Ryan Adams
Project4

Project 4

Findings:

Throughout this assignment we use several regressors, the normal equation, and a neural network regression (nonlinear regression) to ascertain the relationship between two attributes X, and Y.

We note that regression has strong dependency on the values that the regressions are being performed upon. As such when we evaluate how "well" a regression is at determining a relationship we look to see how close the R2 score is to 1 and how close the MSE is to zero. That is to say that if we saw and R2 score of 1 with and MSE of zero, these two columns would be perfectly correlated. And we would in fact expect to see that if a column is correlated against itself.

Based on my running I created the following two tables to demonstrate results from the housing data set and California Renewable Production 2010-2018 dataset.

Housing data sample : Column 1 vs. Column 2

Regressor	MSE	R2
Linear	33.5858624533	0.284972532493
RANSCAR	57.5160866757	0.224490865878
Ridge	33.5858624534	0.284972532489
Lasso	33.5877045426	0.284933315264
Normal Equation	168.459731804	na
Neural MLP Regressor	29.4333325029	0.346965030627

We can see here that the MSE is fairly tightly bounder with a range from 29 to 168. And R2 from .22 to .34. From this we can say that there is some correlation between columns 1 and 2.

California Renewable Production 2010-2018 Column 1 vs Column 7

Regressor	MSE	R2
Linear	5626084.68323	0.0158402890687
RANSCAR	8429877.43507	0.474621554206
Ridge	5626084.68323	0.0158402886291
Lasso	5626084.68574	0.284933315264
Normal Equation	5851830.27404	na
Neural MLP Regressor	5648414.24313	0.0203246880445

Based on these results, we see MSE values being very large, and R2's close to 0. We can conclude that this data is not very correlated.