

*Toronto*

**Module 4**

Dhairya Dave

*College of Professional Studies****,*** *Master of Professional Studies in Analytics.*

***Subject****: ALY6015* ***NUID****: 002110382*

Under the guidance of

**Dr. Prof. Alex Maizlish**

**Introduction**

Why do we need to use regularization in our model? How does regularization create an impact in the model?

Overfitting, multicollinearity, and heavy computational resources are some of the issues that arise while dealing with linear/multiple regression with several characteristics. With these issues, building a model with increased precision is challenging. Regularization becomes extremely crucial in solving these challenges. Regularization is also used when the model lacks stability and generalisation. Two effective regularisation approaches are Ridge and Lasso. These function by penalising the size of feature coefficients and reducing the standard error between predicted and observed values. The primary distinction between Ridge and Lasso is how the variables are penalised.

Ridge does L2 regularisation, which means it imposes a penalty equal to the square of the coefficient magnitude. Ridge has the drawback of include all p predictors in the final model, and the coefficients are decreased but not zero. Ridge regression is calculated using the formula below.

Lambda increases -> flexibility decreases -> bias error

Objective = RSS + α \* (sum of square of coefficients)

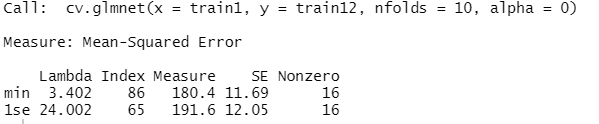
It is possible that the model will be overfitted if the coefficients are not penalised (that is regression line will be passing through all the points thus, it will have minimum residual error).

**Splitting the dataset into test and train dataset.**

Here, in this project the test and train dataset are divided into 30 % and 70% that is 70% goes to training set and 30% does to test set. Therefore, datatrain variable has the training set which has 548 rows and setdata has test dataset which has 229 rows.

