Yahya Alhinai

EE5373

November 12th, 2019

Multi-factor Regression Models

OBJECTIVE FOR THIS LAB:

- This lab explores in greater depth the training and testing process for multi-factor regression models.

1. Explanation of how the graphs were developed and what they show:

The graph produced in this report are meant to show the effect of various values of f-value to the trained data in comparison to the actual data. F-value is representing the percentage of the trained data to produce a multi-factor regression model to test it on the rest of data that has not been included to produce the regression model. Starting with training 10% of data values to produce a regression model up to 90% of data values. Then analyzed the data produce by variant values of f-value and see if there is a trend or a correlation between the produced mean and its confident interval and the set f-value for that produced data.

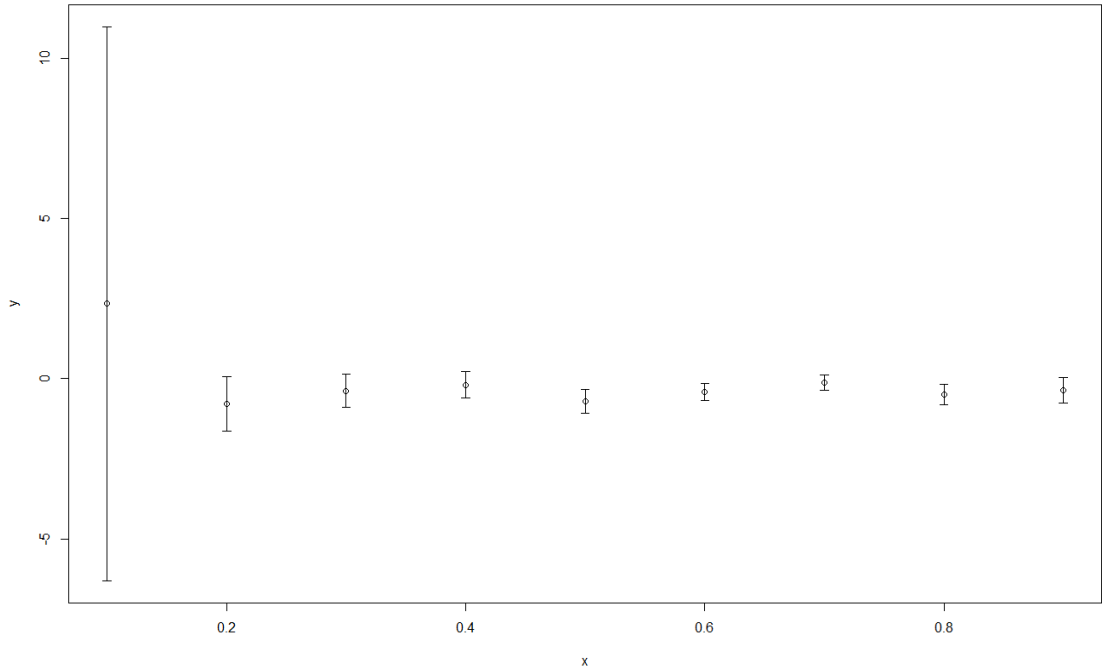
The graphs were developed by testing the predictive capability of a specific model trained on portion of the data set (f-value). This procedure of random sampling a portion of data and then comparing the predicted values with the actual values to see how good our predictive model. This produce a delta vector of the difference (actual values - predicted values) for each element. The closer to zero means the model explain the actual values better. In this lab we conducted k times, in our case k=100, run of different samples producing different predictive model each time which means different delta vector. The we concatenated all the produced delta vectors together to have a increase the precision of mean the confident travel produced because the more data we feed into t-test the higher precision it produced.

The graphs in are RStudio were created by having two loops each other. The inner loop to collect concatenate delta vectors k times each vector produced by different sampling model. The outer loop is meant to vary the value of f from 0.1 to 0.9 for each k sets of delta vectors. After running a t-test for all f values we get the following:

