CS 498 AML ONL: HW6

Ramya P Narayanaswamy (rpn2), Qingkang Zhang (qzhang72)

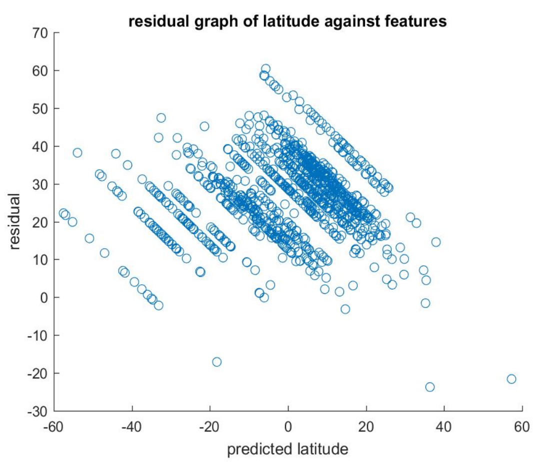
Problem 1

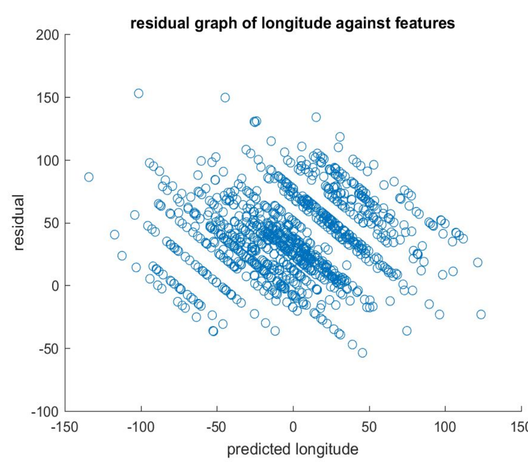
First, build a straightforward linear regression of latitude (resp. longitude) against features. What is the R-squared? Plot a graph evaluating each regression.

Code : problem1a.m, prob1.R : This section was in matlab and in R

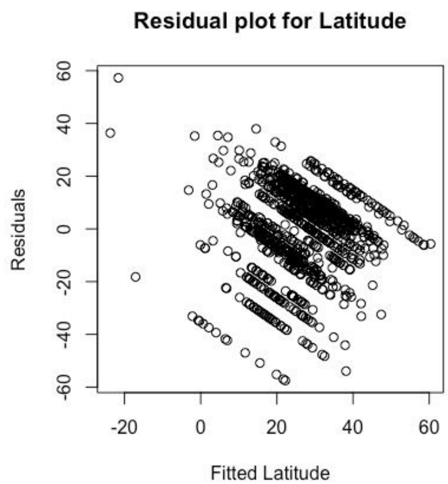
R-squared values for latitude and longitude regression are 0.293 and 0.365 respectively
Residual plot for each is shown below and the plot resembles random noise, indicating this is a decent regression

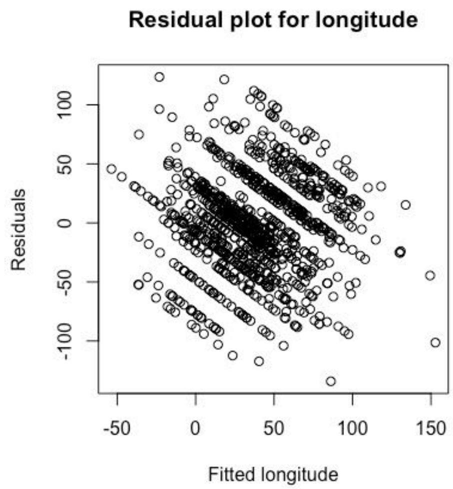
Matlab plots:





R plots :





For comparing with rest of the sections, R equivalent plots are used

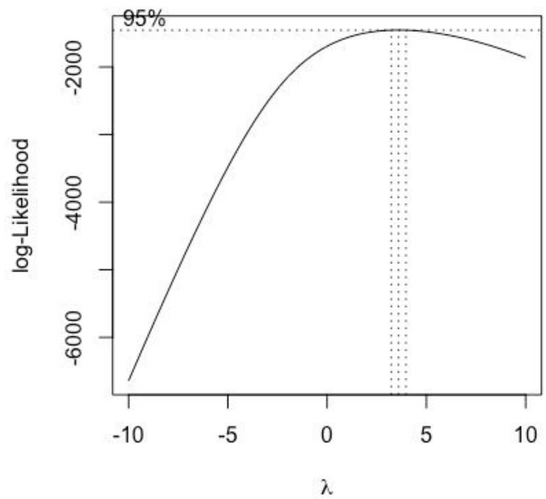
Does a Box-Cox transformation improve the regressions? ***Notice that the dependent variable has some negative values, which Box-Cox doesn't like. You can deal with this by remembering that these are angles, so you get to choose the origin.*** *why do you say so? For the rest of the exercise, use the transformation if it does improve things, otherwise, use the raw data.*

