A decorative pattern on the left side of the slide, consisting of a grid of overlapping circles. Each circle is formed by two intersecting arcs, creating a series of lens-like shapes that fill the area.

MACHINE LEARNING METHODS REGRESSION ANALYSIS ON HEDGE FUND X: FINANCIAL MODELING CHALLENGE DATA

Lily (Lizheng) Zhou

https://github.com/LilyLizhengZhou/Project_StatisticalLearning_RegressionAnalysis

<https://vimeo.com/420817512>

INTRODUCTION

Project Description

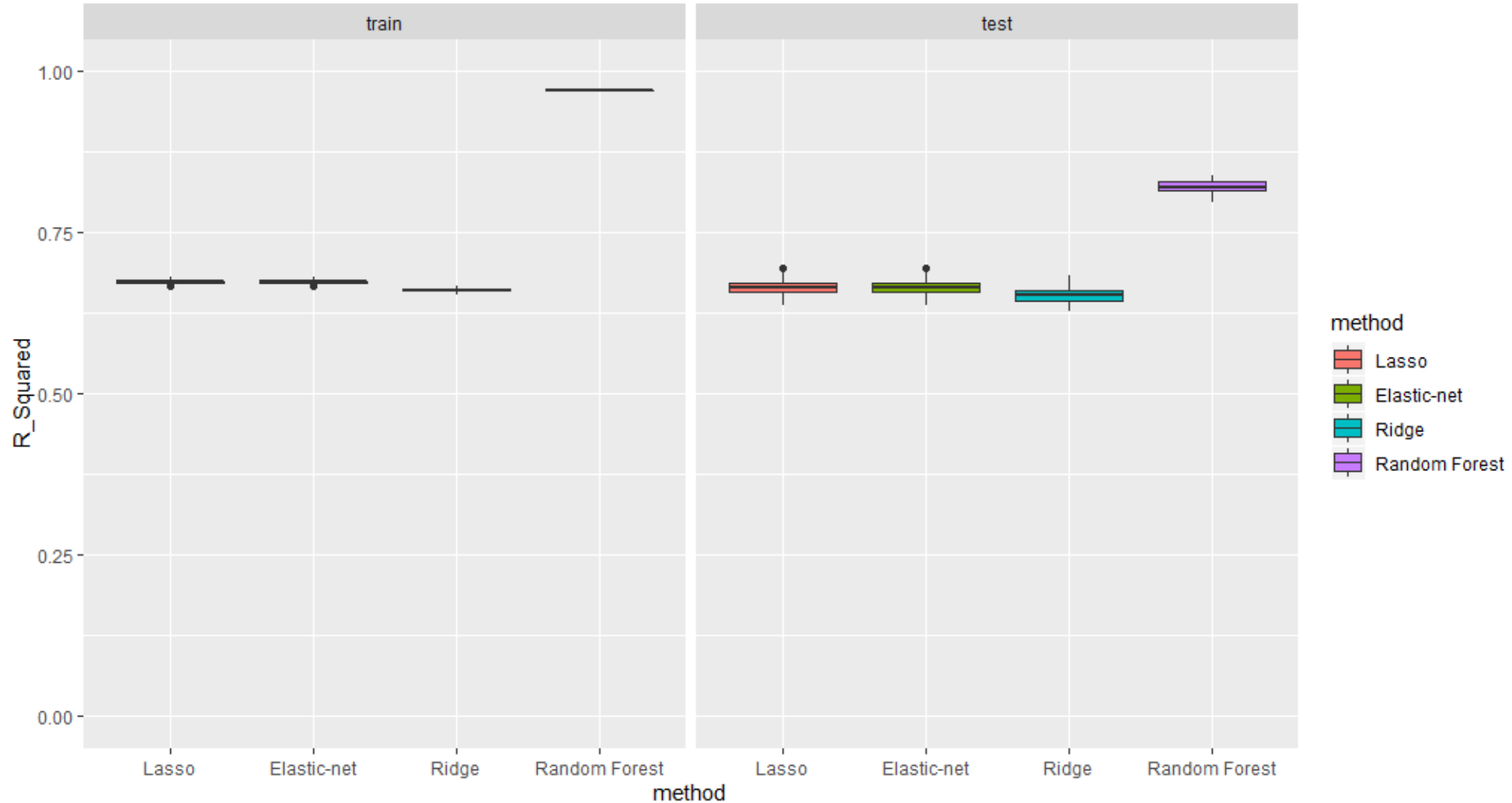
- This project is aimed to use regression analysis to predict a numeric financial response variable based on 88 predictors with 4 methods: Lasso, Elastic-net ($\alpha = 0.5$), Ridge and Random Forrest.

