Roberto Romero

ME 532 Progress Report

The data from the cardiovascular dataset has been converted into a .mat file for easier application in python. The data has been uploaded to the preliminary program. The matrix is full rank so a unique solution to the least square problem exist. A simple least square program was created as a starting point. An error rate of 50.03 % was achieved. An attempt to improve the accuracy of the program was to exclude the features smoking, alcohol intake, and physical activity. These features were removed because they were provided by the subject and can be subjective, since their response was a binary yes or no. This did not improve the accuracy as the error rate was still 50.03 %. From here a least squares SVD and ridge regression models were made. These did not improve the accuracy of the least square classifier either. For all models, the error rate was the same, 50.03%. None of these models are better at estimating decision than flipping a coin. This can mean that the selected features are not good indicators for predicting heart disease, there must be further manipulated to provide better results, or that the least squares approach is not ideal.

Moving forward, some more work will be done to improve the linear classifier. I am hoping to have at least one case with an error rate lower than 50.03%. I am also working on the SVM algorithm. I expect the SVM program to have lower error rates. If the SVM has error rates around 50%, a more in depth analysis of the available data will be done to determine if it is appropriate for determining heart disease.