Nikilas John, Atmin Sheth

Dr. Mazidi

CS 4375

1. **how SVM works, and how SVM kernels work, your impression of the strengths and weaknesses of SVM**

Support vector machines are a method of classification that can be used for many different classification types. This algorithm takes root in computer science principles as well, which was interesting to me. This algorithm is being taught to us because of its success in scenarios such as binary classification, multi-class classification, and in some instances, regression. In the most common scenario, binary classification, we plug the data into the algorithm and it will output a graph. Contained within this graph will be a hyperplane (colored orange). Then, there will be two margins (colored green) on each side of the hyperplane, and this is where we get our goal with this algorithm from. We need an output that has margins not too narrow and not too wide from the hyperplane. Ideally, all of the instances of data would fall nicely on either side of the margins, or at least on the right side of the margin. But in the case that it wouldn't, the instances would be labeled as slack variables. The strengths of this algorithm are that it can work in a lot of different scenarios and perform well. The disadvantage is that the SVM algorithm does not perform well when the classes are overlapping, which is when a logistic regression algorithm would be better.

2. **how Random Forest works, how the other 2 algorithms you used work compared to the simple decision tree, your impression of the strengths and weaknesses of these ensemble techniques**

Random Forest works as an ensemble method. The distinguishing feature of this parallel ensemble method is that it uses multiple trees on a subset of data and then combines those results to create a strong learner. The different trees use different subsets of data as well as different subsets of features to decrease the variance in the algorithm. XGBoost compares to the simple decision tree because it also uses trees, but runs on your computer using multithreading. This makes this algorithm much quicker to execute and can perform regression and classification. This makes XGBoost one of, if not the, best algorithms to boost datasets. The last algorithm used in the ensemble methods was the adabag algorithm. Adabag (Adaboost) works by iteration, increasing the weight of each training example on each loop, and decreasing the weights of correct observations. After the iterations of all of those learners, their weighted errors are used to weigh the learners, with "more accurate learners given higher weights." There doesn't seem to be an advantage to using this algorithm, especially when fastAdaboost exists. The disadvantage of this algorithm is that it is slow compared to the other boosting algorithms.