

Caravan Insurance

Predicting which customers will
buy Caravan Insurance

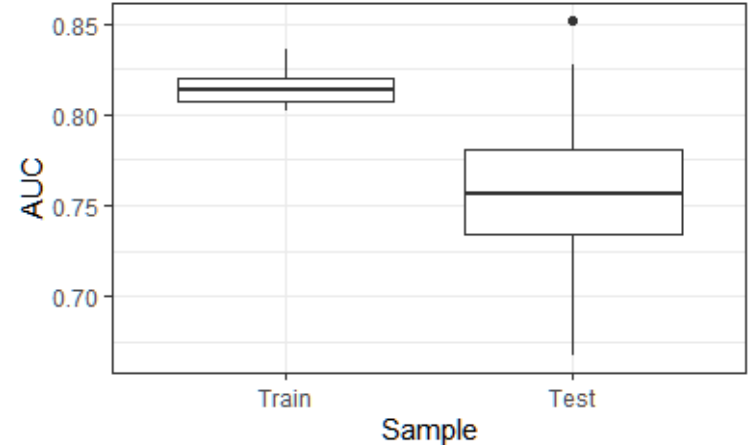


Data Description

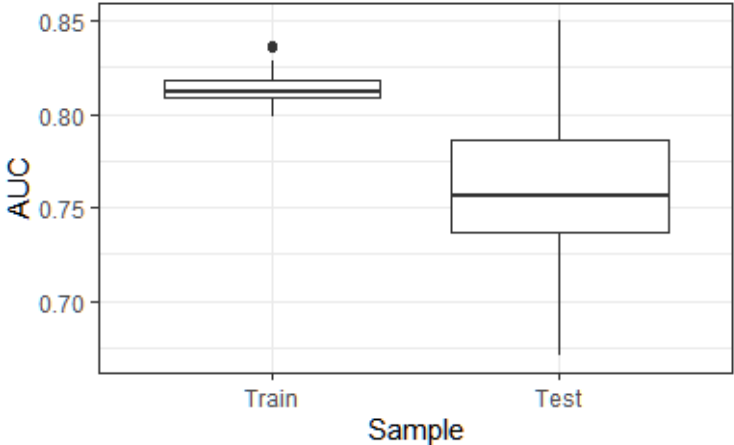
- $N = 5,822$
- $P = 85$
- Total Positive Cases: 348
- N:P Imbalance = 15.73:1
- 2 Feature Types
 - Socio-Economic
 - Product Usage
- Features are multinomial
 - # of Levels range between 2 & 40

