Checkpoint 2

Tyler Trotter

I found much of the pre-processing strategies I had considered in the last checkpoint were ineffective at boosting the accuracy of the classifiers used. This is because dimension reduction for 0-vectors, as I had suggested, doesn't improve the linear separability as I had tried to run this for the Perceptron which tests for linear separability of the data. However, for this update, I used a Decision Tree algorithm and dimension reduction has the effect of improving accuracy and computational efficiency. I found this pre-processing to be interesting because it shows that pre-processing is an exercise that can be specified to the dataset used, the algorithm used and the task it is trying to accomplish. Though I was hoping the dimension reduction strategy would be useful generally, given how sparse the data is, it provided the aforementioned interesting insight into how model training and prediction works.

As I forge onward to prepare for the conclusion of the project, I will finish implementing ensembles for both Decision Tree and Perceptron algorithms. Additionally, I will employ the shown-in-class library classification methods. Lastly, I will write an SVM for the final homework which will conclude the algorithms required for the project. In the end, I will have 6 algorithms and they will be fairly decent given the proper pre-processing. That being said, I will get to the pre-processing only if time permits as I have a pretty difficult semester conclusion ahead of me with projects in all of my classes, exams each week for the remainder of the semester and what I have done regarding pre-processing may well conclude my efforts on that front. Nonetheless, I have learned quite a bit so far as I have implemented the different models, used them to tackle the same problem with roughly equal levels of success, and will finish my thoughts on all of this once it is completed in two weeks' time.