Code : prob1b.R

Ans : Adding a constant to angles, rotates the blob without affecting individual points. The points which are closer in original blob remain closer after shifting the origin. The same applies for far away points. Thus, adding a constant does not affect the results for box cox transformation. 90 was added to latitude and 180 was added to longitude. The function boxcox from MASS library and lm() were used in code. The residual plot and R2 were transformed to original coordinates

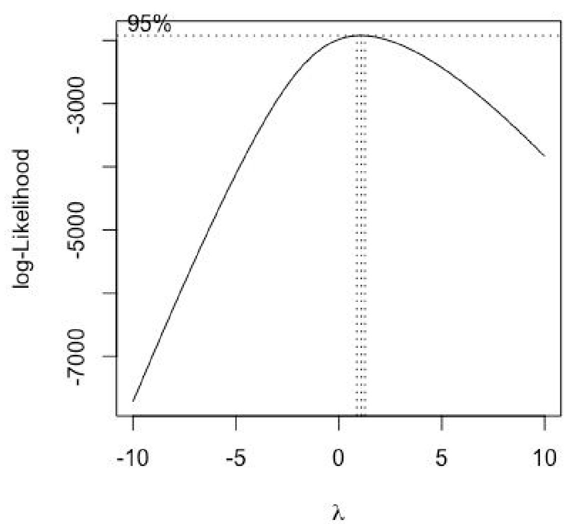
Latitude :
Best Box cox lambda : 3.6. This lambda is chosen as it maximizes the log-likelihood
R2 (after transforming to original coordinates) : 0.248

The plot for log-likelihood vs lambda for latitude is shown below



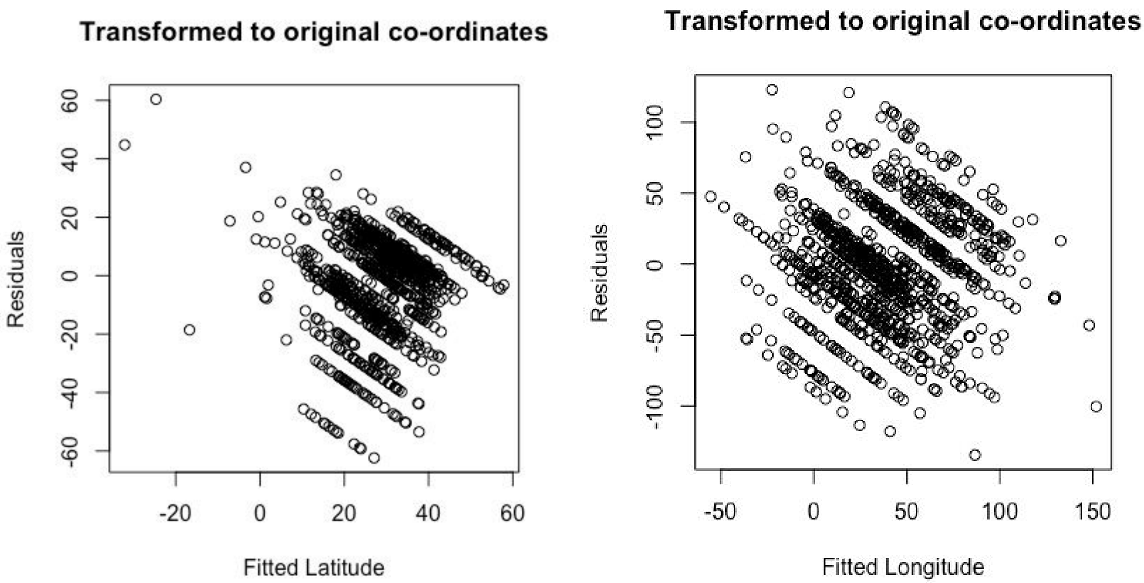
Latitude BoxCox

Longitude:
Box cox lambda : 1.1. This lambda is chosen as it maximizes the log-likelihood
R2 (after transforming to original coordinates) : 0.363



Longitude BoxCox

Residual plots of Boxcox after transformation is shown below



Conclusion : The residual plots are almost similar for boxcox vs non-boxcox. They both resemble random noise phenomenon. The values of R-squared for both latitude and longitude did not improve with boxcox. Thus, boxcox did not bring in significant advantage to this problem. So for all the below sections, no boxcox was used

Part3 :