AUC Box Plots

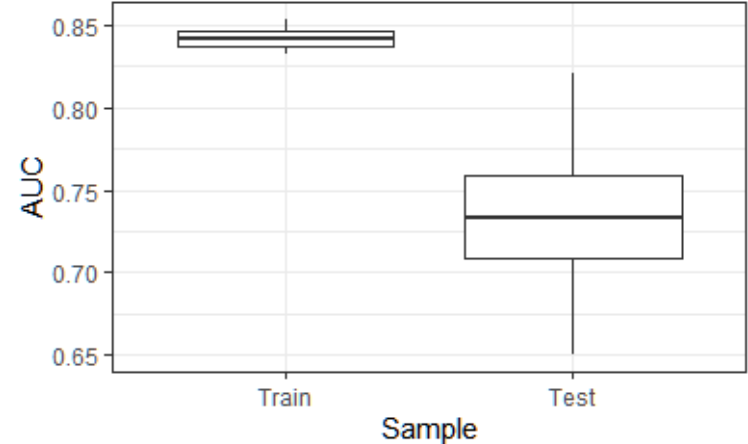
Elastic Net AUC Boxplots



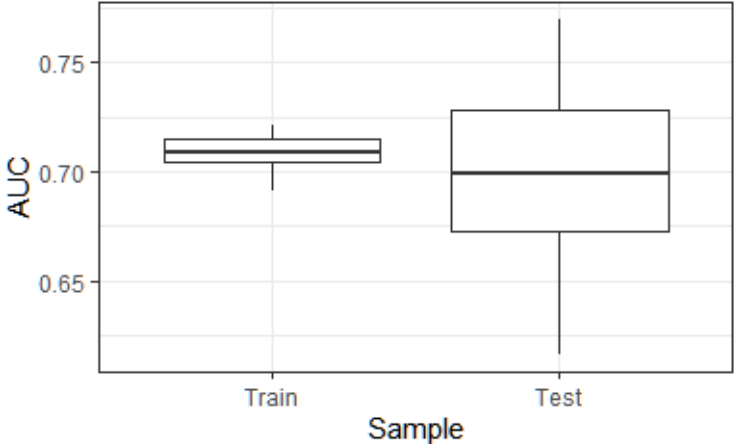
Lasso AUC Boxplots



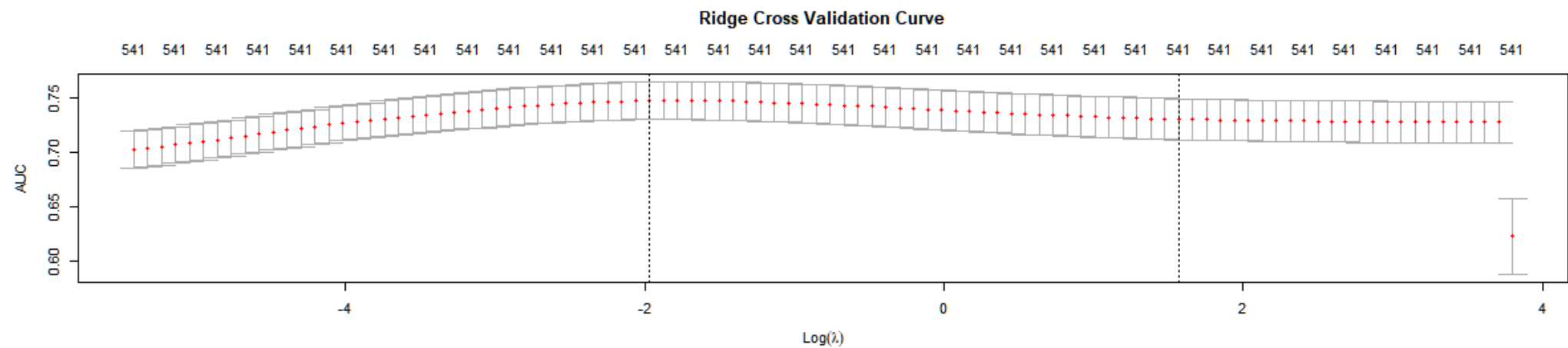
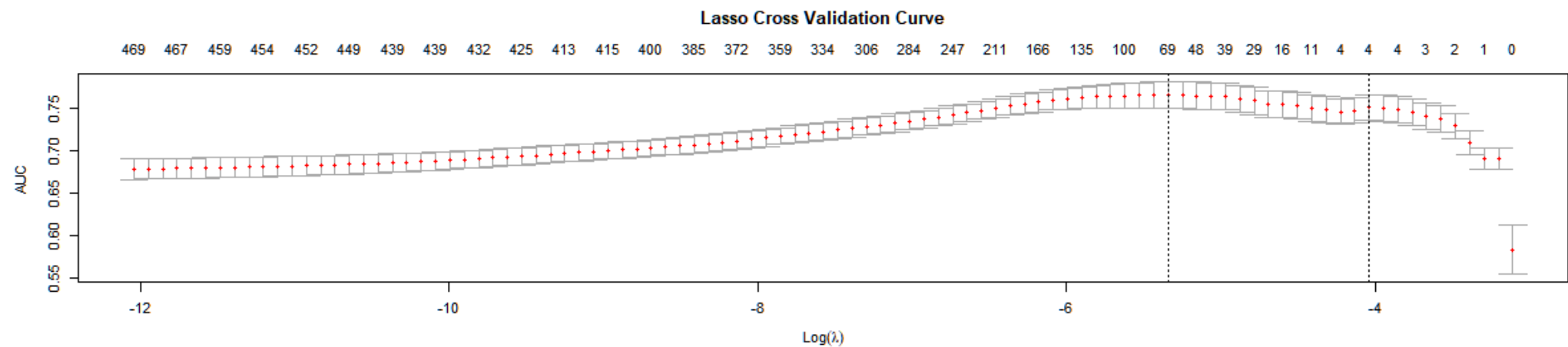
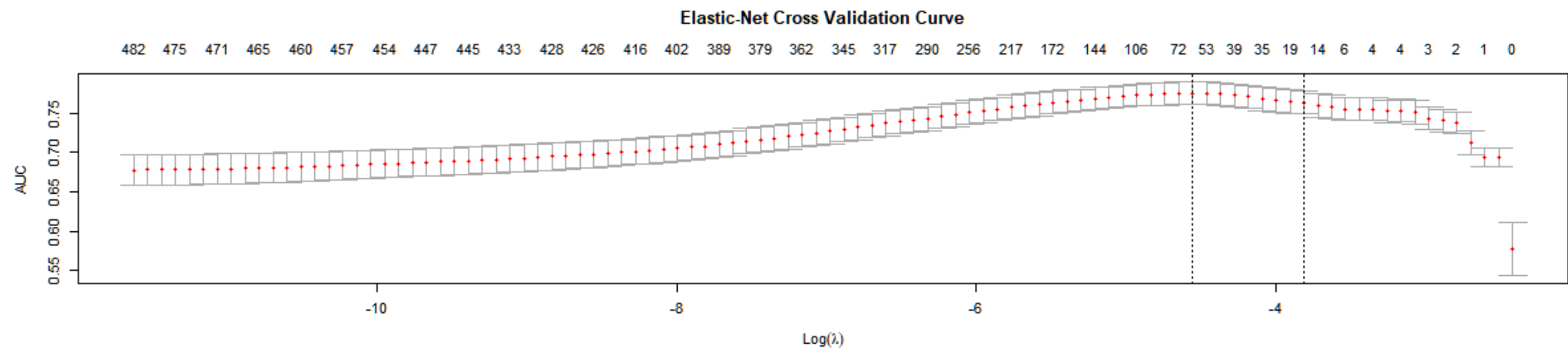
Ridge AUC Boxplots



