

# Project 4 Report

---

## Results

---

```
linreg datasets/housing.data.txt
Training Size (instances): 404
Testing Size (instances): 102
Train Time (s): 0.0010534149478189647
Test Time (s): 0.0006580009940080345
Data slope: -0.070, intercept: 34.319
MSE train: 42.585, test: 22.278
R^2 train: 0.502, test: 0.171
```

```
ransac datasets/housing.data.txt
Training Size (instances): 404
Testing Size (instances): 102
Train Time (s): 0.011309213994536549
Test Time (s): 0.0007798140286467969
Data slope: -0.070, intercept: 34.319
MSE train: 73.027, test: 118.721
R^2 train: 0.146, test: -3.415
```

```
ridge datasets/housing.data.txt
Training Size (instances): 404
Testing Size (instances): 102
Train Time (s): 0.001593180000782013
Test Time (s): 0.0010018310276791453
Data slope: -0.070, intercept: 34.319
MSE train: 42.585, test: 22.277
R^2 train: 0.502, test: 0.171
```

```
lasso datasets/housing.data.txt
Training Size (instances): 404
Testing Size (instances): 102
Train Time (s): 0.0009100159513764083
Test Time (s): 0.0006537479930557311
Data slope: -0.070, intercept: 34.319
MSE train: 42.619, test: 22.998
R^2 train: 0.502, test: 0.145
```

```
nonlinear datasets/housing.data.txt
Training Size (instances): 404
Testing Size (instances): 102
Train Time (s): 0.0007939990027807653
Test Time (s): 0.0006739479722455144
MSE train: 22.694, test: 28.444
R^2 train: 0.735, test: -0.058
```

-----

```
linreg datasets/all_breakdown.csv
Training Size (instances): 54067
Testing Size (instances): 13517
Train Time (s): 0.003546730033122003
Test Time (s): 0.007904438010882586
Data slope: 6.701, intercept: 141.999
MSE train: 1058246.451, test: 1012001.116
```

```
R^2 train: 0.025, test: 0.028
```

```
ransac datasets/all_breakdown.csv
Training Size (instances): 54067
Testing Size (instances): 13517
Train Time (s): 0.15652674197917804
Test Time (s): 0.011446275981143117
Data slope: 6.701, intercept: 141.999
MSE train: 1237500.466, test: 1152224.602
R^2 train: -0.140, test: -0.107
```

```
ridge datasets/all_breakdown.csv
Training Size (instances): 54067
Testing Size (instances): 13517
Train Time (s): 0.001984770002309233
Test Time (s): 0.0016921749920584261
Data slope: 6.701, intercept: 141.999
MSE train: 1058246.451, test: 1012001.116
R^2 train: 0.025, test: 0.028
```

```
lasso datasets/all_breakdown.csv
Training Size (instances): 54067
Testing Size (instances): 13517
Train Time (s): 0.0030993890250101686
Test Time (s): 0.0021569780074059963
Data slope: 6.701, intercept: 141.999
MSE train: 1058246.453, test: 1012000.774
R^2 train: 0.025, test: 0.028
```

```
nonlinear datasets/all_breakdown.csv
Training Size (instances): 54067
Testing Size (instances): 13517
Train Time (s): 0.012229336949530989
Test Time (s): 0.0023268309887498617
MSE train: 984166.807, test: 958451.689
R^2 train: 0.093, test: 0.079
```

## Analysis

The Mean Squared Error (MSE) value tells us how much error there is in the prediction based on the data, and seems high for most of the runs. It tells us how off of the line the data fits. We can compare our test and train MSE to see if we are over fitting or under fitting based on the difference between the two. When test is higher, we may be over fitting.

The  $R^2$  data is easier to read, since we know 1.0 means a better fit. For the housing data, the  $R^2$  values are closer to 0, which means we had better predictions there. For the other dataset, the number is closer to 0 for most of them, meaning it was not as good of fit (we want 1).