Code : prob1c_lat.R, prob1c_long.R

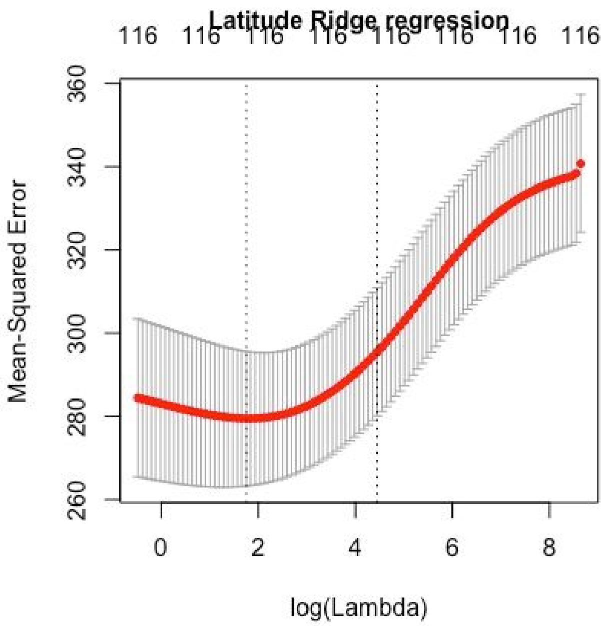
Cv.glmnet was used to get the lambda that produces minimum mean square error. The parameter nfolds is set to 20. No box cox transformation was used

Latitude:

Unregularized regression :cv.lm from library DAAG was used and 20 fold cross-validation was done
MSE = 286
Number of parameters : 117 (116 are weights for independent variables and one is for intercept)

Ridge regression :
Lambda min = 5.778
MSE corresponding to lambda min = 279.466
Number of parameters : 117 (116 are weights for independent variables and one is for intercept)

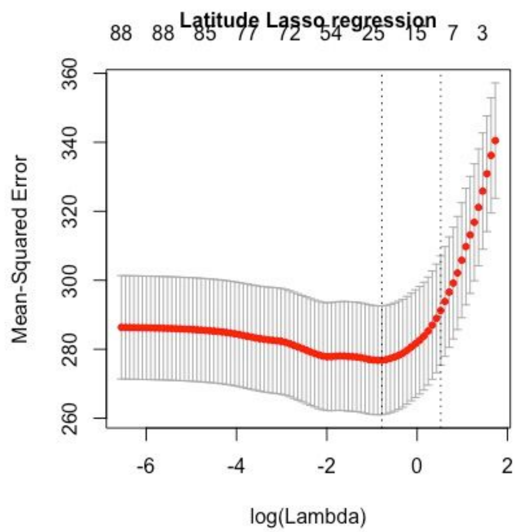
Conclusion : The region to the left of lambda in the below plot is almost constant. There is a small improvement in MSE. Overall, we conclude that ridge regularization for this set of data has a very minor improvement and the effect could be better with unseen data sets



Lasso regression :
Lambda min = 0.458
MSE corresponding to lambda min = 276.790

Number of nonzero parameters is 22. Here one value is for intercept and 21 coefficients are for weighing independent variables

Conclusion : The region to the left of lambda in the below plot is almost constant. Error-wise there is minor improvement when compared to unregularized part. The significant result here is we are able to achieve small errors with reduced number of independent variables, as seen from the plot. We conclude that the Lasso regression might be useful as it uses only a small subset of independent variables (21) and produces a somewhat smaller error.

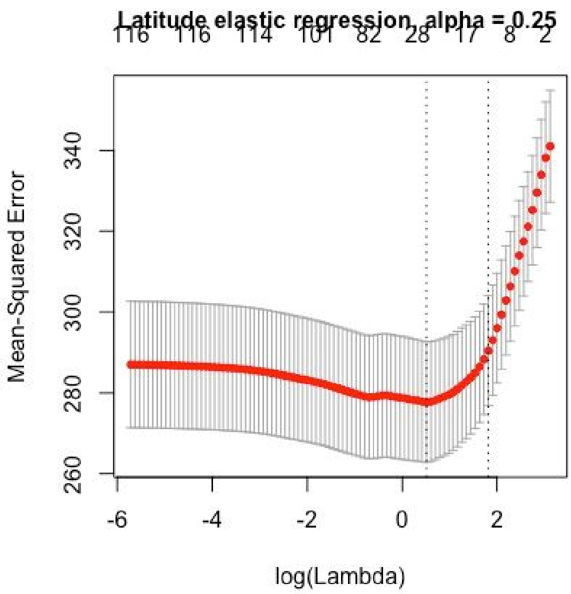


Elastic Regression (Alpha = 0.25):

Lambda min = 1.669
MSE corresponding to lambda min = 278.489

Number of nonzero parameters is 23. Here one value is for intercept and 22 coefficients are for weighing independent variables

Conclusion : The region to the left of lambda in the below plot is almost constant. Error-wise there is a minor improvement when compared to unregularized part. The significant result here is we are able to achieve small errors with reduced number of independent variables, as seen from the plot. We conclude that the elastic regression with 0.25 might be useful as it uses only a small subset of independent variables (22) and produces a somewhat smaller error.

