**ML ASSIGNMENT 2**

**SRI SAI MANISH KONDAPALLI 801020355**

* I have used scikit learn sample dataset to load the data.
* Created training and testing data sets using train\_test\_split with a test size of 0.3(30%) i.e used 45 records for testing and rest for training. I have used this particular training and testing set because this combination gave the highest R Squared value and minimum Mean Absolute Error when compared to other test set sizes of 0.33, 0.5, etc.
* Used PolynomialFeatures with a degree 3 in order to fit\_transform the training and testing set. Chose this degree 3 because other degrees either gave a lower R squared value, score or a high Mean Absolute Error.
* Now this polynomial data is fit onto the Ridge Regression Model ( Regularized Linear Regression). I have chosen this over elastic and lasso regression models because of the low scores obtained from these models when compared to Ridge Regression Model.
* And then I calculated R2 square and Mean Absolute Error to check the efficiency of model.
* Then scatter plot is plotted Actual vs Predicted.
* The above combination of parameter values and model selection gave the best score when compared to others thus making me want to prefer this model.
* **Files Required:**

ml\_assignment2.ipynb

* **Libraries Required:**

matplotlib

sklearn

* **Tool Used:**

Jupyter Notebook

* **References:**

http://scikit-learn.org/stable/auto\_examples/linear\_model/plot\_ols.html

http://www.ritchieng.com/machine-learning-polynomial-regression/

http://scikit-learn.org/stable/modules/generated/sklearn.linear\_model.Ridge.html