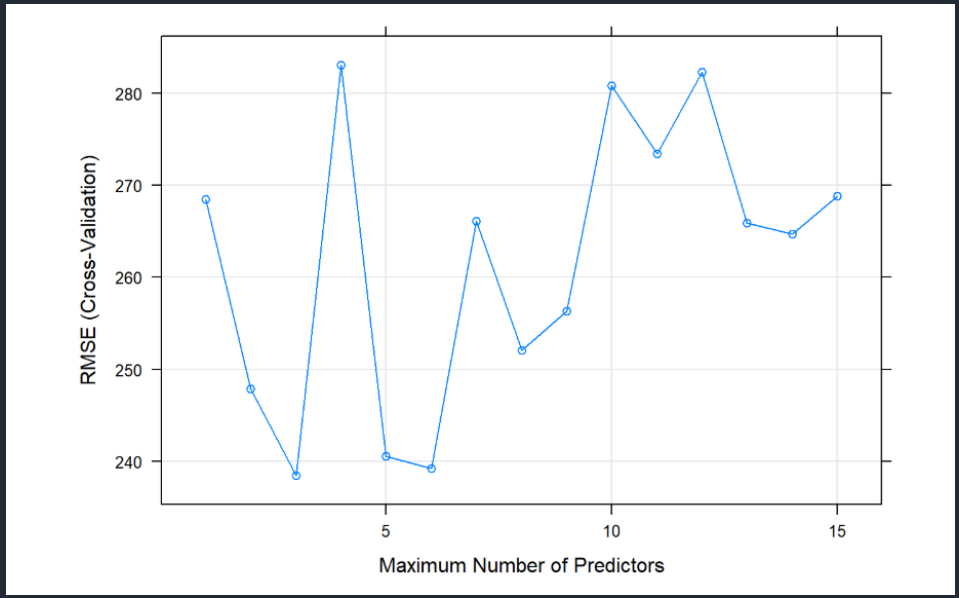
HW-5

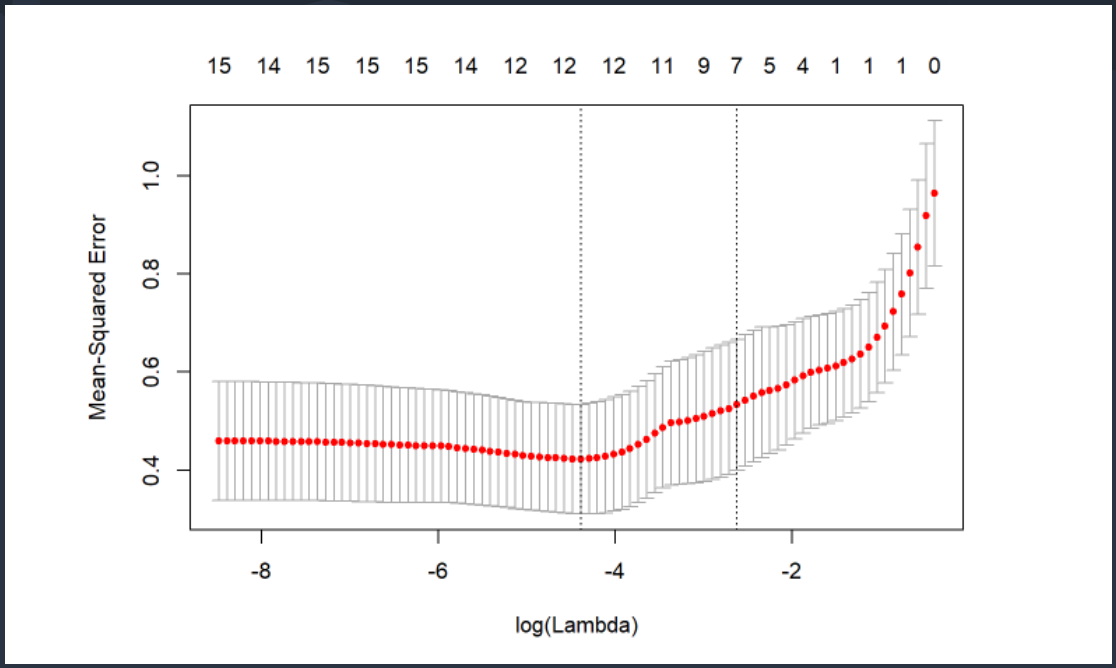
Responses & Summary

Question 11.1

- Created the Stepwise Regression model
  - Predictors were used to create a Stepwise based model
  - Predictors used: M + Po1 + U2 + Ineq + Prob
  - R-Square of new model is 0.65