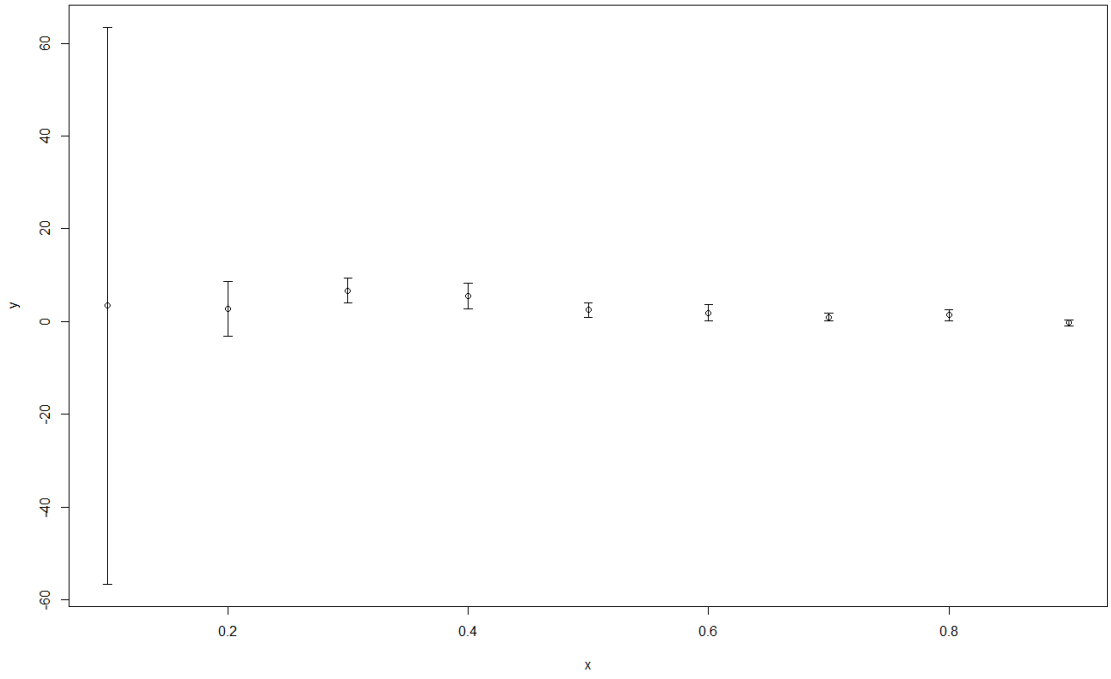
- Mean
- Low-bound
- High-bound

2. All of the graphs are included:

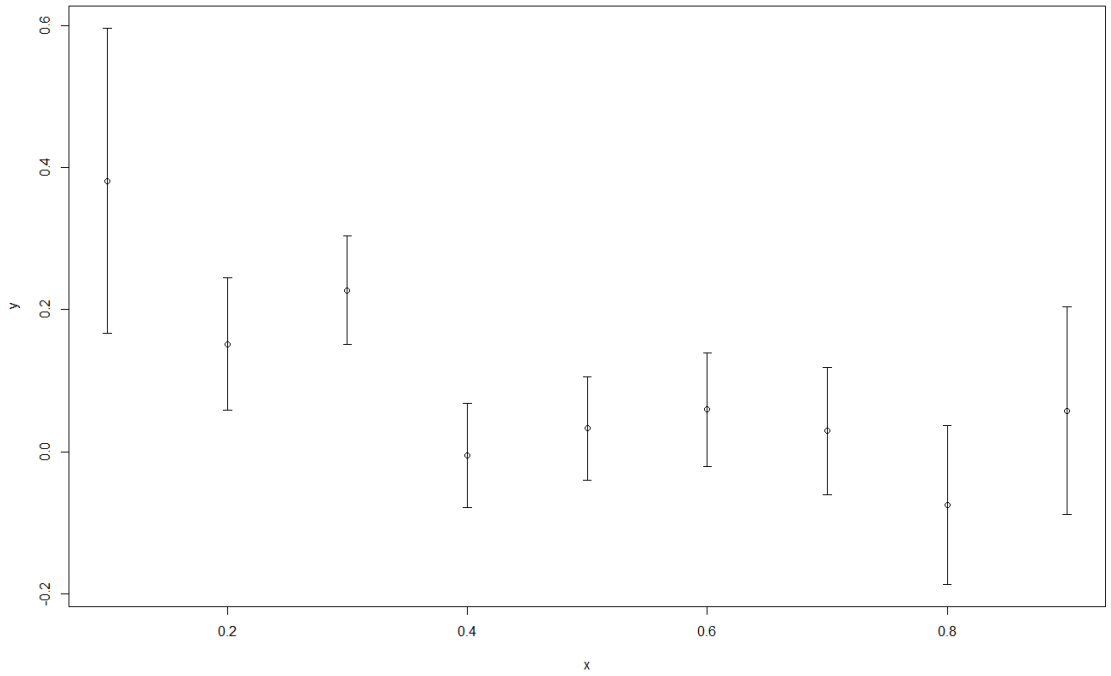
INT1995:



