Roberto Romero

# ME 532 Progress Report 2

From the first update I tried a support vector machine approach to the data. I did this by importing svm from sklearn in the python database. This allows me to create a support vector machine model from the data. I used the support vector classification (SVC) model. The program takes a long time to create the model, so only a few test were ran as of now. Seven different models were created using 90% of the data to train and they were tested on 10% of the data. The lowest error rate from these seven models was 0.49 and the highest was 0.515. The seven models averaged an error rate of 50.02%. This is similar to the error that I was getting from the least square solution (50.03%). I tried using the classifier learner app in MATLAB to see what results I can get. MATLAB also took a very long time to create a model, and couldn't give me any results. This leads me to believe that in order to make progress I have to manipulate the data more.

Moving forward I will reduce the amount of data points I use to see if I can figure out which features are not as important. Currently I am using 70,000 data points with 12 features. I will reduce to 10,000 data points to see if there are any features that do not influence the outcome as much. Removing one feature removes an entire column of 70,000 data points that will make computations faster. Once this is done I'll revisit my least square and svm code with the new manipulated data. After that I will create a neural network model and then compare the results of all three models. I am also planning on going to office hours this week to see if I can get any recommendations on manipulating the data.