Mohit Garg

2012CSB1020

CSL407 Machine Learning

Homework 1

Q : Pick a value for ʎ and examine the weights of the ridge regression model. Which are the most significant attributes? Try removing two or three of the least significant attributes and observe how the mean squared errors change.

ANS : For a given value of ʎ , Most significant attributes are corresponding to the constant term , and the 5th and 6th column in the input data .

After Transforming the first column ,

$$y(x) = w0 + w1 * x1 + w2 * x2 + \cdots + w10 * x10$$

$$J = \sum_{n=1}^{N} (w^t * x_n - t_n) \; + \; \lambda * || \, W \, ||$$

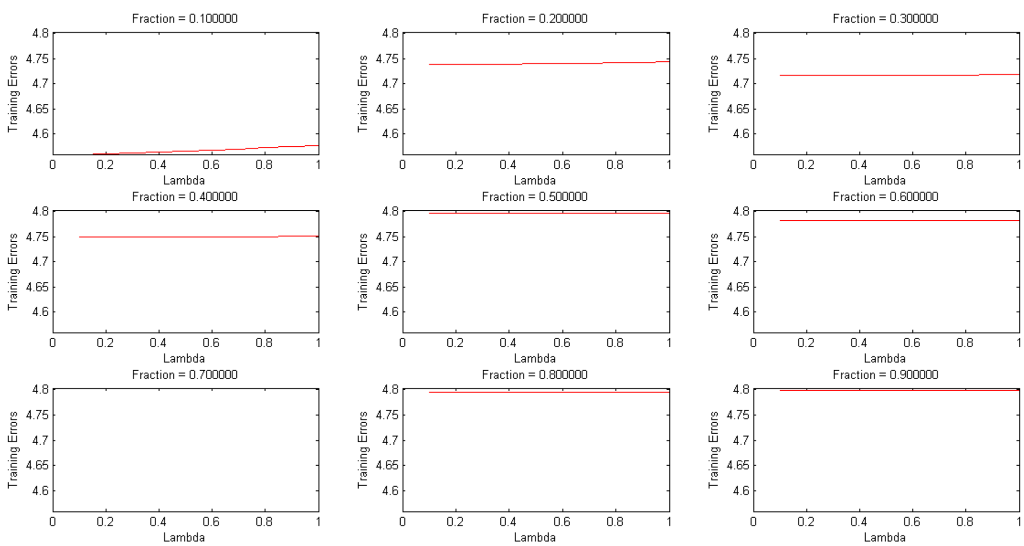
Therefore the most important attributes are corresponding to w0 , w7 and w8 .

While the least significant terms are the attributes encoding male , female or infant . After removing the terms corresponding to these terms , error slightly increased .

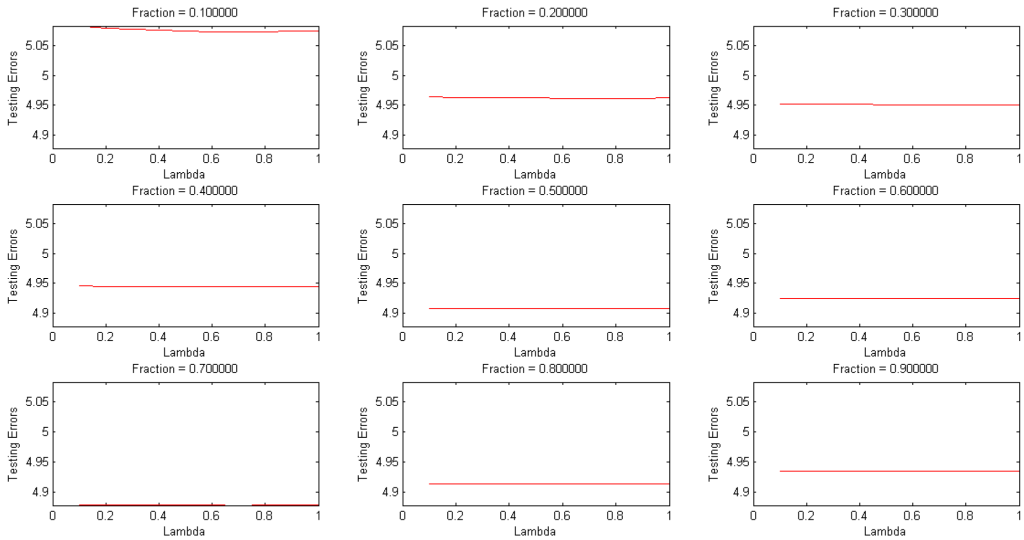
Q: Does the effect off ʎ on error change for different partitions of the data into training and testing sets?

