Problem 1

Data preparation

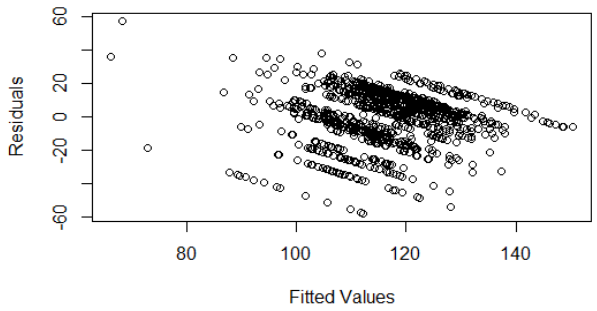
- Latitude and longitude were shifted 90 so the boxcox would not get affected, it doesn't matter if we do so from the start for all the problems.

Part A

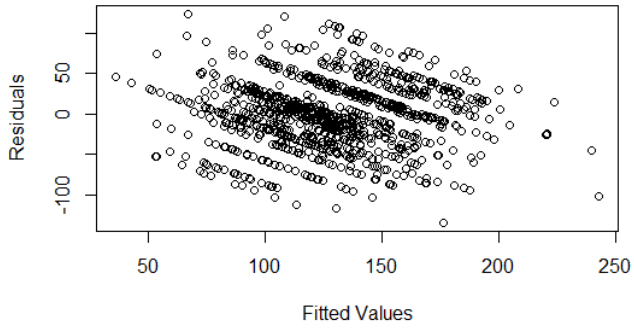
The latitude model had an R-Squared equal to 0.291, while cross validated mean square error was 267.

The longitude model had an R-Squared equal to 0.37. Cross validated mean square error over the testing data was 1907

Note: Number of folds is 10 for all the models



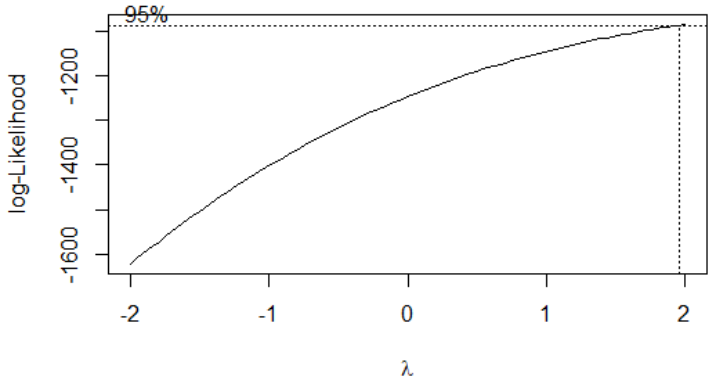
Residuals against values for latitude regression



Residuals against values for longtitude regression

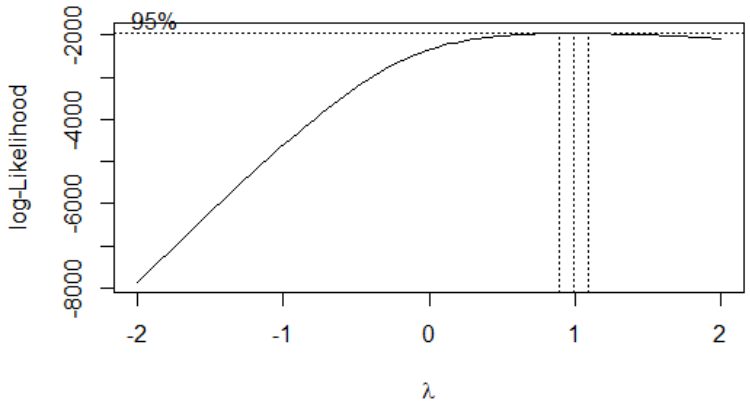
Part B

For the BoxCox Lambda selection over the latitude, I applied the boxcox function and it gave me best lambda equal to 2 as follows:



I then applied the transformation over the data and built a regression model over the transformed data, then converted the data back to calculate the R-Squared was 0.266. MSE over the testing data was 268.