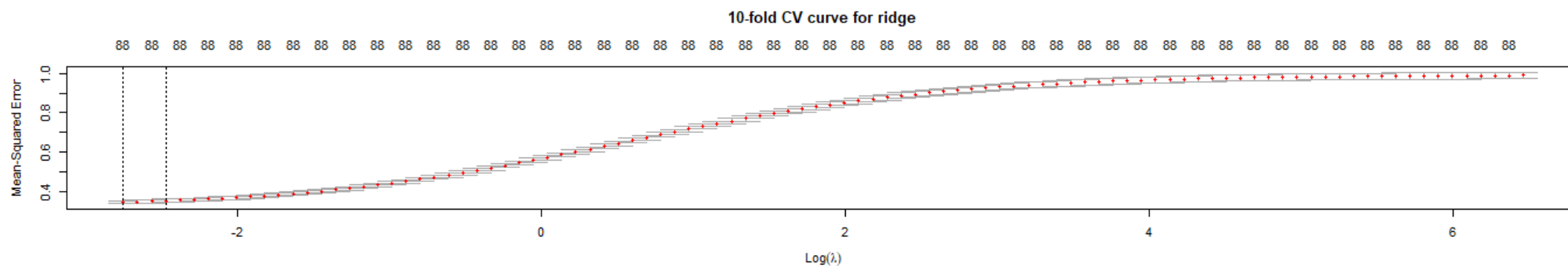
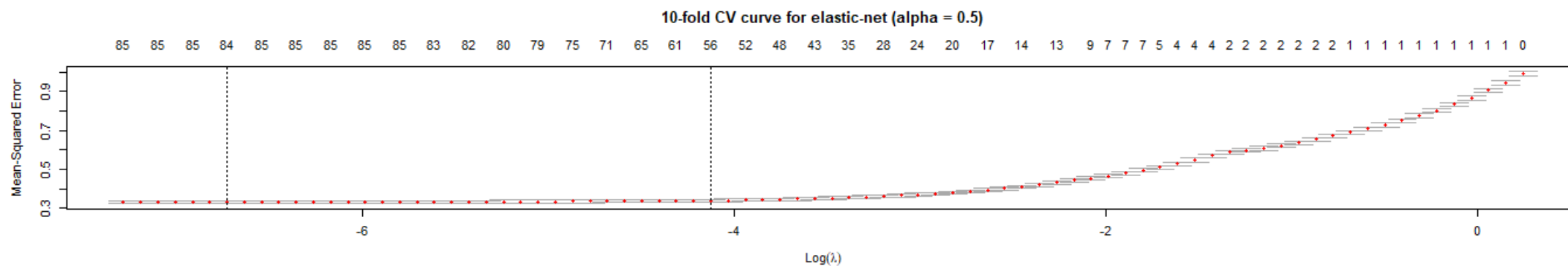
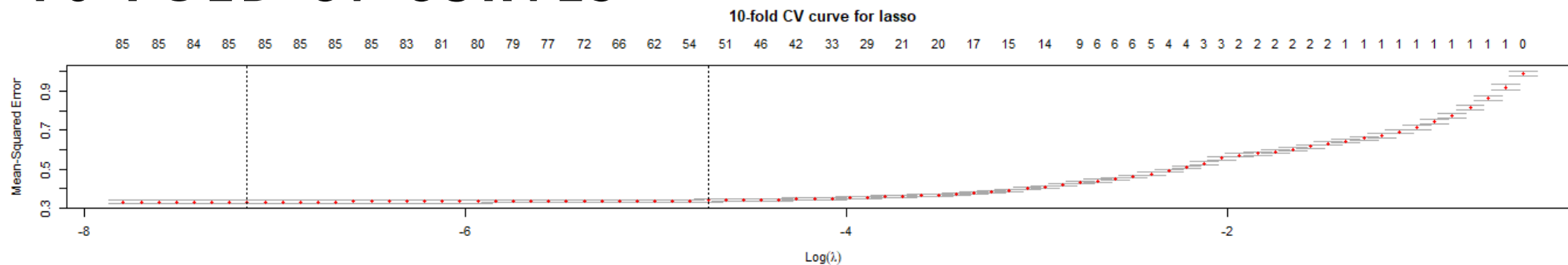
Data Description