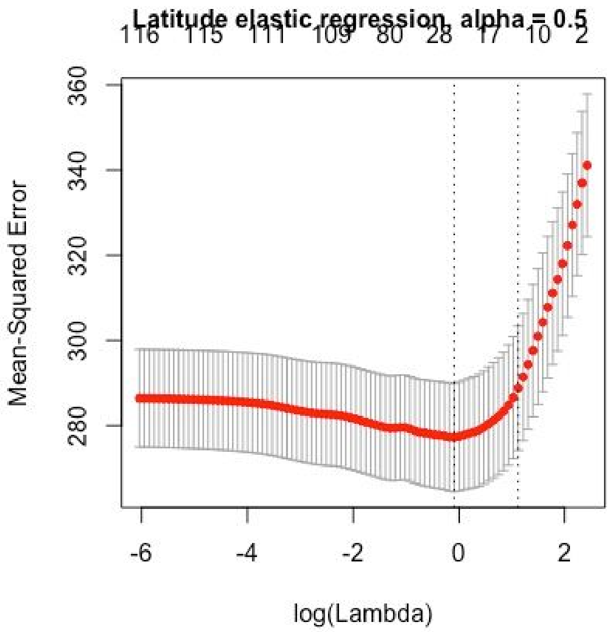


Elastic Regression (Alpha = 0.5):

Lambda min = 0.915
MSE corresponding to lambda min = 278.177

Number of nonzero parameters is 23. Here one value is for intercept and 22 coefficients are for weighing independent variables

Conclusion : The region to the left of lambda in the below plot is almost constant. Error-wise there is a minor improvement when compared to unregularized part. The significant result here is we are able to achieve small errors with reduced number of independent variables, as seen from the plot. We conclude that the elastic regression with 0.5 might be useful as it uses only a small subset of independent variables (22) and produces a somewhat smaller error.

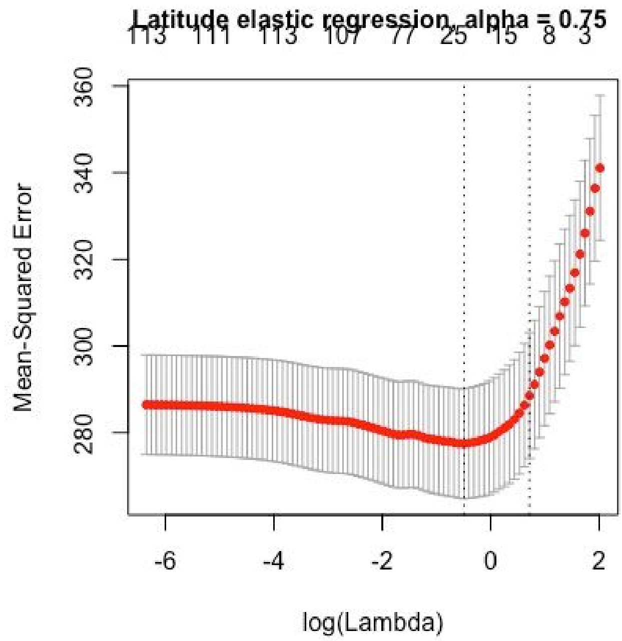


Elastic Regression (Alpha = 0.75):

Lambda min = 0.610
MSE corresponding to lambda min = 277.428

Number of nonzero parameters is 22. Here one value is for intercept and 21 coefficients are for weighing independent variables

Conclusion : The region to the left of lambda in the below plot is almost constant. Error-wise there is not much an improvement when compared to unregularized part. The significant result here is we are able to achieve small errors with reduced number of independent variables, as seen from the plot. We conclude that the elastic regression with 0.75 might be useful as it uses only a small subset of independent variables (21) and produces a somewhat smaller error.



__Overall conclusion__: Adding a small amount of lasso (i.e elastic) reduces the number of variables required for regression. Thus, elastic net with small alpha in range 0.25 is sufficient for latitude regression. When comparing the minimum mean squared errors, the values are close-enough between various regression models.

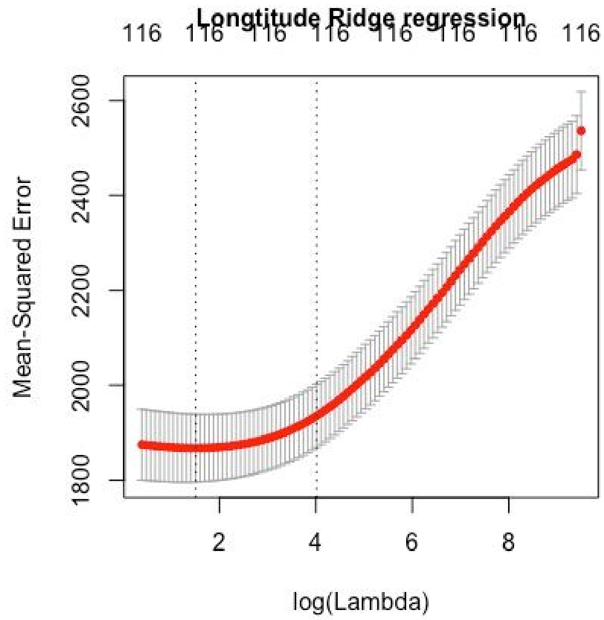
__Longitude:__

Unregularized regression :cv.lm from library DAAG was used and 20 fold cross-validation was done