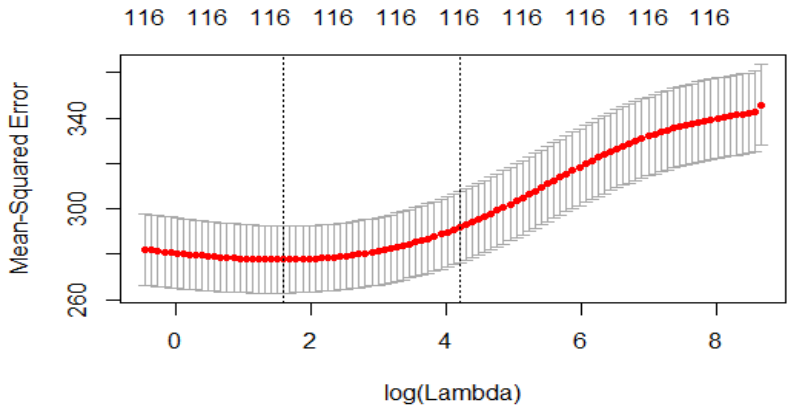
As for longitude regression, best lambda was 1 as shown in the figure below, so it doesn't make since to try it as there's no actual transformation for the data, just shifting. R -Squared and MSE to be reported here are identical to the original model, 0.37 and 1907.16 respectively.



Decision is not to use BoxCox furthermore since it hasn't added any value in neither the R-Squared nor the MSE for both regressions.

Part C.1 (Ridge regularization)

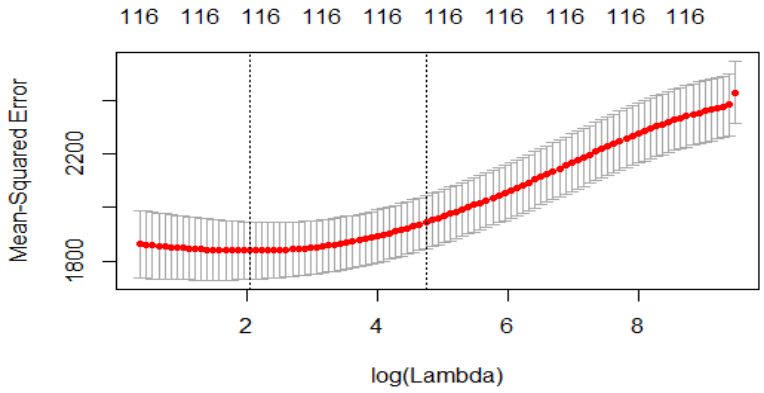
For the case of latitude regression with a ridge regularization, min lambda was 4.94 (deducted from the below figure), and it resulted in R-Squared equal to 0.21. CV error is 278



Lambda selection and MSE over training data for Ridge Latitude regression

For the case of longitude regression with a ridge regularization, min lambda was 7.69 (deducted from the below figure), and it resulted in R-Squared equal to 0.291.

CV Error is 1875.2.



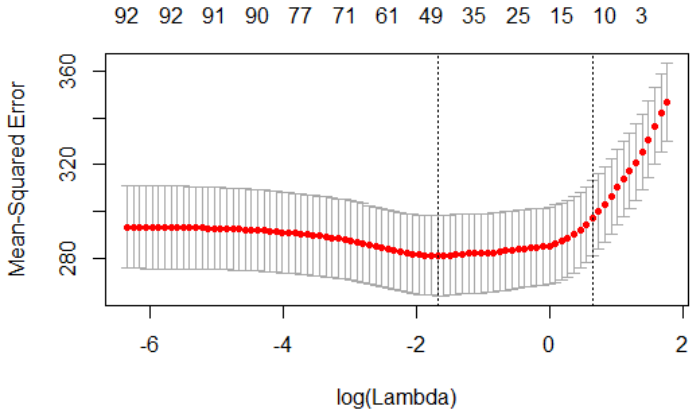
Lambda selection and MSE over training data for Ridge Longtitude regression

Conclusion is that regularization is better than the original regression in the longitude regression only in the CV error, not the R-Squared though. I would say that the CV error is amore trustworthy metric.

Part C.2(Lasso regularization)

For the case of latitude regression with a lasso regularization, min lambda was 0.18 (deducted from the below figure), and it resulted in R-Squared equal to 0.26.

