

Kernel and Ensemble Methods

SVM:

Support Vector Machines(SVM) are a type of machine learning algorithm which attempts to divide the data points by a single optimal line, called the hyperplane. Though SVM is capable of multiple classification, it excels at binary classification. How the SVM divides the data is determined by the kernel it uses. For this assignment we tried three kernels and used them for both regression and classification. The first, linear, is the simplest in that it draws a straight line between the data and classifies each datapoint based on which side it falls on. The second kernel was polynomial, which is similar to linear save in that it can draw lines of higher dimensions to better classify the data. The final kernel was radial, which could encircle groups of data points for classification. SVM is a powerful algorithm, its classification ability is impressive. The only thing I noticed about SVM that could detract from it overall is the long run times it took, especially when tuning.

Ensemble Algorithms:

For this portion we tested three ensemble algorithms, Random Forest, Boosting, and XGBoost; we also used a simple decision tree for comparison. The Random Forest algorithm generates many random decision trees in parallel based on subsets of the data, then it averages the results of each to create a final decision tree. Boosting uses the entire dataset to generate decision trees sequentially which are combined until a final tree is reached. XGBoost is similarly a boosting algorithm using gradient decision trees. The results of all three algorithms were comparable to the decision tree. The only notable outstanding algorithm is XGBoost, which had a faster run time than the other two ensemble algorithms. Random Forest and Boosting took considerable time to run while XGBoost took mere seconds.