ANS : Yes the effect of ʎ on error change for different partitions of the data into training and testing sets i.e. the effect of ʎ on error change for different values of fraction .

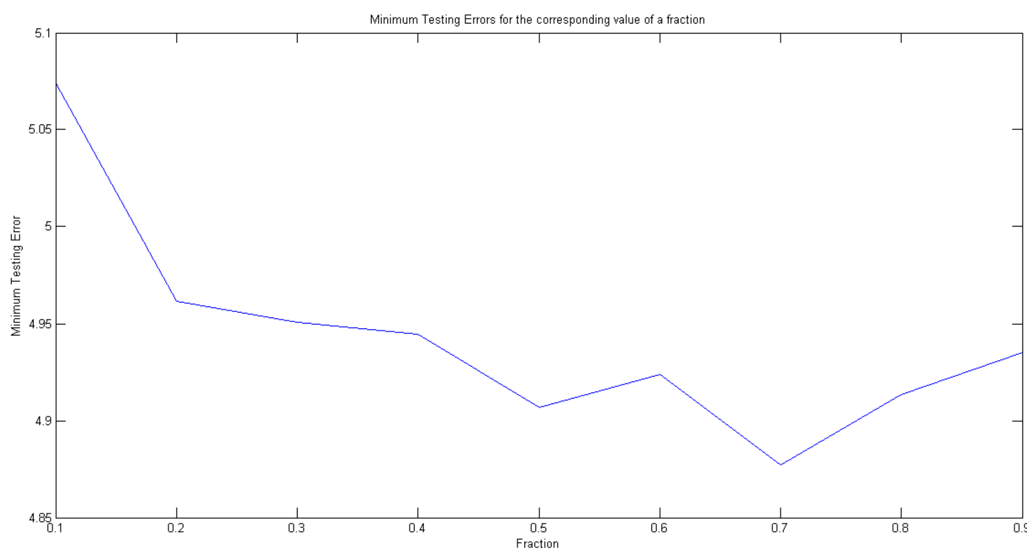
For training data , for the same value of ʎ , the error increases as the value of fraction increases .



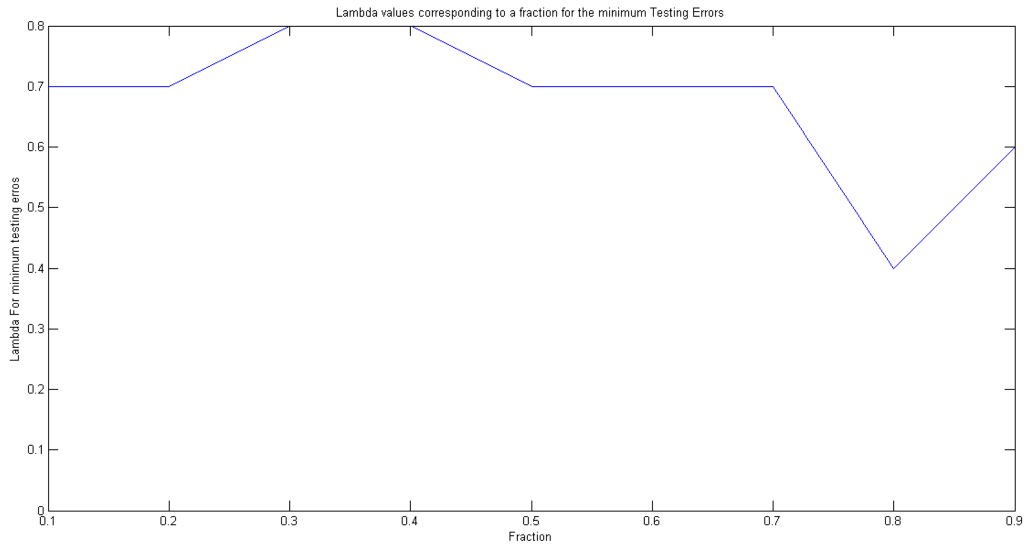
For testing data , for the same value of Λ error decreases as the value of fraction increases .



Also for the given values of fraction , the minimum testing error occur for diiferent values of lambda .



Minimum Error for the given value of fraction



Lambda Value corresponding to the minimum value of error for a given value of fraction

Thus , the effect of λ error changes as the value of fraction changes .