MSE = 1906
Number of parameters : 117 (116 are weights for independent variables and one is for intercept)

Ridge regression :
Lambda min = 4.497
MSE corresponding to lambda min = 1867.664
Number of parameters : 117 (116 are weights for independent variables and one is for intercept)

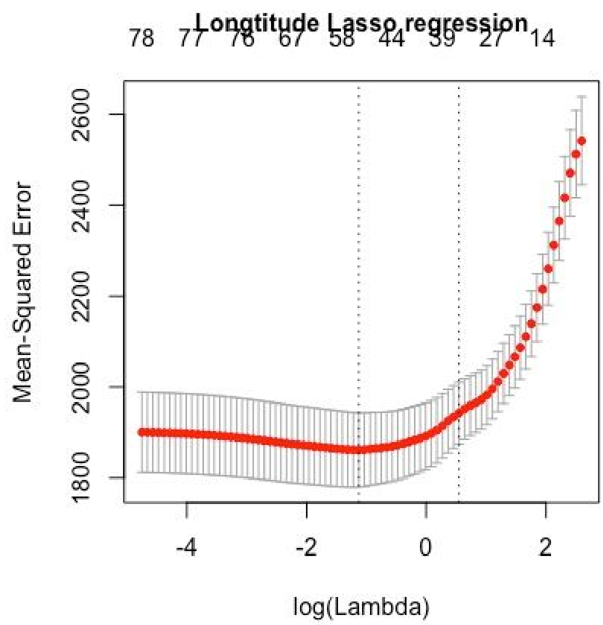
Conclusion : The region to the left of lambda in the below plot is almost constant. Overall, we conclude that ridge regularization for this set of data has a small improvement. However, we may see significant improvement with unseen data sets



Lasso regression :
Lambda min = 0.325
MSE corresponding to lambda min = 1861.659

Number of nonzero parameters is 54. Here one value is for intercept and 53 coefficients are for weighing independent variables

Conclusion : The region to the left of lambda in the below plot is almost constant. When compared to unregularized version, there is a small improvement in mean square error. The significant result here is we are able to achieve small errors with reduced number of independent variables, as seen from the plot. We conclude that the Lasso regression is useful as it uses only ~half the number of independent variables (53) and produces a smaller error.

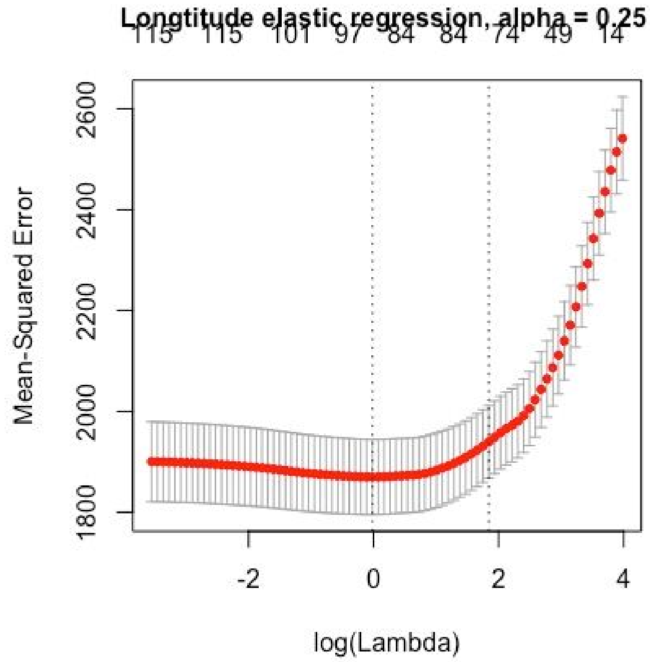


Elastic Regression (Alpha = 0.25):

Lambda min = 0.982
MSE corresponding to lambda min = 1869.898

Number of nonzero parameters is 92. Here one value is for intercept and 91 coefficients are for weighing independent variables

Conclusion : The region to the left of lambda in the below plot is almost constant. When compared to unregularized version, there is a small improvement in mean square error. We conclude that the elastic regression with alpha = 0.25 is not quite useful when compared to Lasso as it uses larger number of independent variables, but slightly better w.r.t unregularized version