- This dataset is a sample of the training dataset used in the DeepAnalytics competition, Hedge Fund X: Financial Modeling Challenge (<https://deepanalytics.jp/compe/53>).
- Data Set Structure: ($n = 10000$, $p = 88$)
 - Response variable (named as y):
 - $c1$: numeric
 - Predictors (named as 1 - 88):
 - $c2 - c88$: numeric
 - target: categorical (with levels: 0 and 1)

BOXPLOTS OF R-SQUARED TRAIN AND TEST

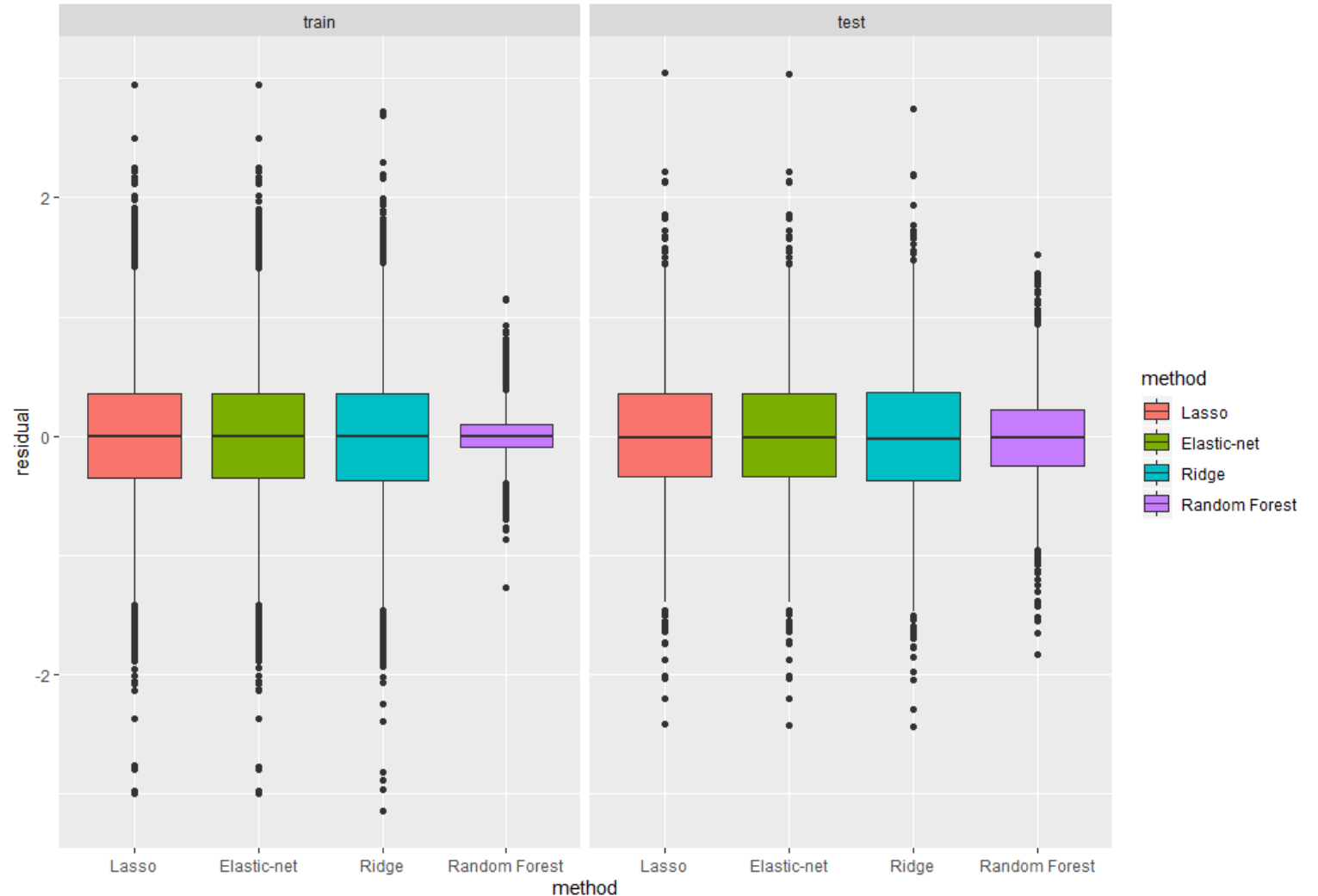
Boxplots of R-Squared Train and Test with 4 Methods (train size = $0.8n$, 100 samples)





