Abed Ahmed (ASA190005) and Dylan Kapustka (DLK190005)

10 October 23, 2022

Homework 5 - Kernel and Ensemble Methods

CS 4375

SVM and Ensemble Methods Narrative

Although it may be used for regression, the Support Vector Machine is a supervised learning technique that is mostly used for classification. The key concept is that the algorithm searches for the best hyperplane that may be used to categorize new data points based on the labeled data (training data). The hyperplane is a straightforward line in two dimensions.

The categorization of a class is often based on the representative qualities that a learning algorithm learns to reflect the most prevalent traits (what distinguishes one class from another) (so classification is based on differences between classes). The SVM operates in the reverse direction. It locates the class samples that are most comparable. The support vectors will be those.

A collection of mathematical operations known as the kernel are used by SVM algorithms. Data is inputted into the kernel, which then transforms it into the desired form. Different kernel functions are used by various SVM algorithms. There are various forms of these functions. For instance, linear, nonlinear, polynomial, sigmoid, and radial basis function (RBF).

RBFs are the most popular kind of kernel functions. since it responds locally and infinitely throughout the entire x-axis.

The inner product between two locations in an appropriate feature space is returned by the kernel functions. Thus, by defining a notion of similarity, with little computational cost even in very high-dimensional spaces.

Let’s take an example of linear SVM for instance. Say we have class A and class B. While most algorithms aim to identify the defining differences between A and B, SVM will attempt to identify the defining similarities.

It will create two support vectors, one lining up with A instances that are most similar to B, and the other one, vice versa. Finally, the confidence of a instance being A is very high up until the first support vector, then, the confidence of A starts decreasing, while the confidence that the instances are B starts increasing.

As for the Ensemble techniques. The Random Forest model is based on the idea that several uncorrelated models (the various decision trees) work significantly better together than they do separately. Each tree provides a classification or a "vote" when using Random Forest for categorization. The categorization with the most "votes" is chosen by the forest. When performing regression with Random Forest, the forest selects the mean of all tree outputs. Boosting is an ensemble learning strategy that creates a strong classifier out of a number of successively weak classifiers. In order to address the bias-variance trade-off, boosting techniques are essential. Boosting algorithms regulate all

the components of a model—bias and variance—and are thought to be more successful than bagging algorithms, which solely account for excessive variance. Gradient Boost is a gradient descent optimization approach to add weak learners with the primary goal of minimizing the loss function. As a result of the generalization, the approach was able to handle regression, multi-class classification, and other tasks in addition to binary classification difficulties. Finally, Let's see how AdaBoost functions. The algorithm creates 'n' trees when the random forest is employed. It creates correct trees with a start node and many leaf nodes. Although some trees may be larger than others, a random forest has no set depth. However, AdaBoost's approach only creates the Stump node, which has two leaves.