**Muhammad Aamir Iqbal**
**Project 4 Report** **October 27, 2018**

In project 4 I used following methods for regression:

1. LinearRegression
2. RANSACRegressor
3. Ridge
4. Lasso
5. Non-Linear Regression using Polynomial
6. Normal Equation

I used Sklearn library for first four methods and used class notes for non-linear regression and normal equation implementation. I used default parameters in most of these methods.
I used following datasets in this project:

1. Housing
2. California-Wind-Power-Generation-Forecasting

I used both datasets for testing fist five methods. I tested normal equation method using housing dataset only. I found missing values in California-Wind-Power-Generation-Forecasting dataset. I tested this dataset using two different approaches:

1. Deleting columns with missing values
2. Replacing blank cell with mean value of these columns

In first approach I deleted following columns:

1. Solar
2. Solar PV
3. Solar Thermal

In second approach I inserted mean value in blank cells in following columns:

1. Solar → Mean value = 163
2. Solar PV→ Mean value = 1647
3. Solar Thermal→ Mean value = 117

**Project 4 Results**

| Housing Dataset | | | | | |
|---|---|---|---|---|---|
| | Scaled | | | | |
| | **MSE Train** | **MSE Test** | **R^2 Train** | **R^2 Test** | **Running Time** |
| LinearRegression | 28.008 | 26.207 | 0.660 | 0.708 | 0.0071871 |
| RANSACRegressor | 87.713 | 74.574 | -0.066 | 0.170 | 0.0467463 |
| Ridge | 37.082 | 42.188 | 0.549 | 0.530 | 0.0155115 |
| Lasso | 82.256 | 90.046 | 0.000 | -0.002 | 0.0154929 |
| Non-Linear Regression | 23.784 | | 0.719 | | 0.0180025 |
| | **Slope** | | **Intercept** | | **Running Time** |
| Normal Equation | -0.159738 | | 20.710393 | | 0.0155168 |

**Muhammad Aamir Iqbal**
**Project 4 Report** **October 27, 2018**

| Solar Dataset with Deleted Columns Scaled | | | | | |
|---|---|---|---|---|---|
| | MSE Train | MSE Test | R^2 Train | R^2 Test | Running Time |
| **LinearRegression** | 997392.459 | 1004483.758 | 0.072 | 0.071 | 0.2995639 |
| **RANSACRegressor** | 1204881.100 | 1216193.751 | -0.121 | -0.125 | 0.427 |
| **Ridge** | 1016450.938 | 1022873.190 | 0.054 | 0.054 | 0.299 |
| **Lasso** | 1074995.993 | 1081128.227 | 0.000 | 0.000 | 0.297 |
| **Non-Linear Regression** | 944925.868 | | 0.122 | | 0.3591526 |

| Solar Dataset with Mean Values Scaled | | | | | |
|---|---|---|---|---|---|
| | MSE Train | MSE Test | R^2 Train | R^2 Test | Running Time |
| **LinearRegression** | 957146.833 | 964682.807 | 0.110 | 0.108 | 0.4216356 |
| **RANSACRegressor** | 1372272.921 | 1376302.895 | -0.277 | -0.273 | 0.5732081 |
| **Ridge** | 991127.732 | 997915.895 | 0.078 | 0.077 | 0.4217758 |
| **Lasso** | 1074995.993 | 1081128.227 | 0.000 | 0.000 | 0.422 |
| **Non-Linear Regression** | 884513.031 | | 0.179 | | 0.5905256 |