BOXPLOTS OF TRAIN AND TEST RESIDUALS

Boxplots of Train and Test Residuals with 4 Methods)



All 4 methods' mean of residuals are very close to zero



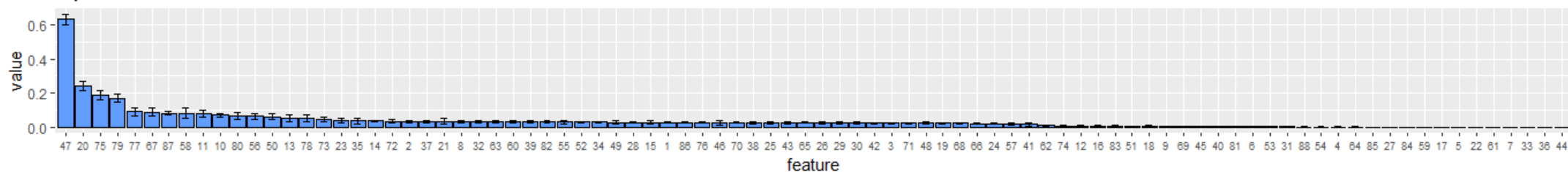
Lasso, Elastic-net and Ridge, their residual variance are also very close



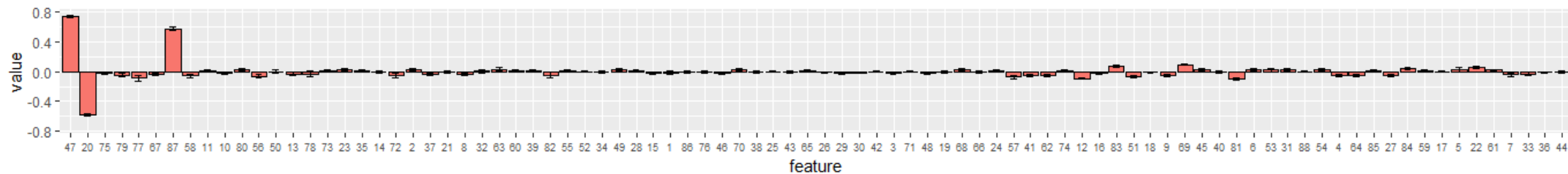
Random Forest has smaller variance compared to other methods; its train variance is smaller than test residuals