- 2nd Lasso Model was created
  - Using glmnet
  - After the model was created, used the optimal Lambda (lambda.1se) to identify the predictors
  - Predictors used: M + Ed + Po1 + M.F + NW + Ineq + Prob
  - R-Square of Lasso model: 0.7517

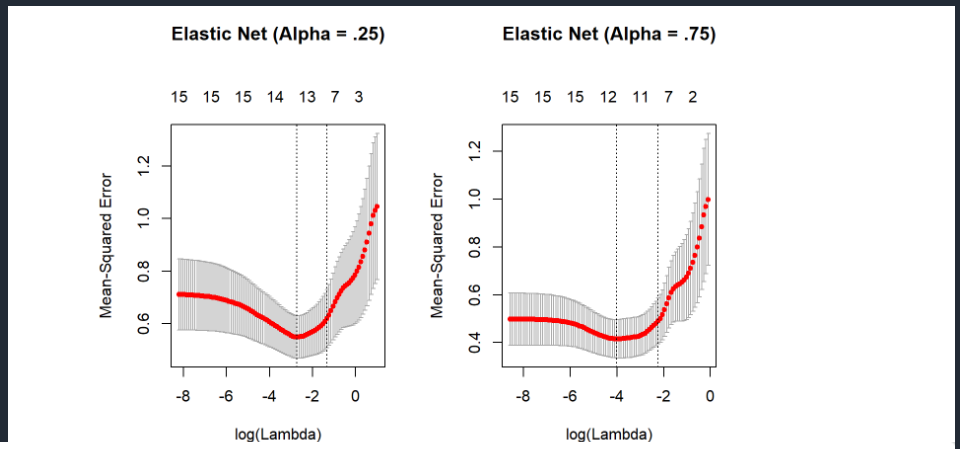


Complete Repository available on GitHub:

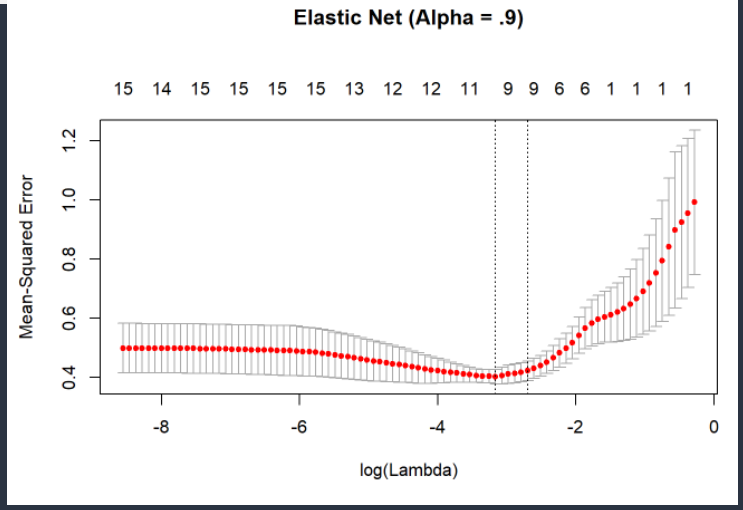
https://github.com/Hizzyth/GTX_Introduction-to-Analytics-Modelling

Question 11.1

- 3rd Elastic Net Model
  - Initially I experimented with alpha value of .25 and .75
  - Upon visualizing the impact, I decided to find the optimized value of alpha
  - For same tuning grid was designed and for loop used to find the lambda & MSE behavior with regards to changing alpha values.
  - Based on that alpha value of 0.9 was used to create the final elastic net based model
  - Final Predictors used: M + Po1 + M.F + NW + Ineq + Prob
  - R-Square value : 0.715



alpha <dbl>	mse_min <dbl>	mse_1se <dbl>	lambda_min <dbl>	lambda_1se <dbl>
0.0	0.4535684	0.5358612	0.10831507	0.69625717
0.1	0.4440554	0.5278638	0.07821224	0.34652883
0.2	0.4392866	0.5293012	0.06226800	0.25137700
0.3	0.4369399	0.5168124	0.04555937	0.18392397
0.4	0.4356648	0.5243523	0.03750102	0.16615280
0.5	0.4342793	0.5224660	0.03292585	0.14588200
0.6	0.4319486	0.5235782	0.02743821	0.13342109
0.7	0.4285387	0.5136500	0.02581149	0.11436094
0.8	0.4258924	0.5044031	0.02258505	0.10006582
0.9	0.4238956	0.5055681	0.02653876	0.09761966



