P8106 Midterm Project: Predicting COVID-19 Recovery Time

Guadalupe Antonio Lopez, Gustavo Garcia-Franceschini, Derek Lamb

UNI's: GA2612, GEG2145, DRL2168

Introduction

EDA

We also examined the pairwise correlations of the variables, and the correlations of the covariates with the recovery time. There were two clusters of strong correlation (height, weight, and BMI; hypertension and SBP), but these covariates were functionally dependent upon each other. There were no other strong correlations between variables, and no one covariate had an exceptional correlation to recovery time.

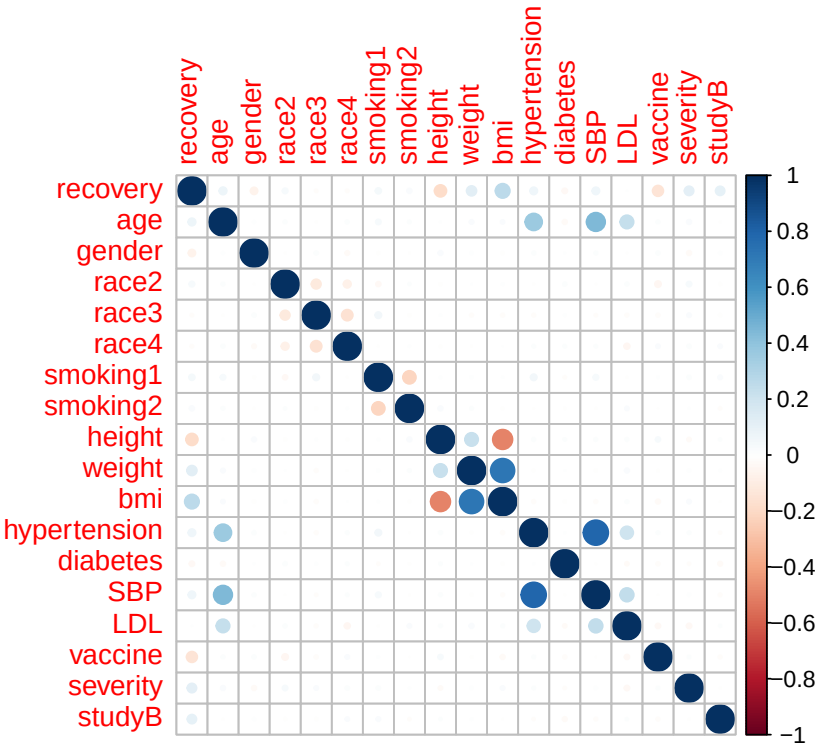

Figure M. Correlogram of study variables.

Model Training

To train the models, we partitioned the data into training and testing sets, with 80% of the data (2400 subjects) being assigned to the training set, and the remaining 20% (600 subjects) being assigned to the test set.

We decided to model the data using four approaches, two linear and two non-linear. For the linear approaches, we selected elastic net and partial least squares regression. For the nonlinear approaches, we selected multivariate adaptive regression splines (MARS) and a general additive model (GAM).

ENet

PLS

MARS

GAM

Results

Remember to talk about `study` as a variable.

Conclusion