CV Error is 275 and number of non-zero coefficients for minimum lambda is 46

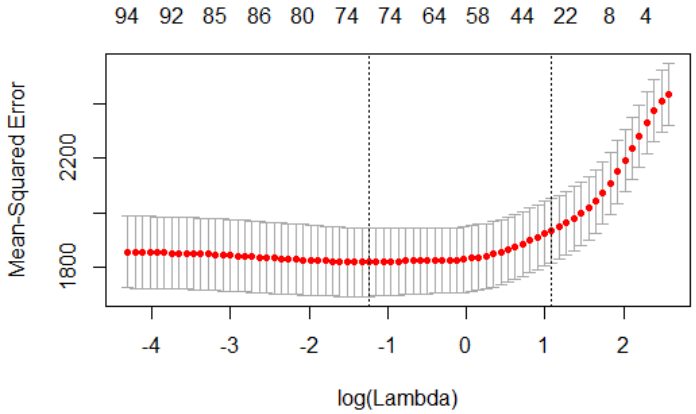


Lambda selection and MSE over training data for Lasso Latitude regression

For the case of longitude regression with a lasso regularization, min lambda was 0.28 (deducted from the below figure), and it resulted in R-Squared equal to 0.33.

CV Error is 1877.1 and number of non-zero coefficients for minimum lambda is 80

Note: Number of non-zero coefficients in case of unregularized model is always all the features (116).



Conclusion is that lasso regularization is better than the non-regularized regression in the longitude regression only (Less CV error)

Parc C.3

Here are the table that represents different numbers for Elastic net regularization against different alpha values, namely 0.25, 0.5, and 0.75.

Regression Task	Min Lambda	R-Squared	CV Error	# of Coefficients
Latitude	1.83	0.168	279	34
Longitude	1.18	0.3	1863	88

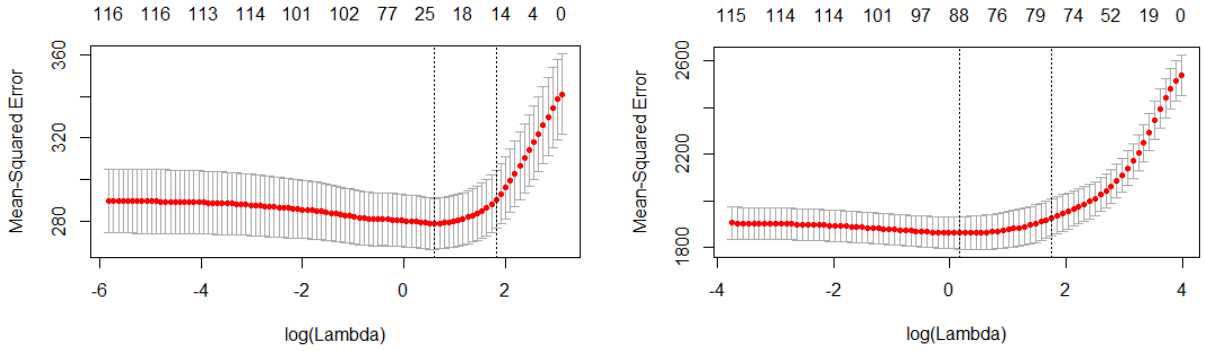
Metrics for Elastic net regularization with alpha equal to 0.25

Regression Task	Min Lambda	R-Squared	CV Error	# of Coefficients
Latitude	0.9	0.17	277	22
Longitude	0.44	0.33	1855	87

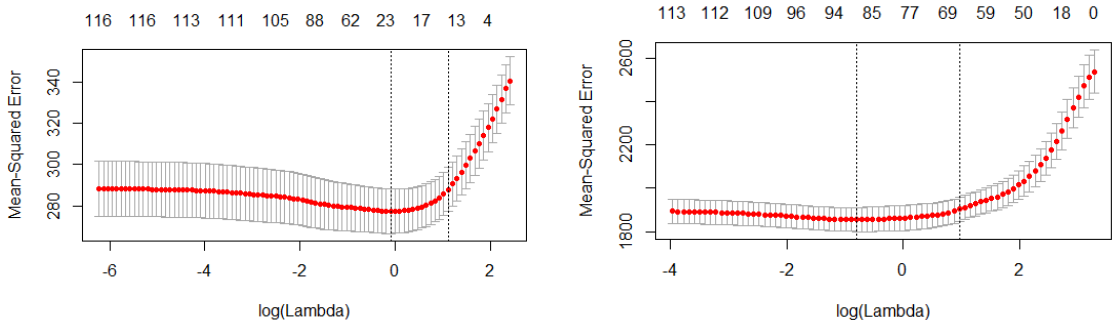
Metrics for Elastic net regularization with alpha equal to 0.5