VARIABLE IMPORTANCE

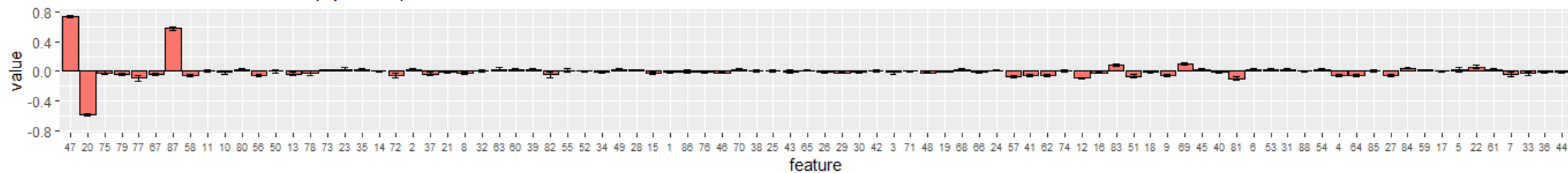
Importance of Parameters - RF



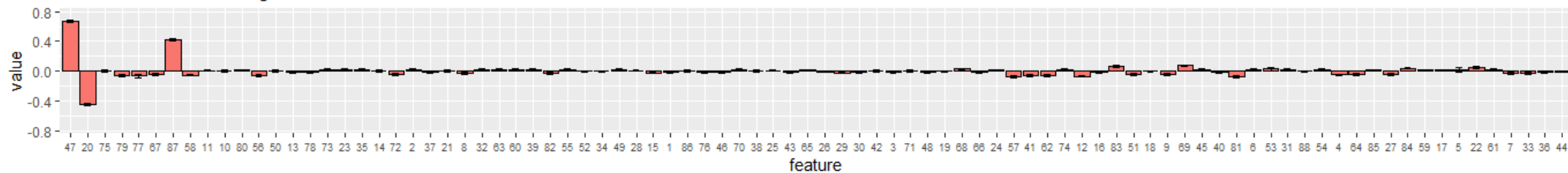
Estimated Coefficients - Lasso



Estimated Coefficients - Elastic-net (alpha = 0.5)



Estimated Coefficients - Ridge



PERFORMANCE VS. TIME

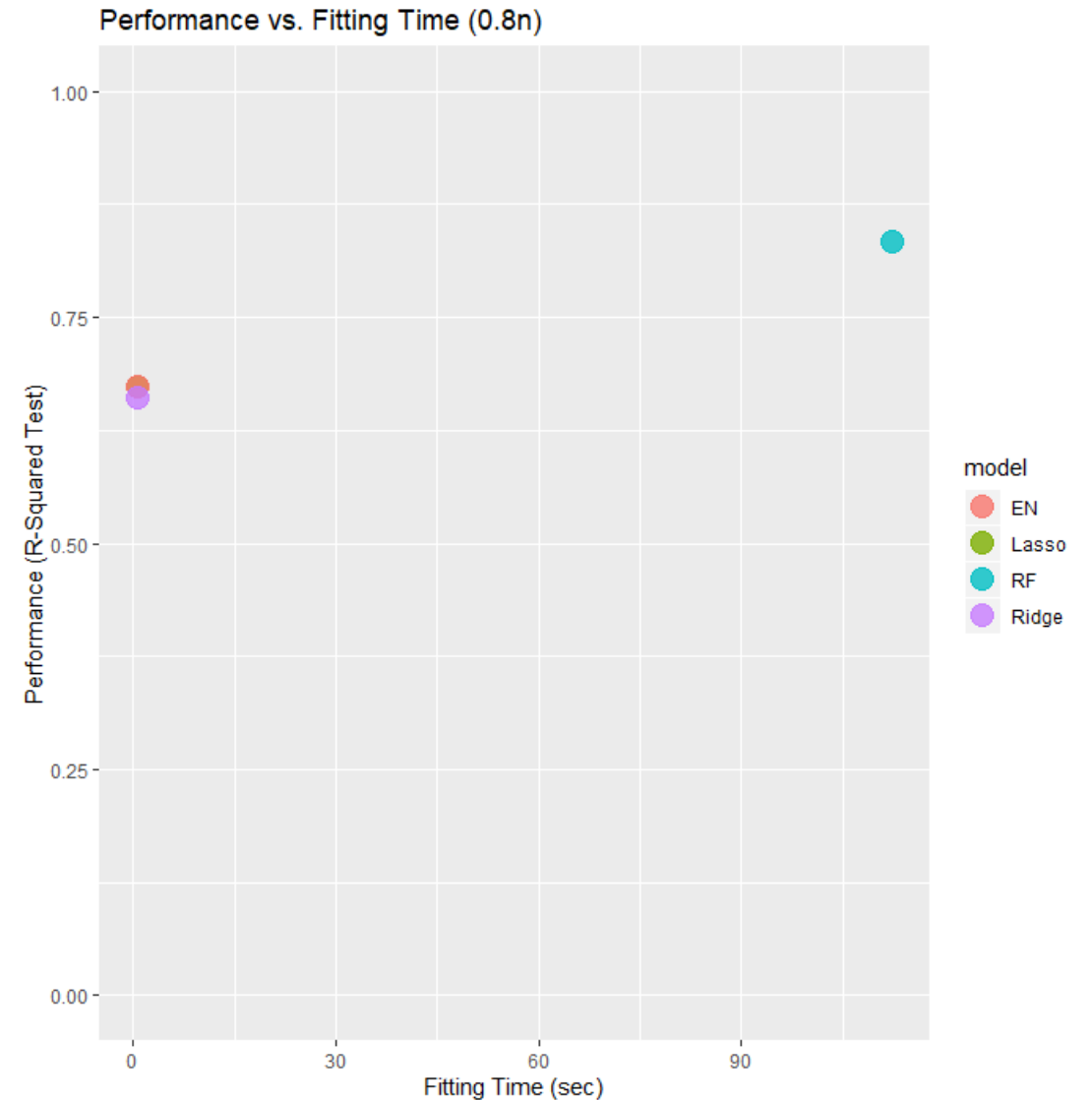
Model	Model Fitting Time (sec)	Model Performance (R-Squared Test)
Lasso	0.66	0.6731127
Elastic-net	0.67	0.6731972
Ridge	0.70	0.6616434
Random Forest	112.20	0.8332545

