Jimmy OSullivan

Mid project report

For my mid-project update, I implemented a Naive Bayes network. I implemented first using all of the data, then using only the continuous variables, and finally with continuous and discrete variables.

The best performing of those, and the one I have shared, is the one with only continuous data. With this, I achieved 98.4% accuracy on a separated test set. However, this accuracy may be biased because of my attack biased training dataset, which makes my balanced accuracy of 84.0% a better measure of performance.

The only outside libraries I used were “ucimlrepo”, which I used to fetch my dataset, and “numpy”, which I used for preprocessing and some basic math formulas, like mean and standard deviation. These could also be done by my implementation, but were much slower (50 minutes vs <1 minute), so I left them as the numpy version for my submission.