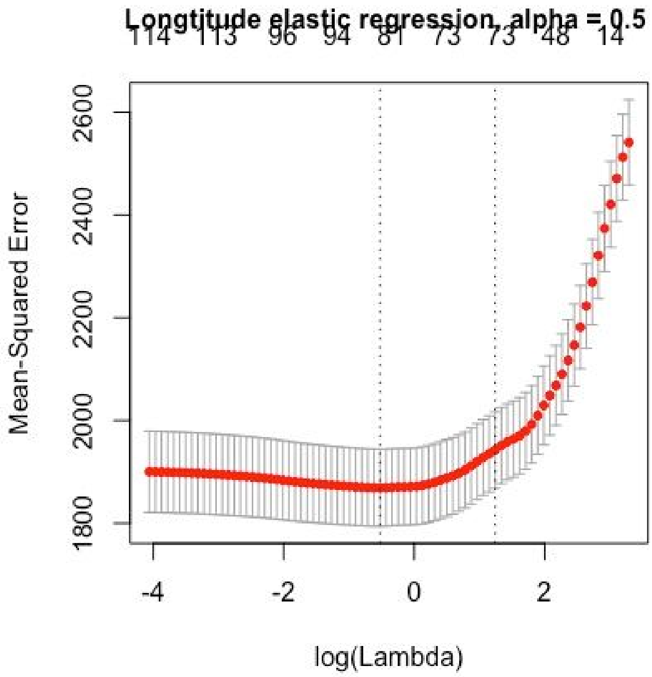


Elastic Regression (Alpha = 0.5):

Lambda min = 0.592
MSE corresponding to lambda min = 1869.07

Number of nonzero parameters is 86. Here one value is for intercept and 85 coefficients are for weighing independent variables

Conclusion : The region to the left of lambda in the below plot is almost constant. When compared to unregularized version, there is a small improvement in mean square error. We conclude that the elastic regression with alpha = 0.5 is not quite useful when compared to Lasso as it uses larger number of independent variables, but slightly better w.r.t unregularized model

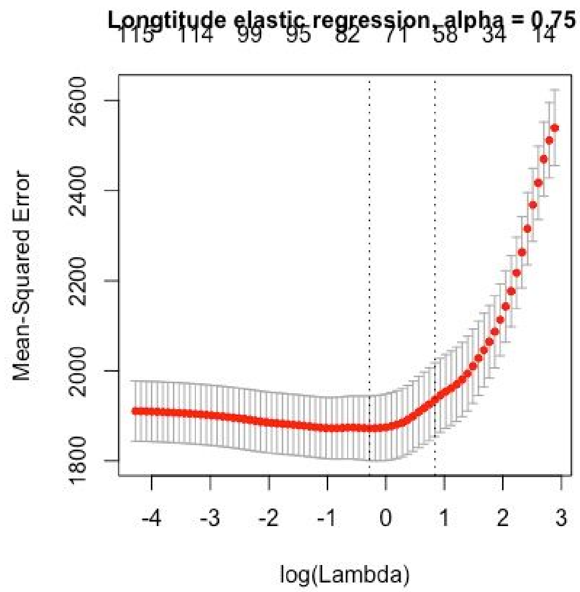


Elastic Regression (Alpha = 0.75):

Lambda min = 0.757
MSE corresponding to lambda min = 1872.047

Number of nonzero parameters is 77. Here one value is for intercept and 76 coefficients are for weighing independent variables

Conclusion : The region to the left of lambda in the below plot is almost constant. When compared to unregularized version, there is a small improvement in mean square error. We conclude that the elastic regression with alpha = 0.75 is not quite useful when compared to Lasso as it uses larger number of independent variables, but slightly better w.r.t unregularized version



Overall conclusion for Longitude: Lasso is considered best when compared to other forms of elastic regression due to lower number of independent variables. However, by just looking at mean-square error, there is a small improvement between regularized and unregularized versions

Part2:

Unregularized version : Part2.R

Regularized versions : Part2ridge.R, Part2lasso.R, Part2elas1.R, Part2elas2.R, Part2elas3.R

10-fold cross-validation was used to obtain accuracy in above codes. createFolds function from caret library was used. Note that all the errors below are **misclassification errors**.

Unregularized : glm and predict functions from library glmnet were used. The following parameter setting was used to enable binomial logistic regression. family=binomial(link='logit'). The training is done using glm and test data is predicted using predict function for all the 10 iterations.

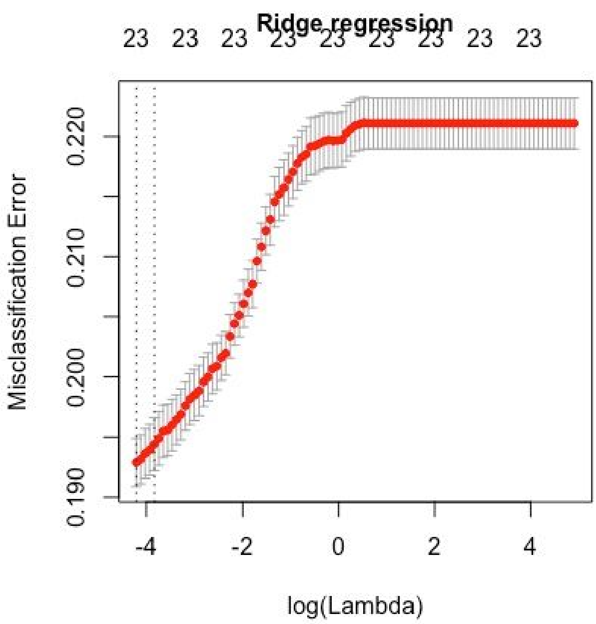
Error in 10 folds = [0.194 0.195 0.177 0.181 0.198 0.181 0.189 0.192 0.198 0.189]
Number of parameters used : 24 (23 independent variables and 1 for intercept)