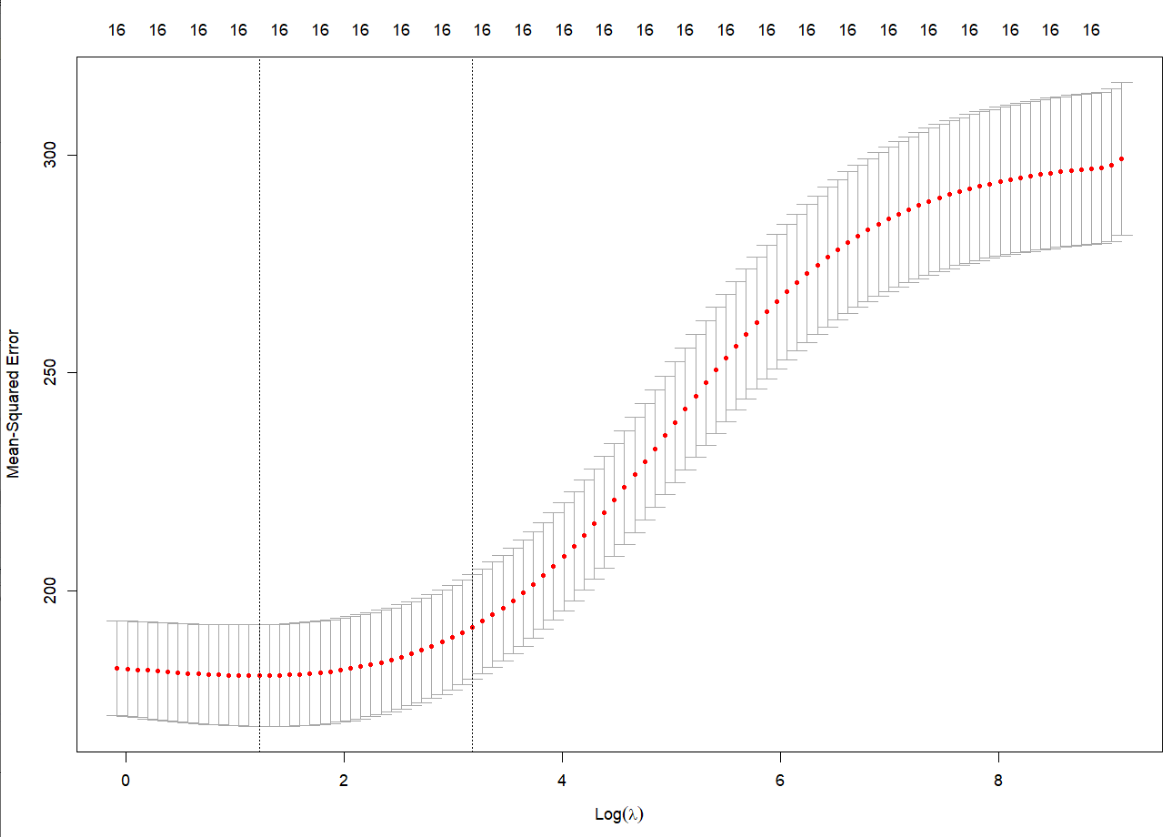
**Estimating lambda for ridge model.**

* Lambda is actually the penalty term which is responsible for regularizing the model. It actually curbs down the multicollinearity int the model.
* Glmnet() function is used to for penalyzing the model. Usually for ridge alpha is set to 0 and for LASSO it is set to 1.
* Lambda.min is the value of lambda that gives a minimum mean of the cross-validation error.
* Lambda. 1se is the maximum value of lambda that produces an error that is one standard error above the minimum.

