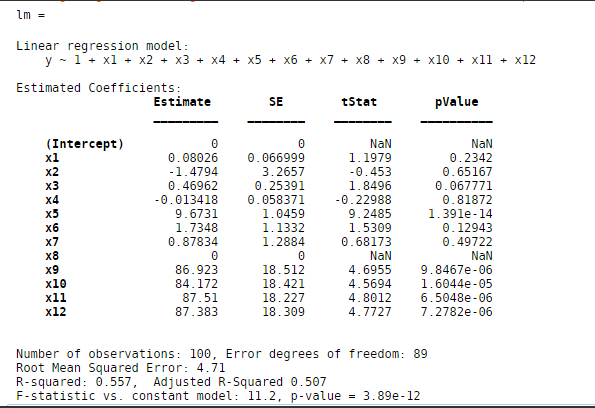
Robert Krall

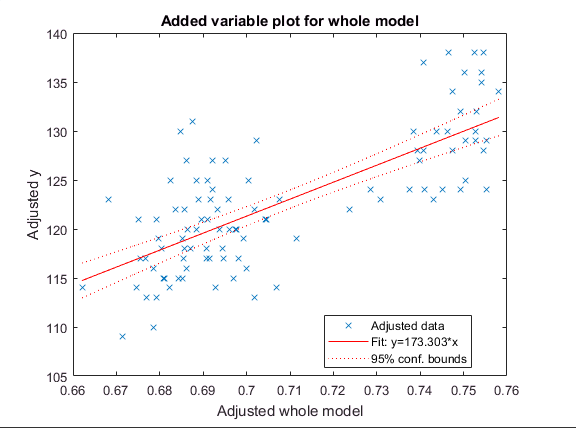
Assignment 2

SEIS 763 – Machine learning

4. What are the regression coefficients (thetas)?

See below for the theta values.



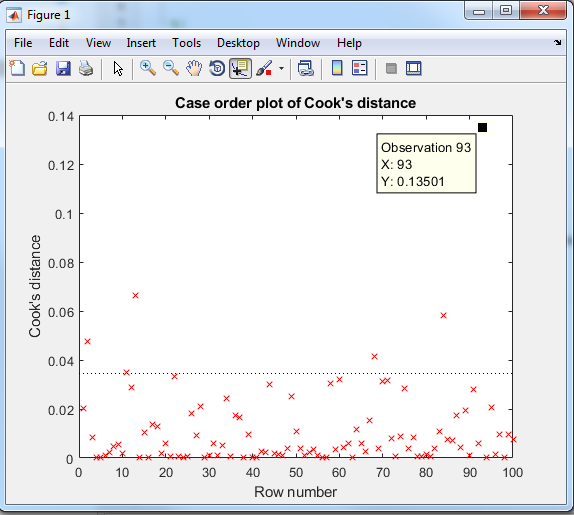


5. How do you interpret those numbers?

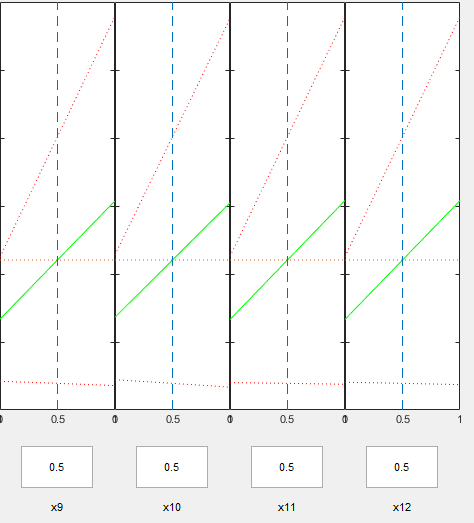
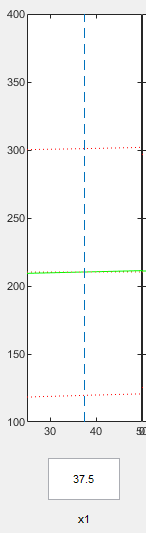
The observations with high coefficients and low pValue would suggest they are not as valuable as the other observations.

6. If you need to identify one outlier record, which record is a potential outlier? How do you reach this conclusion?

The outlier in this dataset is observation 93. This is done by using a cook’s distance plot using Matlab. We can quickly see the outlier in the graph below.



7. If you need to identify one or few useless features (independent variables or predictors), which one(s) will you choose? Why do you reach this conclusion?



To identify useless features I would use a combination of 2 things. I would want to dive deep in those observations with a high coefficient but a low pValue. To do this I would create a plot slice map as seen below.

The graph on the left shows that x9, x10, x11, x12 observations have a significant impact on y. If we look at observation x1 or age (graph on the right) it would have no signification impact on y.