Kernel and Ensemble Methods

SVM – How it works

SVM, support vectors machine, uses support vectors to find a margin that separates classes in various scenarios, including binary and multiclass classification.

SVM Kernel – How they work

SVM kernels can be used when data is not linearly separable, in which kernel functions will map the linear function to another form (such as polynomial or radial) so the data can then become linearly separable. For example, in a radial kernel, the decision boundary would be circular, and points would be classified on whether they appear inside or outside this boundary.

Impression of SVM's Strengths and Weaknesses

Strengths of SVM observed would be how it can be used for various scenarios, both in regression and classification analysis. When it comes to accounting for data that are far from the hyperplane, SVM is robust as well. For these specific points, the lost function ends up being 0 in SVM while being a very low number in logistic regression. But SVM works better than logistic regression when classes do not overlap as much. A weakness observed would be that if classes do overlap, it might end up performing worse than logistic regression.

Decision Tree– How it works

A simple decision tree formulates a set of rules for predictions based on the data given. It's a single decision tree.

Random Forest – How it works

Random Forest is a bagging ensemble method for classification and regression. They construct multiple decision trees trained on subsets of data and combine their results to make predictions.

XGBoost – How it works

XGBoost is an implementation of the gradient boosted trees algorithm. It's an ensemble method that uses a series of many decision trees that minimize the error of the previous tree. This is how it creates a strong learner.

Adaboost- How it works

Adaboost is a type of boosting algorithm that uses decision stumps as weak learners. Weights are assigned to observations in the data based on how easy they are classified and the final result is an average of the weighted outputs of all individual learners where their weight is proportional to their accuracy.

Strengths and Weaknesses of Random Forest, XGBoost, and Adaboost

Random forests can be a strong algorithm because they reduce possible overfitting and are flexible for classification and regression. They also can work well with unbalanced data. However, it has a huge drawback in that it is very computationally expensive.

XGBoost is quick, robust and performs well in both regression and classification scenarios. It also prevents overfitting. However, it doesn't perform well with sparse unstructured data and is sensitive to outliers.

Adaboost is relatively easy to use with less hyperparameters to tune and is easier to understand than XGBoost. However, it is sensitive to outliers like XGBoost and is also much slower.