STA9891 – Final Project

A binary classification of Diabetes Patients and their Hospital Outcomes

By Brian Contreras and Troy Whittemore

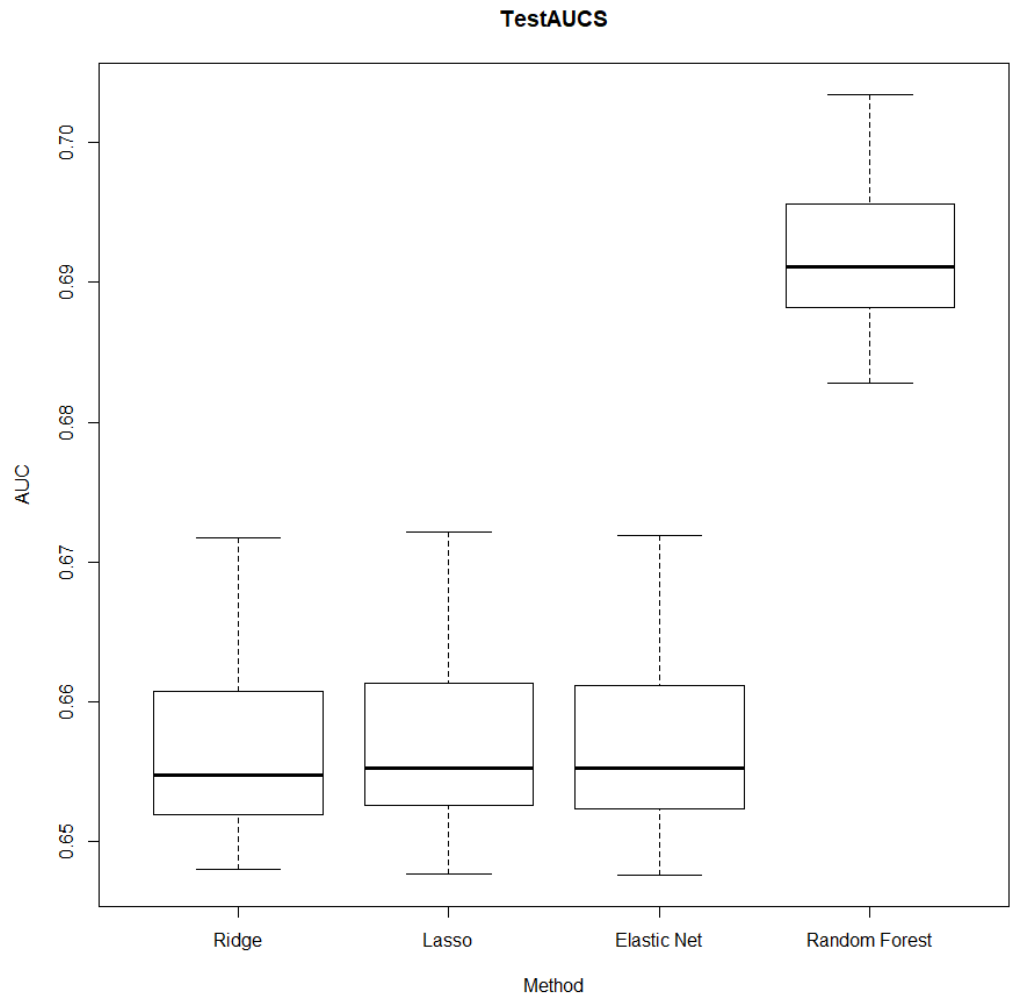
Data Overview

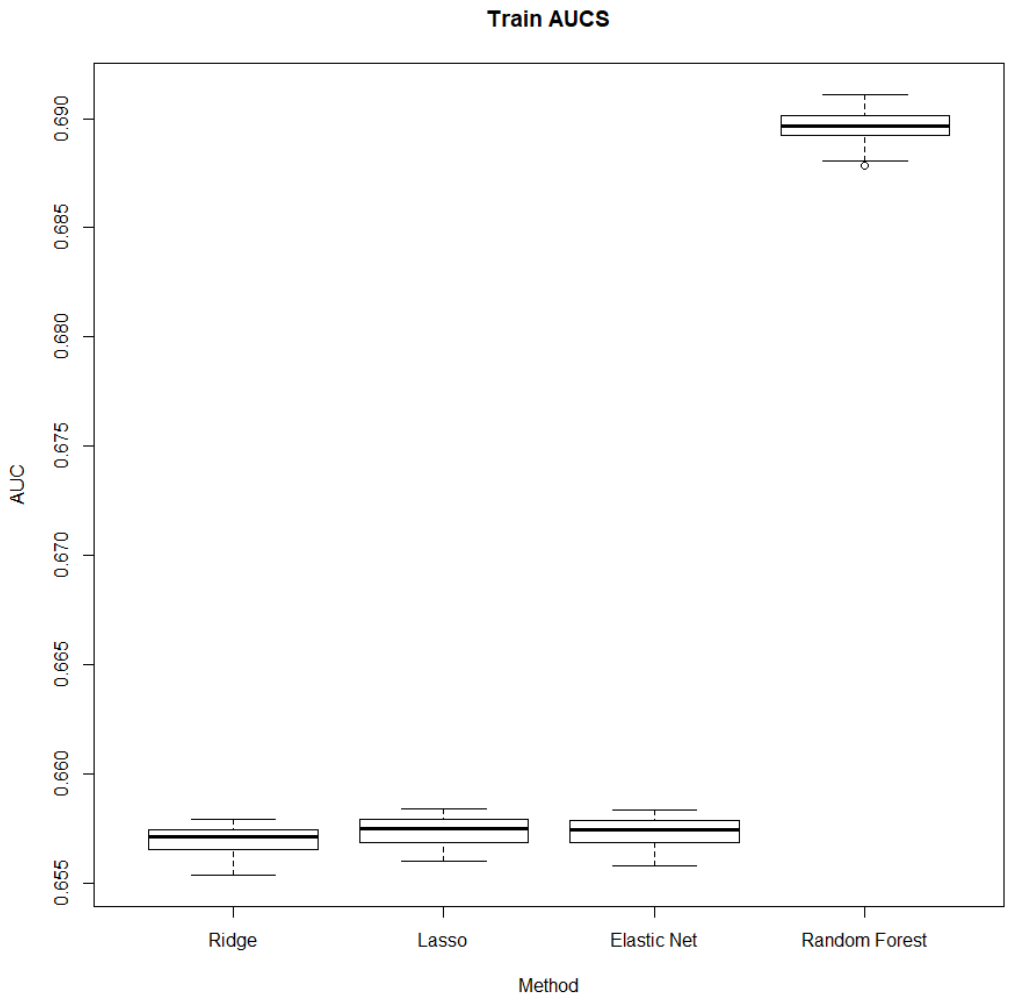
- This data covers diabetic, inpatient encounters in hospitals with length of stays between 1 and 14 days.
