Draft Project Summary

Hypothesis: Can we use Machine Learning to predict daily Air Quality Index (AQI) in Berkeley, CA? The variables we will test are:

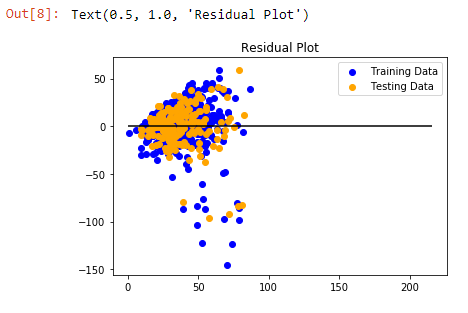
1. Wildfire data: Daily Acres Burned by wildfires in the Northern CA region;
2. Daily Precipitation;
3. Daily Average Humidity;
4. Daily Average Wind Speed;
5. Daily Average Wind Direction;
6. Daily Average Temperature;

The dataset includes daily measurements of the variables above over the period from 01/01/2017 through 12/31/2018 (2 years – 730 rows).

**Model Test Summaries**

**Basic Multivariate Linear Regression**

* LinearRegression (copy\_X=True, fit\_intercept=True, n\_jobs=None, normalize=False)
* Training Score: 0.28012777253831334
* Testing Score: 0.36348094999959646
* R2 Score: 0.30255



* **Lasso Model** (least absolute shrinkage and selection operator) - performs both variable selection and regularization in order to enhance prediction accuracy;
  + MSE: 0.6641789181804544
  + R2: 0.36264406893552614
* **Ridge Model** - the most commonly used method of regularization of ill-posed problems. if no “x” satisfies the equation or more than one “x” does—that is, the solution is not unique—the problem is said to be ill posed.
  + MSE: 0.6633075210405341
  + R2: 0.36348027454253684
* **Elasticnet Model** - a regularized regression method that linearly combines the L1 and L2 penalties of the lasso and ridge methods.
  + MSE: 0.663868935823033
  + R2: 0.3629415325987346