Random Forest AUC Boxplots



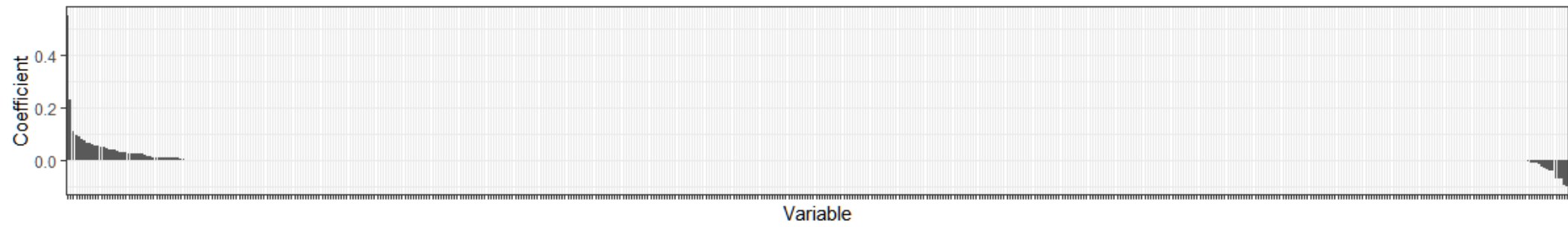
Method	Train	Test
Elastic Net	0.8153	0.7595
Lasso	0.8139	0.7595
Ridge	0.8421	0.7347
Random Forest	0.7092	0.6993