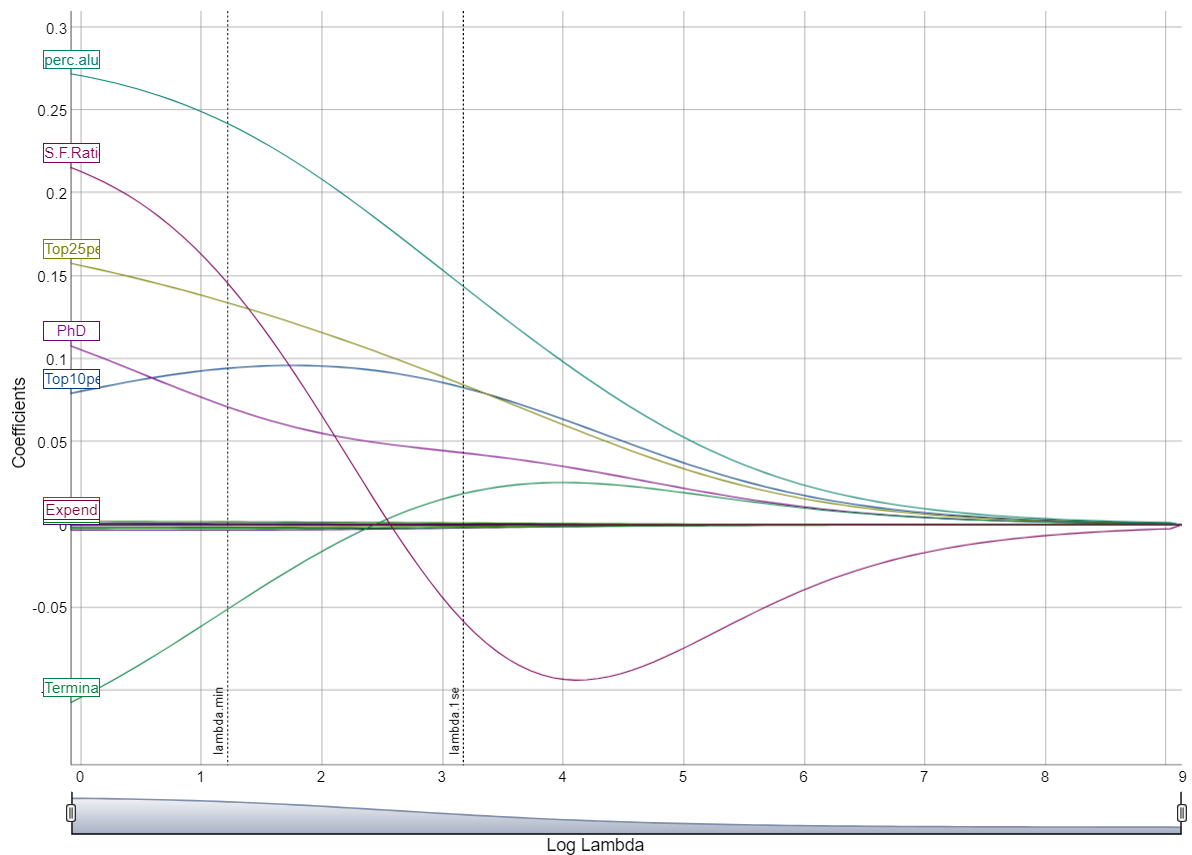


* In the figure, it is seen that we lambda min is 3.402 and lambda.1se is 24.002
* As alpha is set 0, it is considered that we have performed ridge regularization.

**Visualization with respect to lambda and mean square error.**

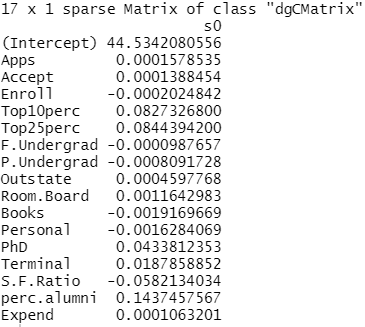


* The cross-validation curve (red dotted line) is presented alongside the higher and lower standard deviation curves throughout the series (error bars).
* Two unusual values in the series are shown by the vertical dotted lines.
* The Mean-Squared Error is displayed along the y axis. On the y axis are the numbers 200, 250, and 300. On the x axis, we have log (Lambda).
* On the x axis are the numerals 0, 2, 4, 6, and 8. The graph has a number line running across it.
* There are a total of twenty-five sixteens. A dotted line with lines that start flat and progressively slope up, starting at 4, appears at the start of the graph. A series of lines run across the graph, each with a red dot in the middle.

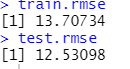


* When lambda is small, the RIDGE solution should be extremely close to the RSS , and the model should include all of your coefficients. The regularisation term has a higher influence as lambda rises, resulting in fewer variables in your model (because more and more coefficients will be zero valued).

**Applying Ridge regression on training dataset.**

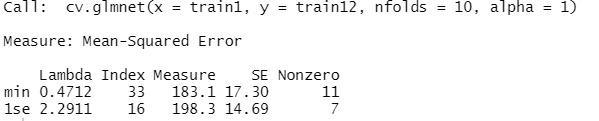


* The takings from this ridge regression are that that when graduate rate rises some of the coefficients like Apps, accept, Top10 perc, Top25 perc, Outstate, Room Board, PhD, Terminal, Expend and perc. Alumni rise as well.
* The root mean square error of an ideal model is expected to be low as lower the rmse value better the performance of the model. RMSE basically is the measure of the avg. distance between model’s projected values and dataset’s actual values.



