CSL603 - Machine Learning
Lab 2

Aditya Gupta
2015CSB1003

September 9, 2017

Linear Ridge Regression

Given X and Y we will find W that minimizes $J(W)$, the error function and are defined as:

$$f(X) = \underbrace{\begin{pmatrix} 1 & x_{11} & \cdots & x_{1D} \\ 1 & x_{21} & \cdots & x_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{ND} \end{pmatrix}}_{X} \underbrace{\begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_D \end{pmatrix}}_{W} = \underbrace{\begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_N \end{pmatrix}}_{Y}$$

$$\min_W J(W) \equiv \min_W \frac{1}{2}(XW - Y)^T(XW - Y) + \lambda||W||^2$$

Which when solved gives us:

$$W = (X^t X + \lambda I)^{-1} X^T Y$$

Observations

The following observations were obtained:

- A particular value of λ(say 0) was choosen and then the magnitude of entries in the weights W was compared and one by one the least significant ones were discarded and the mean squared error changed in the following way:
- The effect of λ on error was observed for different partitions of the data into training and testing sets. The average mean absolute error for 100 repetitions for splitting-fractions varying from 1% to 99% and lambda values from 0 to 100 was observed. The surface corresponding to the average mean absolute error observed was:

- The above surface can be plotted into different graphs for few particular values of splitting-fractions and varying lambda and obsreving the change in average mean absolute error. These figures were observed:
- Now we noted the minimum average mean squared testing error for each training set fraction values. Also the corresponding λ value was observed. These figures were observed:
- The correspondence between the actual and predicted values was also observed. For perfect prediction, this should correspond to a straight line through origin at 45° degrees. The following figures were obtained: