CSL 467: Homework #1

Due on Friday, August 15, 2014

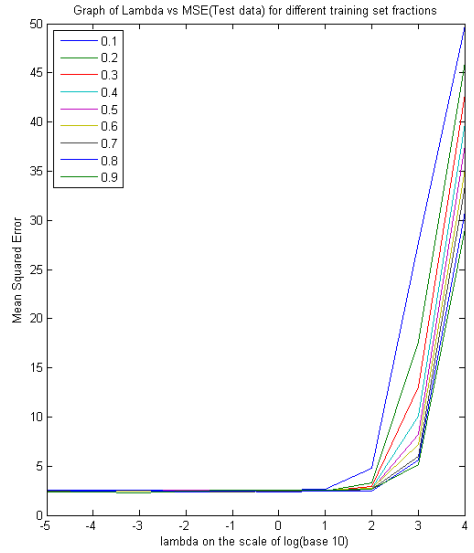
Narayanan C Krishnan

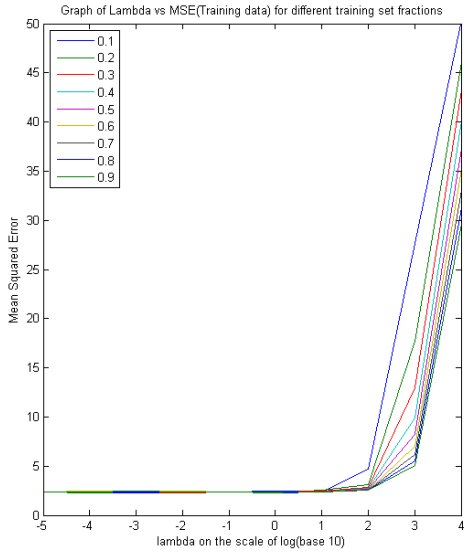
Harsimran Singh

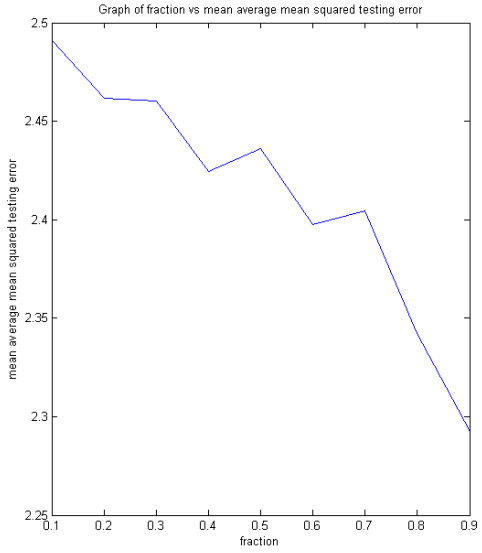
Contents

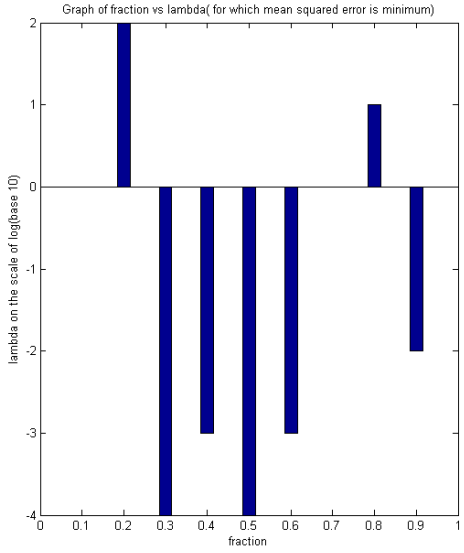
Problem 1

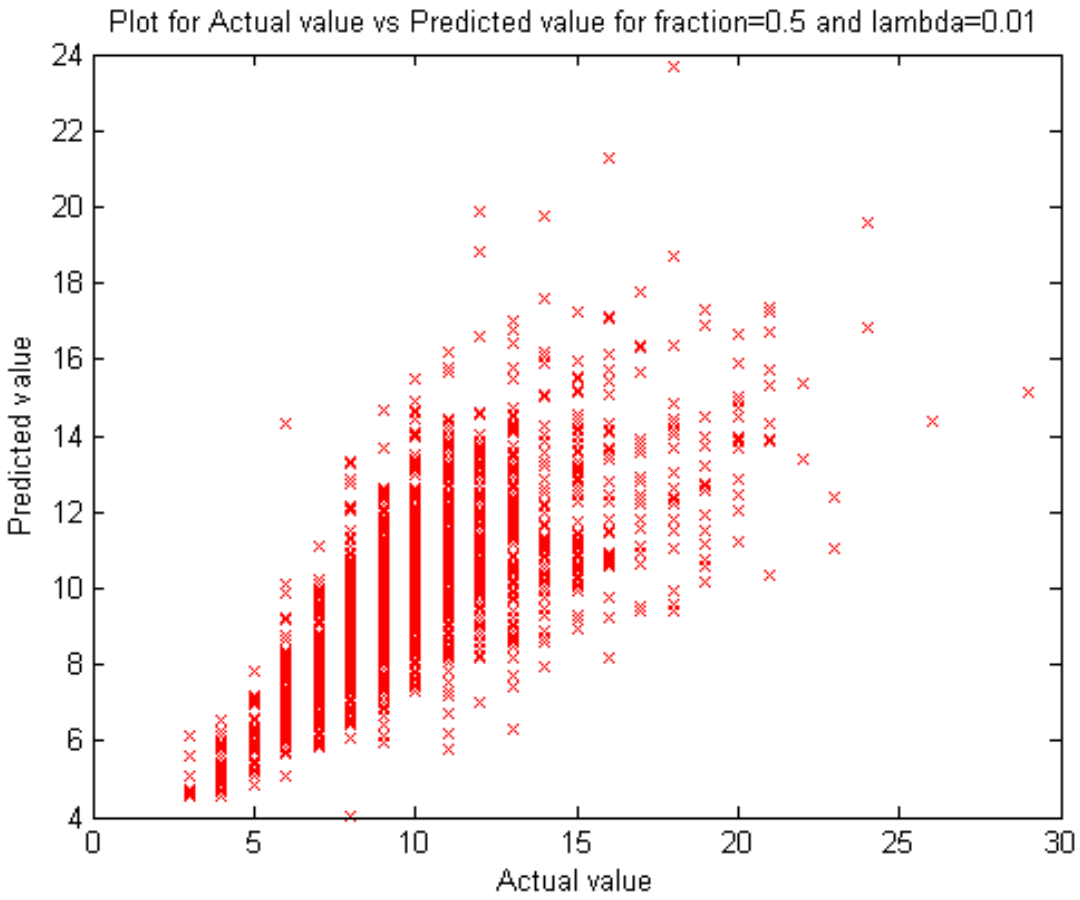
When we remove two or three least significant attributes, the mean squared error sometimes change very little and sometimes it increases a bit. For example, in one case it increased from 2.5204 to 2.8807.
I varied lambda from 10^{-5} to 10^{4} and observed that for all the values of fraction splits, mean squared error goes up with increasing lambda.










Plot for Actual value vs Predicted value for fraction=0.5 and lambda=0.01
Predicted value
Actual value
24
22
20
18
16
14
12
10
8
6
4
0
5
10
15
20
25
30

Problem 2

(a)

The training residual sum of squares for linear regression should be more than that of quartic regression in general. In quartic regression curve can be flexible enough to resemble the spatial distribution of training points, which will be a rather difficult task in linear regression.

(b)

The training residual sum of squares for linear regression should be less than that of quartic regression in general as a result of over-fitting in case of quartic regression. As we know that original relationship between X and T is linear, so it should fair better on test set than that of quartic regression.

(c)

The training residual sum of squares for linear regression should be more than that of quartic regression in general. The reason being same as part (a).

(d)

We don't have enough information to answer this question. It depends on how far is the actual relationship between X and T from linear. Closer this relationship is to linear, more probable is the case where linear regression will have less residual sum of squares than quartic regression.