Nick Handelman (nrh8883)

April 30, 2012

CMPS 509

Program 2: Linear Regression

**Introduction and Description of Program**

My program implements simple linear regression using the ordinary least squares (OLS) regression model. As such, it operates on a scalar independent variable and outputs a scalar dependent variable. Its goal is to fit a straight line through the observed input data points in such a way as to minimize the sum of the squared residuals, that is, the vertical distances between the data points and the fitted straight line.

First, the program makes one pass through the data file to determine the number of samples. Next, it reads in the data from the file which consists of two columns: the first holds the independent variable (x) and the second holds the dependent variable (y). In the file, the two data values on each row are separated by whitespace. They are stored as doubles in the program.

The fitted straight line in simple OLS takes the form of:

where and α equal:

with n being the number of input data samples.

In calculating the simple linear regression equation, the program first calculates:

It then applies these pre-calculated summations as necessary in calculating equations β and α.

**Conclusions**

While a linear regression equation can be calculated for almost any data set, it may not offer the best representation of the data. In the case of a data set wherein the input (x) values are all equal, the equation can't be calculated in the first place, due to a division by zero. In the case of a data set that demonstrates a parabolic trend, a non-linear regression equation would be a better choice. There are methods for determining a "goodness of fit", but I did not implement any of these in my program.

The simple OLS regression model does have two advantages. First, it is very easy to calculate a regression equation for this model, as described previously. Second, due to its simplicity, this regression model offers a good introduction to the subject of linear regression.