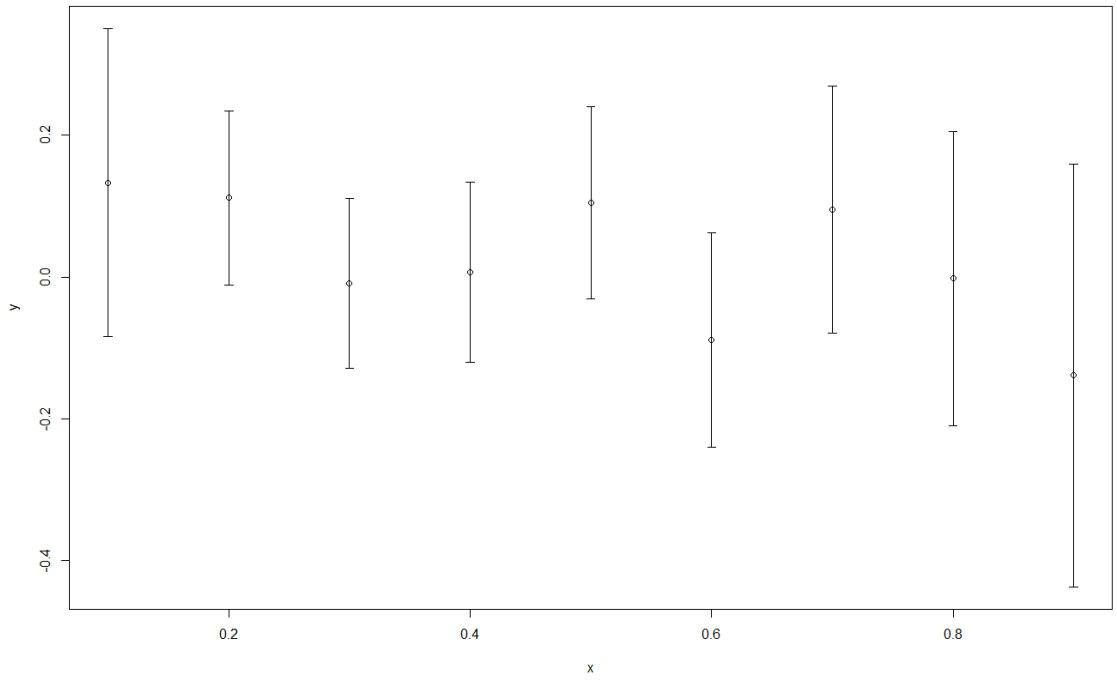
FP1995:



INT2006:



FP2006:



3. Clear explanation of the effects of varying f

The most profound observation that is seen in all four graphs is the offset of mean and the range width of the confident interval at f = 10% for 1995's processes. The processor INT and FP of the year 1995 tend to have similar trend. Both graphs started with a wide range of the confidant interval at value of f = 0.1 and the mean point is shifted from the value zero. Indicating the model that is train with 10% of the data is not perfectly describing the data set.

As the value of f increases, the mean tends to settle at zero which means the predictive model is describing data set well. As well as, the range shrink down considerably which indicates high precision of describing the data set. In conclusion, the higher f-value leads to have more data to be trained and therefore produce a better model with tight confidence interval and a mean around zero.

4. Other interesting observations

Ideally the higher f-value is the better the model produced. This is generally true but, it has been observed in this lab that after a certain f-value threshold, the quality of the model reaches a saturation state. After reaching around 0.4 to 0.5, the quality of the model stays the same with some margin for systematic error.

Also, it has been noticed that the INT model produced gives better results than the one modeled for FP. This is assumed to be because the lack of data for FP model that is fed into this process which make the model less accurate in depicting the actual data.

Finally, there is an exponential improvement in the regression model produced for the processors in between the year 1995 and 2006. As time goes, a better processor architecture is being used which indicates the processer's dependency factors increases. As the processor's dependency factors increases, a better prediction is made when train data which lead to a better model.