

STAT 239 Homework 5 - Final

Tarek Tohme

June 11th, 2018

Classification Method

I fitted 6 different models on the training data: SVM with polynomial kernel degree 1, SVM with radial kernel, GBM with interaction depth 1, Random Forest with $m = 1$, LDA and K-means clustering. The parameters for each model were determined by cross-validation, which also served to estimate each model's prediction error. Then, I predicted the response with each model, and determined the final response with a weighed majority vote of all the responses. The weights were the estimates of each model's test error obtained by CV (normalized such that the response for an observation is 1 if the weighed sum of the votes for that observation is greater than 0.5).