Brandon Garza

CPTS 315

HW3

Analytical Part

Q1:

1. The decision boundary for voted perceptron is non-linear because it has many weight vectors due to it analyzing many points and choosing the best.
2. The decision boundary for average perceptron is linear because it has a single weight vector due to it analyzing all points together.

Q2:

To update the perceptron algorithm to handle an importance weight you would simply multiply the importance weight to the learning factor when updating the weights.

Q3:

To adjust the perceptron algorithm, you could introduce a set of hyperparameters, one for the positive examples and one for the negative examples. The hyperparameter for the positive examples would be scaled higher than the negative hyperparameter proportional to the difference of examples. These hyperparameters would be multiplied in the update function resulting in a higher rate of correction in the positive examples than the negative examples.

Q4:

When calculating the results of the “Data Mining Fight” between (K > L) the result will be either a positive value or a negative value. This means that it will be a binary classifier. I would use the pairs of candidates as training examples and have the perceptron algorithm attempt to guess the winner of the fight. If the guess was incorrect, I would use the feature vector of the winning feature vector to update the weight vector w. I would continue with this for each pair of candidates.

Q5:

The decision boundary of CHOICE will be in the form of sign(w \* x + b) which is linear because the examples will either be closest to the center negative or the center positive meaning that the boundary will lie when the example is equally far from either center. When graphed this results in a linear boundary equally distanced from each point. The value of w will be the minimum of |x – C+| and |x – C-| and the value of b should be the distance from either center to the decision boundary or (C+ - C-)/ 2. All the training examples should lie on the linear decision boundary between the two center points.

Q6:

One of the largest fields in computer science is Machine Learning which has allowed computations to be much more efficient for applications such as recommender systems, stock trading, and spam filters. Machine learning lowers the amount of manual programming needed for applications and allows the algorithms to adapt when new data is presented to them. Although there are many resources available for learning and applying machine learning algorithms, there is a lack of information on shortcuts and tricks that allow for ease of use which this article seeks to amend.

There are three main aspects to learning algorithms that can help you decide which one to use for a given application. These aspects are Representation, Evaluation, and Optimization. Representation dictates the way that data or information is presented to the learner. The set of classifiers for the inputs must be in a format that is able to be read by the computer, the set of classifiers are in the hypothesis space of the learner. Evaluation describes the evaluation function that the learner uses internally to determine which classifiers are good ones and which are bad ones. Lastly, Optimization describes the way that the algorithm chooses the classifier that scores highest.

It is important to remember that the overall goal of Machine Learning is to be able to generalize for data beyond the test data. This means that although a learner adapts to the test data perfectly, it can still be a bad algorithm because it is unable to correctly predict further examples. This shows that data is not everything in machine learning, there are other factors that will assist in generalization such as dependencies and similarities between data. These extra factors also assist in decided best choice of algorithm for the given data.

When a learner is overly accurate on the training data but does not generalize to testing data, it is said to be overfit. The generalization error can be represented in two aspects, bias and variance. Bias is the learner’s tendency to learn a wrong thing multiple times, and variance is the learner’s tendency to learn random things that are not applicable. Different learning algorithms fall prey to these aspects of overfitting such as a linear learning that has high bias and a decision tree that has high variance. Another issue in Machine Learning comes when increasing the dimensionality of the feature vectors. When dimensionality increases so does the difficulty of generalization so it is not always helpful to add more features to the input examples as this will only increase the difficulty of creating a meaningful generalization.

One of the most important factors in the success of a machine learner is features. To have a better chance at success, it is helpful to have features that are independent and correlate well with the class as opposed to features that have a complex relation to the class. Most of the time spent on a machine learning task is in choosing meaningful features from raw data to facilitate successful learning. After using this method to design a good learner, there is still a chance that is not as accurate as you would like. To improve it further, you can either try to write a better algorithm, or you could simply gather more data for the learner. In a lot of cases a dumb algorithm with lots of data can out perform a smarter algorithm with less data.

These points are important to fully understanding learning algorithms and to efficiently using them. A lot of the issues presented in the paper are not always readily available, there seems to be a lot of information on ideal cases but this article did a great job of presenting a set of solutions to real world issues in machine learning.