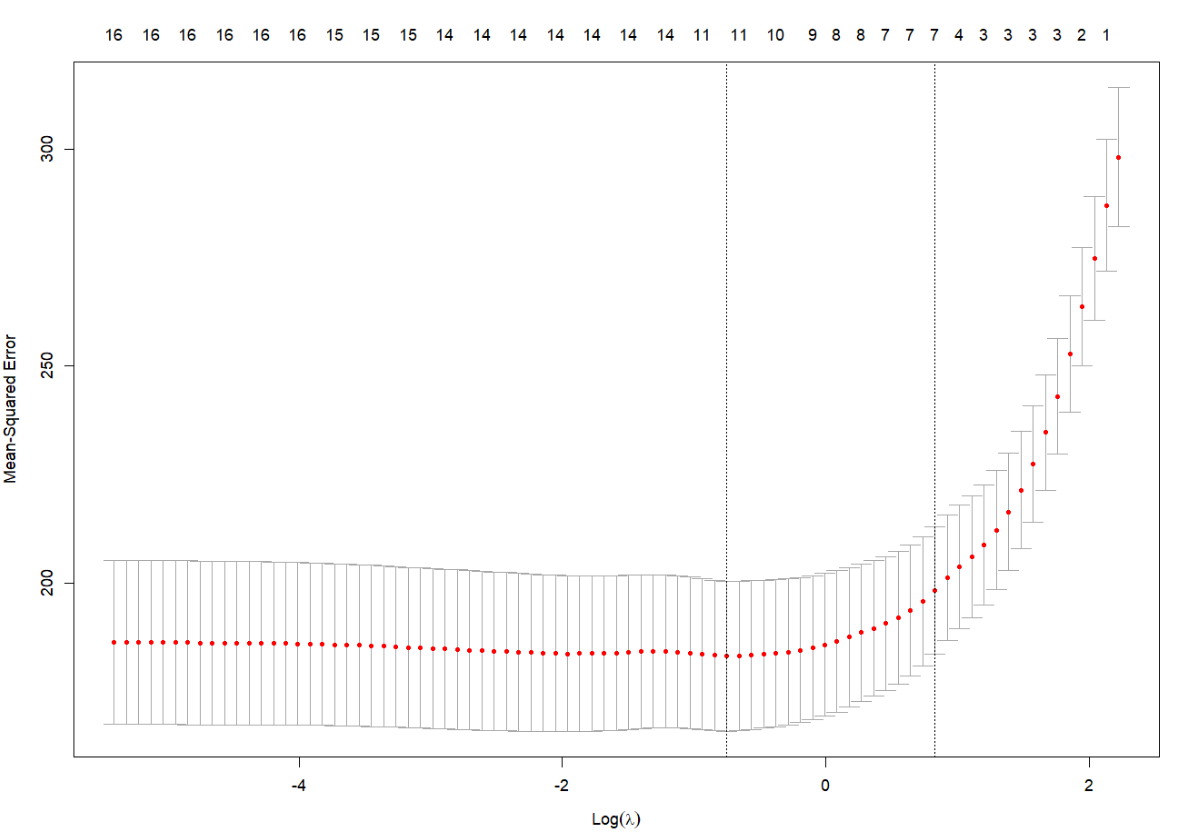
The train RMSE value is 13.707 whereas test RMSE value is 12.530. Comparing the values it is clearly seen that the model for test dataset is served better than the train dataset only because the RMSE value is lower than train RMSE value.