Average accuracy with 10-fold cross-validation : 81.1%

Regularization: The outer loop is 10-fold cross-validation. To find a good lambda value, cv.glmnet is used on the training data (in every iteration). The predict is done on test data using lambda that resulted in minimum classification error. The following are required parameters set for cv.glmnet type.measure="class",family="binomial". The predict parameters are s = lambda.min,type = "class". Ridge, Lasso and 3 values are tried for Elastic regression.

Ridge Regression: Alpha = 0

The cv.glmnet plot for ***one of the runs*** of outer-loop for obtaining lamda is shown here as a sample. The lambda min, misclassification error produced by min lambda for all 10 runs are shown below



Lambda min= [0.0148 0.0149 0.0147 0.0149 0.0148 0.0147 0.0148 0.0146 0.0149 0.0148]
CV misclassification error = [0.201 0.197 0.183 0.183 0.197 0.184 0.193 0.200 0.201 0.194]
Number of parameters in all 10 runs = 24 (23 independent variables and 1 for intercept)

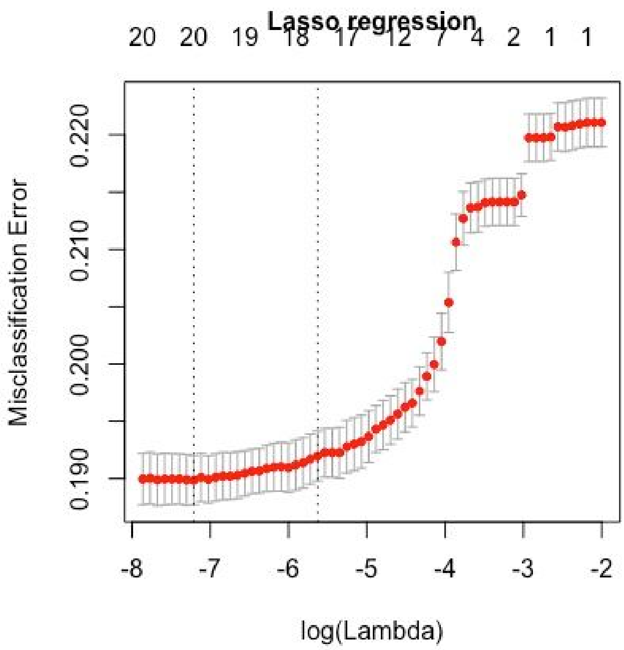
Average accuracy with 10-fold cross-validation : 80.7%

Lasso Regression: Alpha = 1

The cv.glmnet plot for ***one of the runs*** of outer-loop for obtaining lamda is shown here as a sample. The lambda min, misclassification error produced by min lambda for all 10 runs are shown below

Lambda min =[0.0003 0.0003 0.0003 0.0004 0.0004 0.0005 0.0005 0.0004 0.0004 0.0007]
CV misclassification error = [0.194 0.194 0.177 0.180 0.198 0.182 0.190 0.193 0.198 0.189]
Number of parameters including intercept = [24 22 23 22 23 23 21 23 22 21]

Average accuracy with 10-fold cross-validation : 81.07%

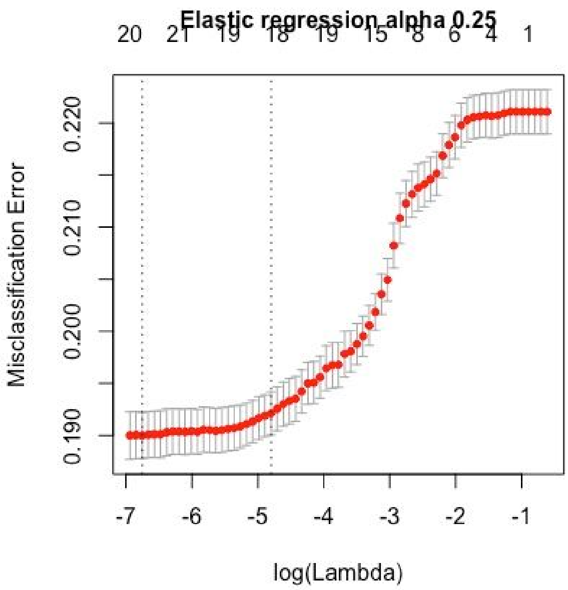


Elastic Regression: Alpha = 0.25

The cv.glmnet plot for ***one of the runs*** of outer-loop for obtaining lamda is shown here as a sample. The lambda min, misclassification error produced by min lambda for all 10 runs are shown below