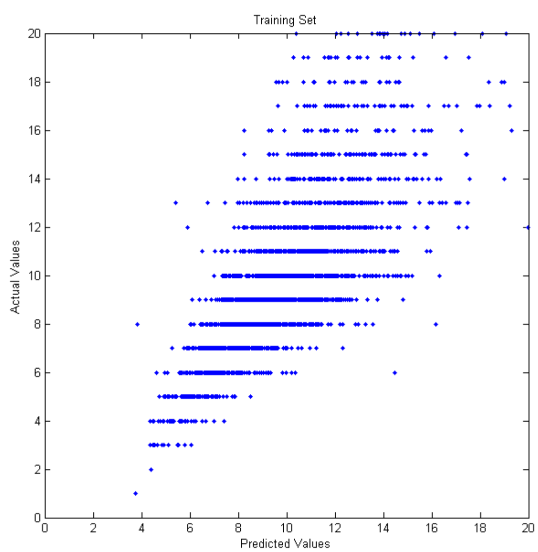
Q: How do we know if we have learned a good model?

ANS : The Best value of fraction and lambda , is found out to be ,

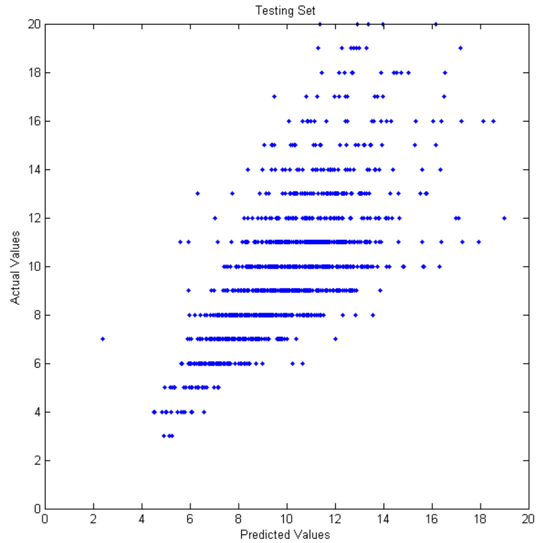
Best choice of Fraction = 0.700000

Best choice of Lambda = 0.700000

Minimum Error = 4.877076



Actual Values of Training Sets VS Predicted Values of Training Sets



Actual Values of Testing Sets VS Predicted Values of Testing Sets

As we can see the actual vs Predicted values lies near to the line corresponding to the angle of 45° , for both the case of the training as well as the testing values .

Q: I have collected a set of data (N=50 observations) containing a single attribute and a target response. I then fit a linear regression model to the data, as well as a separate quartic regression, i.e. y(x) = w0 + w1x + w2x2 + w3x3 + w4x4 .
(a) Suppose that the true relationship between X and T is linear, i.e. t = w0+w1x+_. Consider the training residual sum of squares (RSS) for linear regression and quartic regression. Would we expect one to lower than the other, would we expect them to be the same, or is there not enough information to tell? Justify your answer.

Ans (a) : For the training RSS , quartic regression will produce better results than the linear one , i.e. RSS of quartic will be lower , since we are considering higher order terms , and this will try to fit the curve better and thus producing better results , than the linear one , because in case of linear function error will be more.

(b) Answer (a) using test rather than training RSS.

Ans (b) : For the test RSS , linear regression will produce better results , i.e. RSS of linear function will be better than the quartic regression . As the original function was linear , therefore by using the quartic regression , we overfitted the training set , and hence the test error will be more , and hence RSS value of quartic regression will be more .

(c) Suppose that the true relationship between X and T is not linear, but we don't know how far is it from linear. Consider the training RSS for linear and quartic regressions. Would we expect one to be lower than the other, would we expect them to be the same, or is there no enough information to tell? Justify your answer.
Ans (c) : Similar to part (a) , for the training RSS , quartic regression will produce better results than the linear one , i.e. RSS of quartic will be lower , since we are considering higher order terms , and this will try to fit the curve better and thus producing better results , than the linear one , because in case of linear function error will be more.

(d) Answer (c) using test rather than training RSS.
Ans (d) : For the test RSS , quartic regression may or may not be better than the linear regression , if the relationship is more closer to the linear than the quartic will cause overfitting and will produce bad results , but if the relationship is more closer to quartic than the linear , than quartic will produce better results , as it will try to better fit the curve .

Q: Consider a data set in which each data point tn is associated with a weighting factor
rn > 0, so that the sum-of-squares error function becomes .Find an expression for the solution ^ w that minimizes this error function.
ANS : Taking the weighting factor inside the bracket it becomes
Summation of (sqrt(r)*t – W(transpose)* sqrt(r) * (x))
Consider a diagonal matrix (R)whose diagonal enteries are corresponding to the sqrt(r) values .
Now , X= (R)*x
T=(R)*t

Also solution = inv(X(transpose)*X)*(X)(transpose)*t
= inv (x (transpose) * r (transpose) * r * x) * (x)transpose *(r) (transpose) * r (transpose)

Let P a diagonal matrix whose diagonal enteries are corresponding to the r values .
Therefore , P = r(transpose) * r
Ans = inv(X' * P * X) * X' * P * T