- The importance of the analysis of this data is to understand the interaction between the variables within hospitals and the readmission rates of the patients.
- n = 98,053 p = 54

  n+ = 41,175 instances of readmission, n- = 47,073 instances of no readmission

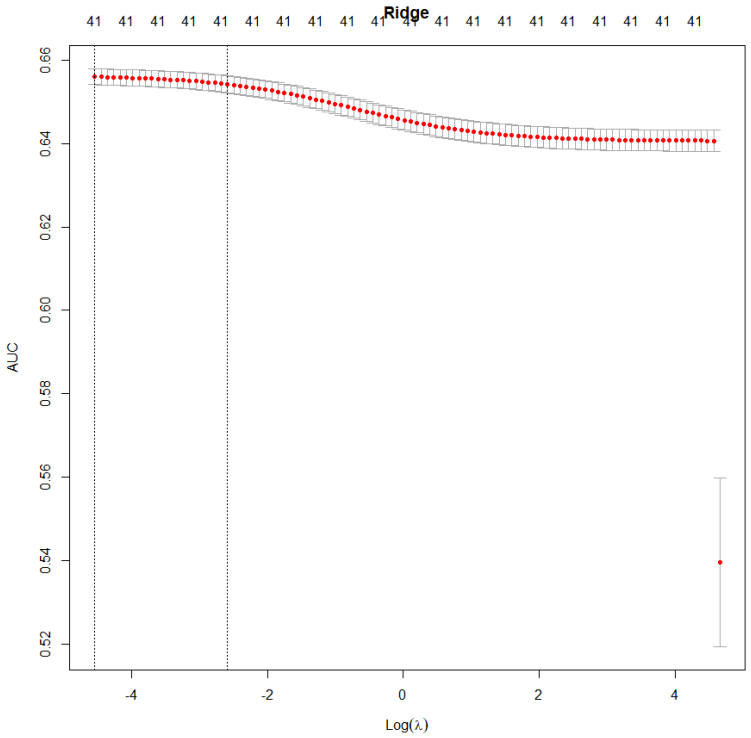
  features include; age, race, gender, time in hospital, medical information, results of different diagnostic tests, medications, and more.