Regression Task	Min Lambda	R-Squared	CV Error	# of Coefficients
Latitude	0.61	0.18	277	21
Longitude	0.29	0.32	1838	90

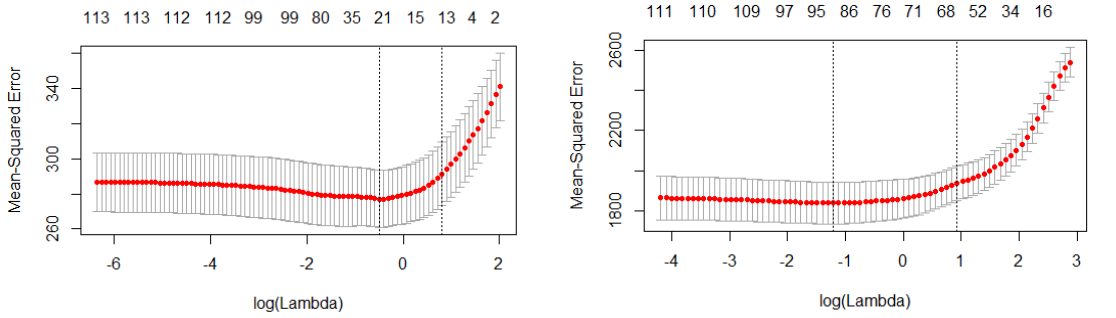
Metrics for Elastic net regularization with alpha equal to 0.75



Mean error against log of Lambda for alpha equal to 0.25



Mean error against log of Lambda for alpha equal to 0.5



Mean error against log of Lambda for alpha equal to 0.75

Conclusion is that elastic net regularization, like other regularization techniques, give better longitude regression based on their lower CV error.

Problem 2

Data Preparation

I am using the bigger file with more features, I removed the first row and the IDs column.

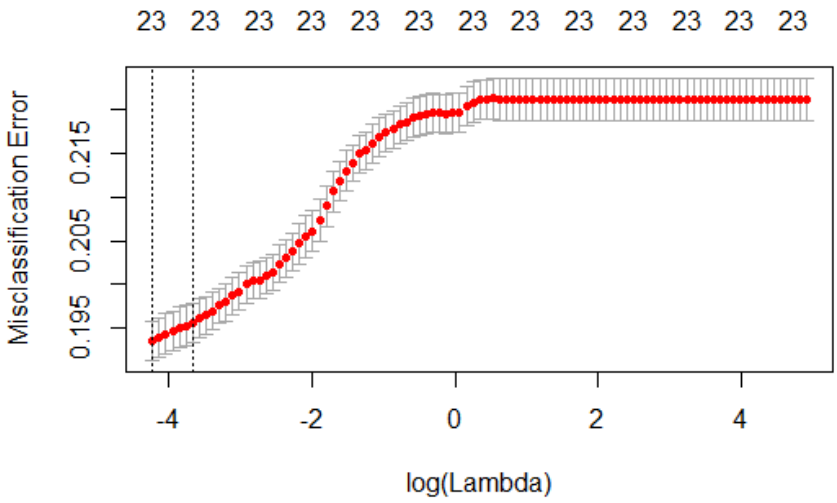
Regression

For regression without regularization, I used glmnet with manual k-folds, where number of folds is 10. For other regularized models, number of folds is 10, by default, and I am using lambda min as it gives me better results, I think it's not overfitting the data as the data is big enough, so no need for lambda.1se