Complete Repository available on GitHub:

https://github.com/Hizzyth/GTX_Introduction-to-Analytics-Modelling

Question 11.1

- Conclusion:

Number of predictors dropped again from Lasso to Elastic Net, also there is drop in R-squared value from .75 to .71. If compared back to Step regression, Elastic net did a better job of picking same number of predictors (6) but the ones with better R-Squared performance.

Overall for this dataset Lasso method proves to be robust.

Complete Repository available on GitHub:

https://github.com/Hizzyth/GTX_Introduction-to-Analytics-Modelling

QUESTION 12.1 : Design of Experiments

Question: Describe a situation or problem from your job, everyday life, current events, etc., for which a design of experiments approach would be appropriate

Answer:

I am a beginner Golfer, and one of the challenges I face is selection of Irons/clubs while playing the game based on distance to hit. If I continue to rely on data points from each of my game, it will take too long for me to get it right.

Instead of that I can design an experiment to get some data points by using all the irons on a golf range, which will serve as a baseline and predict based on my playing frequency/consistency how much distance each of the clubs will give me.

QUESTION 12.2 : Fractional Factorial Design

- Designed the Factorial fractional design using FrF2 library
- Assigned 10 features to 16 fictious houses

https://github.com/Hizzyth/GTX_Introduction-to-Analytics-Modelling

```
FFD<- FrF2(16,10, factor.names = list(Yard ="", Solar_Roof="",Italian_Kitchen ="", Golf_Course ="", Pool ="", Neighborhood=
"", Schools="",Gated="",HOA="",Parks=""))

FFD
```

	Yard <fctr>	Solar_Roof <fctr>	Italian_Kitchen <fctr>	Golf_Course <fctr>	Pool <fctr>	Neighborhood <fctr>	Schools <fctr>	Gated <fctr>	HOA <fctr>
1	1	1	1	-1	1	1	1	-1	-1
2	1	1	-1	-1	1	-1	-1	-1	1
3	1	1	-1	1	1	-1	-1	1	-1
4	-1	-1	-1	1	1	1	1	-1	1
5	-1	1	1	1	-1	-1	1	-1	1
6	1	1	1	1	1	1	1	1	1
7	-1	1	-1	-1	-1	1	-1	1	1
8	-1	-1	1	1	1	-1	-1	-1	-1
9	-1	1	1	-1	-1	-1	1	1	-1
10	-1	-1	1	-1	1	-1	-1	1	1

1-10 of 16 rows | 1-10 of 11 columns Previous 1 2 Next

	Yard <fctr>	Solar_Roof <fctr>	Italian_Kitchen <fctr>	Golf_Course <fctr>	Pool <fctr>	Neighborhood <fctr>	Schools <fctr>	Gated <fctr>	HOA <fctr>
11	1	-1	-1	1	-1	-1	1	1	1
12	-1	-1	-1	-1	1	1	1	1	-1
13	1	-1	-1	-1	-1	-1	1	-1	-1
14	1	-1	1	1	-1	1	-1	1	-1
15	-1	1	-1	1	-1	1	-1	-1	-1
16	1	-1	1	-1	-1	1	-1	-1	1

11-16 of 16 rows | 1-10 of 11 columns Previous 1 2 Next

QUESTION 13.1 : Probability Distributions

Question: For each of the following distributions, give an example of data that you would expect to follow this distribution (besides the examples already discussed in class).

Answer:

a. Binomial : College acceptance rates

b. Geometric : New antivirus definitions rolled out, number of days the virus can keep up with cyberattacks before it needs another rollout.

c. Poisson : Amount of rainfall expected over weekend

d. Exponential : Change in Temperature trends over last 3 years

e. Weibull : Time before I need to replace the battery of my laptop