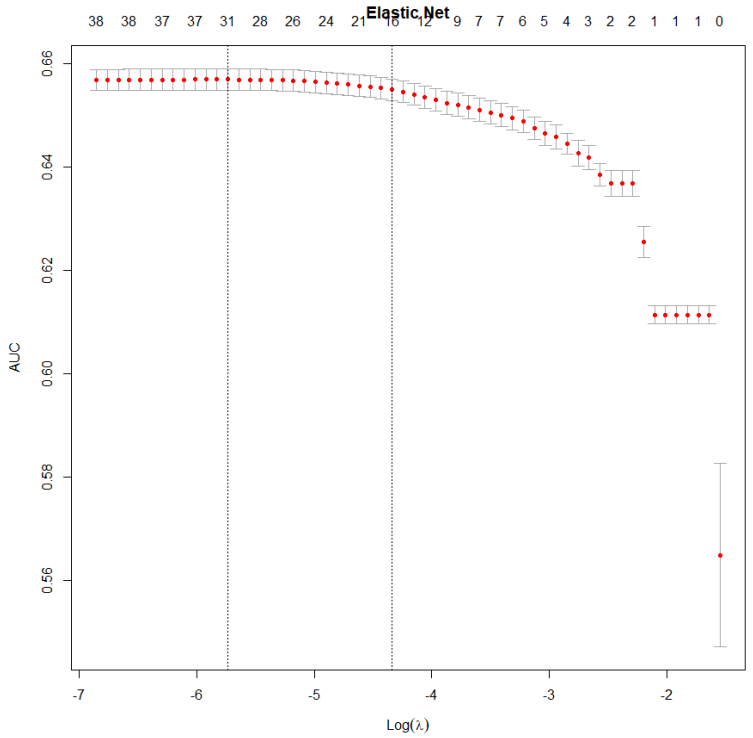
AUC Boxplots

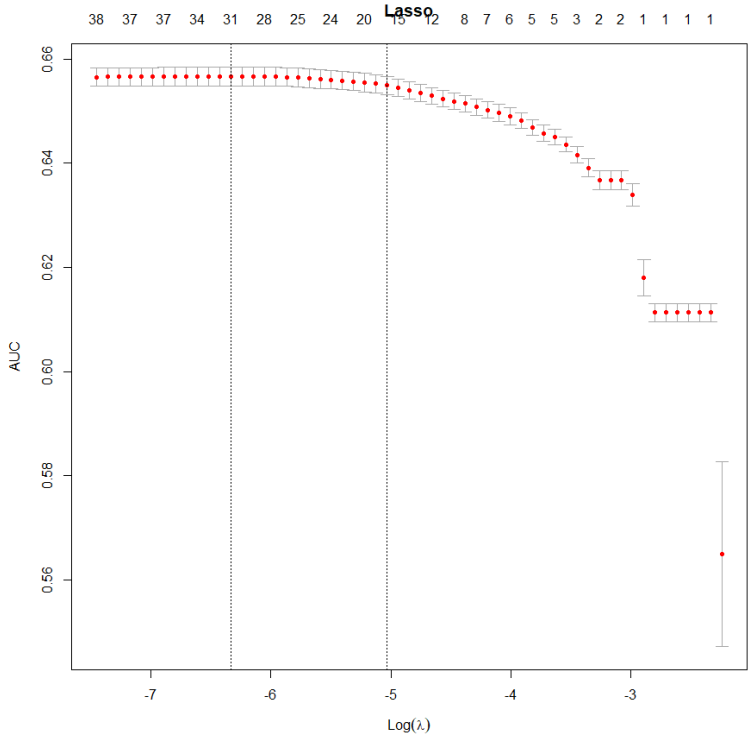




10-fold CV curves and Time







- Time = 83.67 seconds

38.19 Seconds

37.55 Seconds

Time to complete regressions

Method	Test AUC	Time (seconds)
Ridge	0.654728	86.52
Lasso	0.6552557	38.67
Elastic Net	0.6552481	39.97
Random Forest	0.6911152	575.43

- Trade off between time and performance?
- Lasso and Elastic Net both performed significantly better than Ridge on this data set.
- Random Forest did have the best Test AUC at the cost of time.

Importance of Parameters

Concluding Remarks