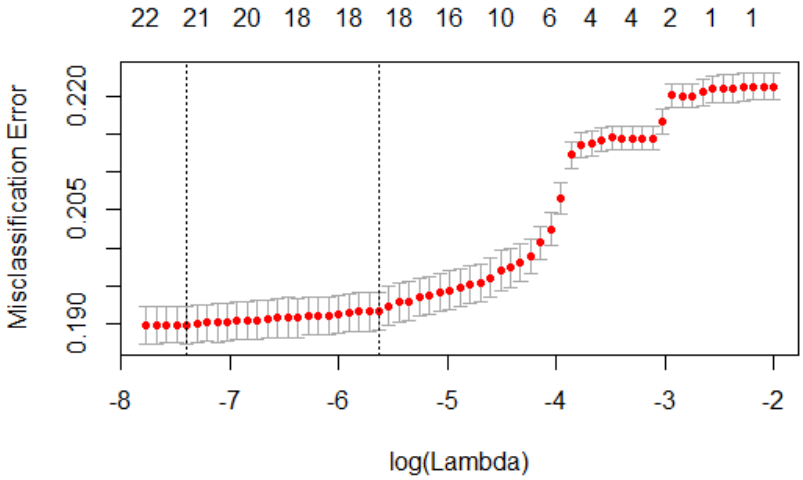
Technique	CV Misclassification error	# of parameters(Coefficients)	Best Lambda(min)
No regularization	19%	23	N/A
Ridge	22.1%	23	0.0148
Lasso	19%	21	0.000611
Elastic net (0.25)	19%	22	0.0014
Elastic net (0.5)	18.9%	22	0.000843
Elastic net (0.75)	18.9%	22	0.000562

Conclusion, by numbers, elastic net regression is giving the minimum misclassification error rate. Although there isn't much difference, but I'd prefer picking that model as it's less susceptible for overfitting (than non-regularized).

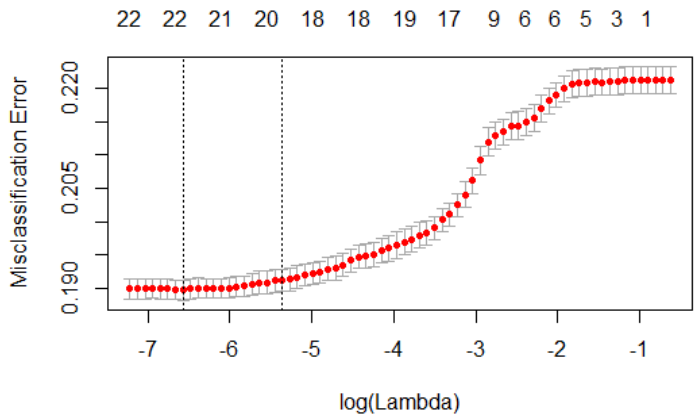
Finally, below are plots for misclassification error rate against different lambdas for the different regression techniques.



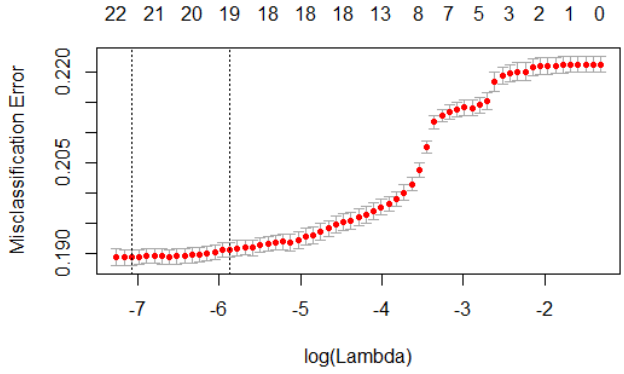
Misclassification error against Lambda for Ridge regression



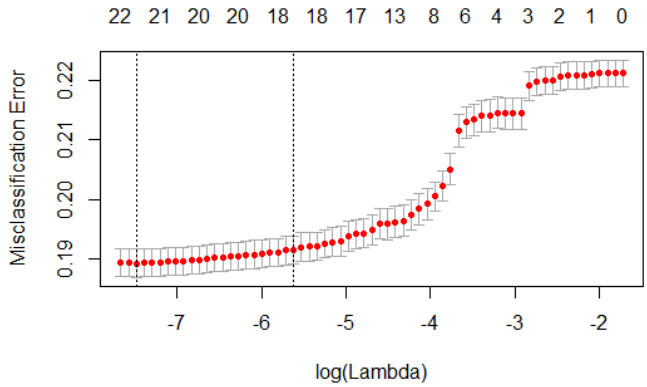
Misclassification error against Lambda for Lasso regression



Misclassification error against Lambda for elastic net regression (alpha=0.25)



Misclassification error against Lambda for elastic net regression (alpha=0.5)



Misclassification error against Lambda for elastic net regression (alpha=0.75)