.192
x4            -1.1590      0.765     -1.515      0.130      -2.662       0.344
x5             6.1684      3.400      1.814      0.070      -0.511      12.848
x6            -4.3751      0.316    -13.841      0.000      -4.996      -3.754
x7             0.0884      0.011      7.930      0.000       0.066       0.110
x8             0.1662      0.179      0.929      0.353      -0.185       0.518
x9             0.0533      0.036      1.474      0.141      -0.018       0.124
x10            0.1046      0.117      0.894      0.372      -0.125       0.334
==============================================================================
Omnibus:                       50.069   Durbin-Watson:                   1.185
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              296.678
Skew:                           0.020   Prob(JB):                     3.78e-65
Kurtosis:                       6.751   Cond. No.                     2.02e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.02e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```