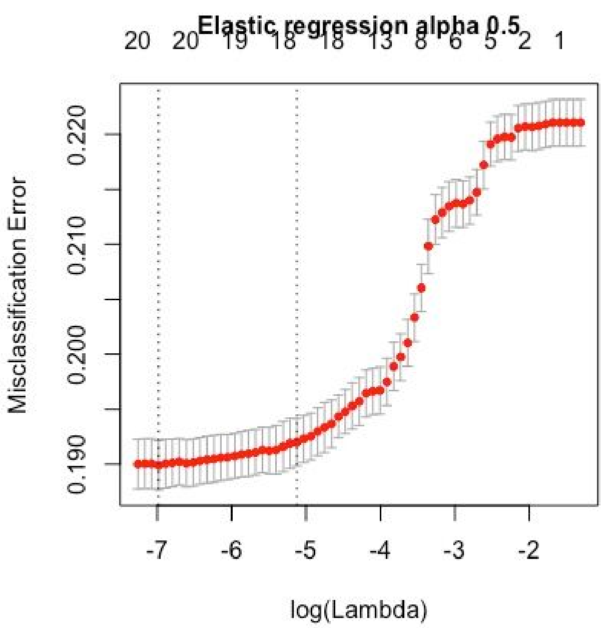
Lambda min =[0.001 0.0008 0.0009 0.0006 0.0008 0.0007 0.0001 0.001 0.001 0.001
CV misclassification error = [0.194 0.194 0.176 0.180 0.197 0.181 0.189 0.193 0.198 0.189]
Number of parameters including intercept = [24 23 23 23 23 23 22 23 22 22]

Average accuracy with 10-fold cross-validation : 81.05%



Elastic Regression: Alpha = 0.5

The cv.glmnet plot for ***one of the runs*** of outer-loop for obtaining lamda is shown here as a sample. The lambda min, misclassification error produced by min lambda for all 10 runs are shown below

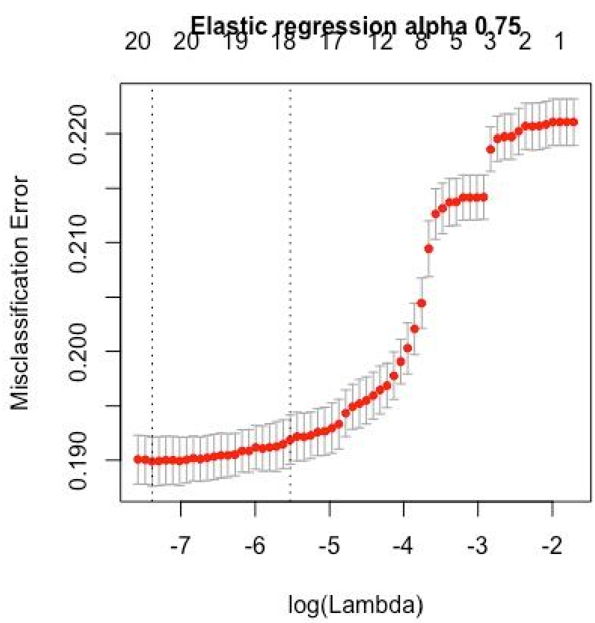


Lambda min =[0.0006 0.0005 0.0009 0.0004 0.0009 0.0004 0.0007 0.0009 0.0008 0.0009]
CV misclassification error = [0.194 0.194 0.177 0.180 0.197 0.181 0.190 0.193 0.198 0.189]
Number of parameters including intercept = [24 23 22 22 23 23 21 23 22 21]

Average accuracy with 10-fold cross-validation : 81.06%

Elastic Regression: Alpha = 0.75

The cv.glmnet plot for ***one of the runs*** of outer-loop for obtaining lamda is shown here as a sample. The lambda min, misclassification error produced by min lambda for all 10 runs are



shown below

Lambda min =[0.0005 0.0005 0.0006 0.0003 0.0005 0.0003 0.0005 0.0006 0.0006 0.0006]
CV misclassification error = [0.194 0.193 0.177 0.180 0.197 0.181 0.190 0.193 0.198 0.189]
Number of parameters including intercept = [24 23 22 22 23 23 21 23 22 21]

Average accuracy with 10-fold cross-validation : 81.07%

Conclusion : Based on the average cross-validated accuracy for regularized and unregularized portion, we conclude that the regularization did not improve this binomial logistic regression. It is also to be noted that the Lasso and elastic did not reduce the number of parameters significantly, hence the advantage of using less parameters does not play out in this case. Furthermore, the optimal lambda values are in small ranges, confirming that regularization is not very significant.