Objective Function: (y-yhat)^2 = t(y-yhat)\*(y-yhat)

Constraints: I am using the number of rows in the test set as the number of constraints, so it keeps changing. But the constraints are basically the data points. 511 in my initial case.

Problem: To create an optimization model using mosek to predict the number of comments received in a particular blog by minimizing the loss function. We are using a quadratic function here as an objective function. The sample of train data is trained on linear model to check the squared error value after predicting from the test data. Then creating a model using Rmosek to find the optimal solution.

Dividing the data: I have used K fold cross validation method to divide the data in to 10 folds and randomly sampled it to train and test in the preprocess file. It is more efficient because of generating more number of datasets which are different from each other.

Building a model: The first model I used is the linear regression model using the lm function, I had to remove column 55 and 60 due to rank deficiency to make the model more accurate. The RSS generated from the fit was 2491630 and the **MSSE 556**

Rmosek Model: This model starts with a list, objective function c , A sparse matrix which contains the data points of the test set. I have put the constraint lower limit as 0 and constraint upper limit as Inf because a blog can have minimum zero comments and maximum Infinity comments. Next up the bounds to the variables are set by finding the minimum and maximum of the data. Then I pass the model in the mosek function.

Interior-point solution summary

Problem status : PRIMAL\_AND\_DUAL\_FEASIBLE

Solution status : OPTIMAL

Primal. obj: 1.0709656936e-05 nrm: 2e-10 Viol. con: 6e-11 var: 9e-13

Dual. obj: -1.6222399360e-10 nrm: 5e+06 Viol. con: 0e+00 var: 3e-08

Basic solution summary

Problem status : PRIMAL\_AND\_DUAL\_FEASIBLE

Solution status : OPTIMAL

Primal. obj: 8.6781909018e-06 nrm: 2e-10 Viol. con: 7e-11 var: 7e-13

Dual. obj: -1.0131299553e-04 nrm: 2e+05 Viol. con: 1e-13 var: 7e-09

$bas

$bas$`solsta`

[1] "OPTIMAL"

$bas$prosta

[1] "PRIMAL\_AND\_DUAL\_FEASIBLE"

$bas$xx

[1] 1.006837e-12 0.000000e+00 -1.467584e-13 -5.542978e-13 1.028866e-11

[6] 3.185862e-11 0.000000e+00 5.078805e-11 2.510361e-12 9.023893e-13

**The mean sum of squres after optimal solution in mosek is 116.2352.**