Assignment 4

Due on May 8, 2020
Machine Learning Fundamentals and Applications
CS559−Spring 2020
Professor In Suk Jang

Daniel Kadyrov
10455680

1 Part 1

Provided in `kadyrov_daniel_assignment4.py` attached.

2 Part 2

2.1 Logistic Regression

Classification through logistic regression uses the sigmoid function to predict the probability a given data entry belongs to a category. Although it is easy and effective especially when feature scaling and hyperparameter tuning is not needed, it performs poorly (as seen in this example) on non-linear data, like for image classification.

2.2 Support Vector Machine

The Support Vector Machine algorithm uses a hyperplane in a dimensional space to separate the data where the number of dimensions is based on the number of features. SVM performs well when there are a lot of easily separable features. SVM, however, is slow and does not perform well when the classes have overlaps. SVM performed second best when compared to the other algorithms.

2.3 Decision Tree

Decision Tree classification uses a treechart like algorithm where internal nodes test an attribute and create branches based on the outcome of the tests. Each node representing a label. Although Decision Trees does not require scaling or normalization and has automatic feature selection, the algorithm can be easily overfit and requires a long time to train.

2.4 Random Forest

Random Forest expands the decision tree algorithm by building multiple decision trees and chooses the most voted prediction as the result. It avoids overfitting and reduces the errors commonly found in decision trees and can handle large amounts of data. The data for Random Forest needs to have predictive features to work effectively. Random Forest performed the best out of all the algorithms used.

2.5 Comparison

Table 1: Comparison of Classification Model Accuracy

kmeans	lr	svc	dt	rf
0.5946	0.9258	0.9417	0.879	0.9693

3 Part 3

Provided in `kadyrov_daniel_assignment4_part3.py` attached.

4 Part 4

I faced issues importing my own handwriting in the same format as MNIST. There are no resources provided on Campus, in the materials required by the course, or online that I could find to make my images work. I will continue looking but to ensure the submission is on time, I had to stop.