**Estimating lambda.in and lambda.1se with the help of cv.glmnet**

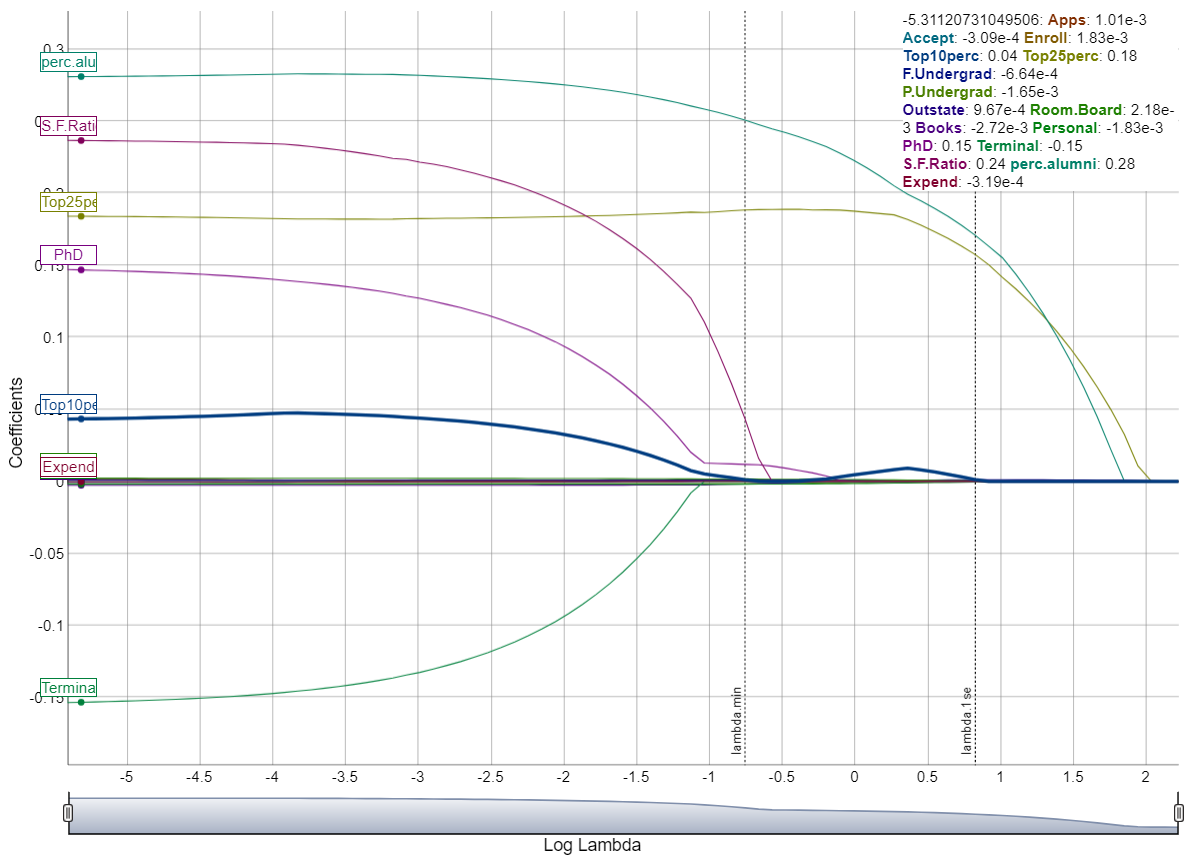


* The lambda.min and lambda.1se value mean have already been discussed.
* We also know that when alpha is 1, R recognises that the model wishes to utilise Lasso as the regularisation approach.
* The value of lambda.min is 0.4712, whereas lambda.1se 2.2911, as seen in the figure. Because lambda.min is closer to zero, we use it to fit our lasso model.

**Visualization of LASSO**



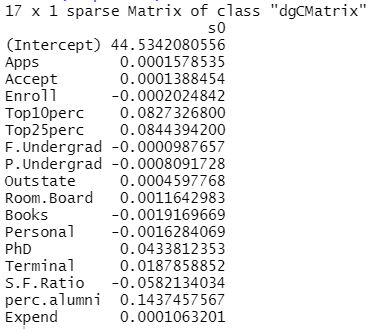
* The cross-validation curve for lasso regression is shown in the graph above. We have reviewed how to understand the graph in question 3.
* As can be seen from the graph, the value of MSE grows with the magnitude of lambda.
* Although, until the log of lambda is 0, the value of MSE remains constant. After then, it progressively rises until it reaches a maximum of about 300 values for MSE.
* The numbers at the top of the graph reflect the number of non-zero regression coefficients, which range from 0 to 16 in this example.



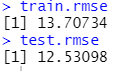
The Lasso model's coefficient route is depicted in the graph above. We know how to interpret a graph already.

* The value of the coefficients gradually decreases until lambda's logarithmic value is -2. After then, the value of the coefficients drops dramatically, and the number of variables on the top axis also changes dramatically.
* This is due to lambda.1se more coefficients are eliminated by the 1st lasso model than by lambda.min.

**Applying LASSO regression on training dataset.**



* When the Lasso regression is fitted, the graduate rate rises. Top10percent, Top25percent, Outstate, Room.Board, and other considerations perc.alumni is also increasing. We can see that there are numerous  coefficients reach 0, they are either positive or negative like Outstate, Room.Board, perc.alumni, Top10perc, Top25perc, Personal and P.Undergrad.
* The rmse (Root Mean Square Error) is used to evaluate the model's performance. average square error). The average separation between the two, the difference between the model's anticipated values and the dataset's actual values is Rmse is used to calculate the value. The better the model's fit, the lower the rmse number, the better.



* The rmse value for the train dataset is 13.70734, whereas the rmse value for the test dataset is 12.53098. The test dataset has a lower rmse value than the train dataset, indicating that it will fit better.

**Conclusion with which model is better.**

* We know there is no overfitting or underfitting by either model after running ridge and lasso on our training and test datasets.
* Ridge regression, on the other hand, was able to provide a lower RMSE value than lasso regression.
* Although the value was only changed by one, it has the potential to impact the model's accuracy.
* Because it functions as a feature selection model, lasso regression is usually preferable to ridge regression.
* However, in the case of correlated features, ridge regression may perform better than Lasso since it employs all of the characteristics, but the coefficients are distributed according to the correlation.