**Fall 2018: EEL - 6935 Smart Grid – Homework 0 2**

Hector Lopez

3. Sustainability on the Brain

Google trend for the search term “sustainability” may have a correlation with the power demand from the data set. We do know that there is a correlation between temperature and power demand. Can the Google trend be a better predictor of power demand?

Use the data **hw2\_ q 3 \_ data.txt** for this problem; Each line of this file contains average values of the Google trend, temperature, and demand for one week:

(a)Plot the trend data versus average demand for all the data points:

(b) Using the trend data alone (not the temperature data), use linear regression to predict electrical demand given the trend data. You can use just linear regression here (adding a constant - valued feature to the input). Evaluate the performance of this model using 70% of the data as the training set and 30% of the data as a test set. Report mean squared error along with 95% confidence intervals on the test set. Also report mean squared error and 95% confidence intervals for a baseline prediction that just predicts the mean demand from the test set (this is essentially linear regression with just the constant term).

=== Run information ===

Scheme: weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4

Relation: googtrend

Instances: 207

Attributes: 3

googletrend

avgtemp

avgdemand

Test mode: split 70.0% train, remainder test

=== Classifier model (full training set) ===

Linear Regression Model

avgdemand =

-0.4365 \* googletrend +

0.0012 \* avgtemp +

2.3742

=== Summary ===

Correlation coefficient 0.6481

Mean absolute error 0.1192

Root mean squared error 0.1477

Relative absolute error 76.5901 %

Root relative squared error 74.8646 %

Total Number of Instances 62

(c) Use both the trend and temperature data to predict demand and compare this to using just temperat ure to predict demand, using the same evaluation methodology as above (i.e., for each method, report mean squared error and 95% con fidence intervals on a test set). For the temperature, you can use quadratic features (temperature and temperature squared).

(d) What can you reasonably conclude from this st udy? Note that for the purposes of this problem you can conclude that predictor A is “significantly” better than predictor B if the mean squared error of predictor B lies above the 95% con fidence interval of predictor A (i.e., it is enough for the mea n of predictor B to lie outside A ’ s error bars; the error bars do not have to be completely separate).