Assignment

For this data you will use the dataset “MI” (MI.Rdata). The dataset contains data about patients that enter hospital with a myocardial infarction. The dataset includes information on the patient at the acceptance in the hospital after the first evaluation. This includes the medical history of the patient, past cardiac problems for example, as well as immediate measurements on the patient, including blood pressure and blood analysis. The response variable (LET_IS) is death of the patient during their stay. In other words, we try to predict from our initial data if the patient is at an elevated risk to die during their hospital stay.

1. Study and describe the predictor variables. Do you see any issues that are relevant for making predictions? Make sure to discuss the dimensionality of the data and the implication on fitting models.
2. Fit and compare an appropriate unconstrained linear model, as well as lasso and ridge regression models. Discuss what you find. What is an appropriate base-level of performance to compare your models to?
3. Among your top predictors, do you see evidence of non-linear effects? How could you accommodate non-linear effects and still use a regularized regression approach? Does adding non-linear effects improve your model?
4. Fit an appropriate Random Forest model. Report a comparison of performance to your linear model and explain any differences in performance. Do you see an important difference in how variables are used for predictions?