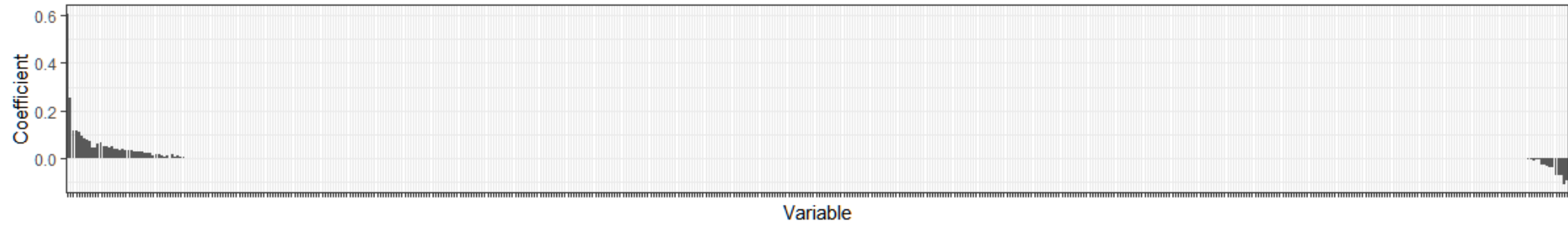
Test AUC vs. All Data Train Time

Method	AUC	Time
Elastic Net	0.7562556	97.27
Lasso	0.7565090	99.15
Ridge	0.7335237	39.02
Random Forest	0.6994190	78.76

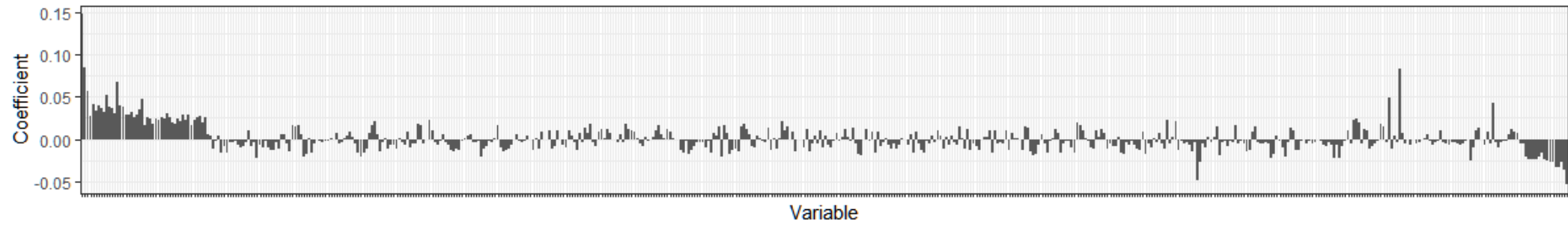
Standardized Elastic Net Coefficients



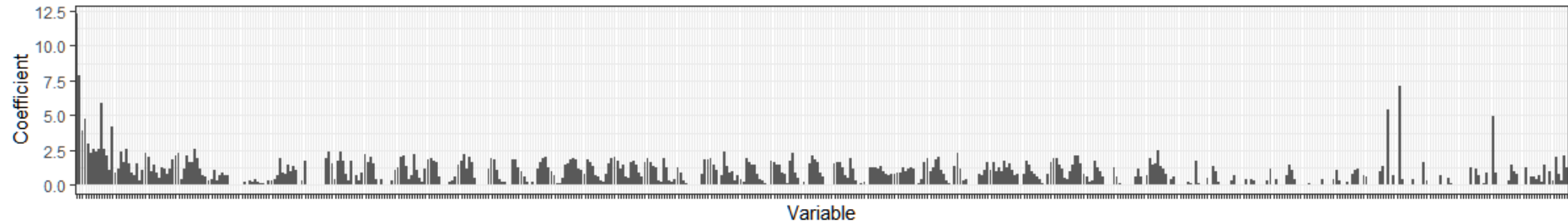
Standardized Lasso Coefficients



Standardized Ridge Coefficients



Random Forrest Variable Importance



Concluding Remarks

- Most Important Variables: High Level Contribution to Car & Fire Insurance, High Level Purchasing Power Class, Ownership of Boat Insurance
- Parsimony is vital
- Cost Benefit Analysis

https://github.com/escarrion/9891_Project_Fall_21