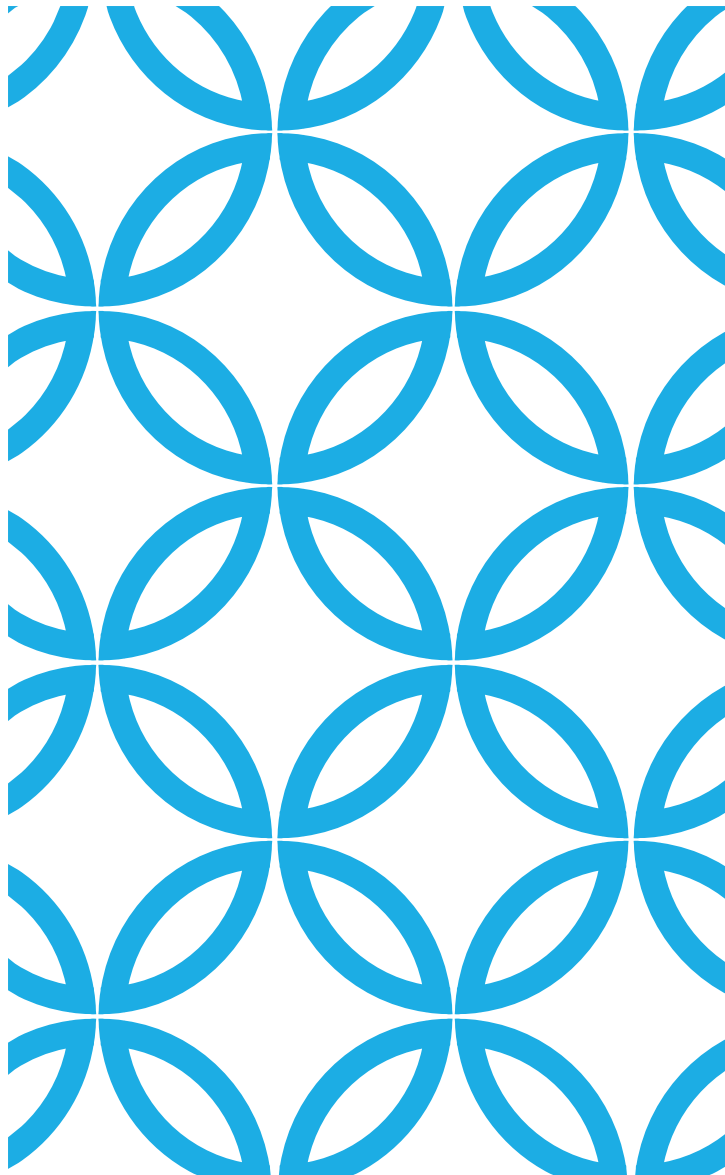


Lasso, Elastic-net and Ridge, their performance and fitting time are very similar



Random Forest performs decent at the cost of high time complexity





THANK YOU
