Final Project Proposal 5/5/2022
Thomas Kismarton

For my final project, I wanted to build a more complicated classification model using some of the more math-intensive algorithms we've learned about, such as SVM. A few of the compeitions on Kaggle piqued my interest, but I found an independent dataset that I think would be more fun to work with: https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease

There are 17 independent variables included in the dataset, as well as the classification variable 'Heart Disease'. The data comes pre-cleaned, which makes it a bit easier to work with, and is presented in CSV format – a format that I'm comfortable working with.

In order to generate training, testing, and validation data, I'll split the list of ~320,000 samples into three separate sets along the percentages described in class (70% training, 10% validation, 20% testing). Performance can be measured easily – the percentage of correctly classified test samples reflects the accuracy of the model.

Overall, this project still seems relatively basic as far as machine learning goes, since I'm sticking to pre-processed datasets and working on classification models. I think this will still be sufficient for a final project given the complexity of the models I'm attempting to build, while still being manageable. I want to build an SVM model, and I'm planning on building a Perceptron alongside it to showcase the improved accuracy of SVM in comparison to Perceptrons. I might build a third model, since the Perceptron is essentially a chunk of the SVM, but I don't want the prohect to get overwhelming.