## Analysis - Assignment 5

## aarigela-singh74-snehchem

# Nearest neighbor

To train, run below command:

python orient.py train train-data.txt nearest

To test, run below command:

python orient.py test test-data.txt nearest

Nearest neighbor is a special case of k nearest neighbor where k=1.

We analyzed the accuracy for various values of k instead of just taking the nearest neighbor, we found that setting the k=25 gives the maximum accuracy which is 71.04% but for k=1, we get the accuracy as 67.23%.

|  |  |
| --- | --- |
| k | Accuracy |
| 1 | 67.23 |
| 5 | 69.14 |
| 10 | 70.3 |
| 15 | 70.20 |
| 20 | 70.5 |
| 25 | 71.04 |
| 30 | 70.20 |
| 35 | 70.41 |
| 40 | 70.2 |

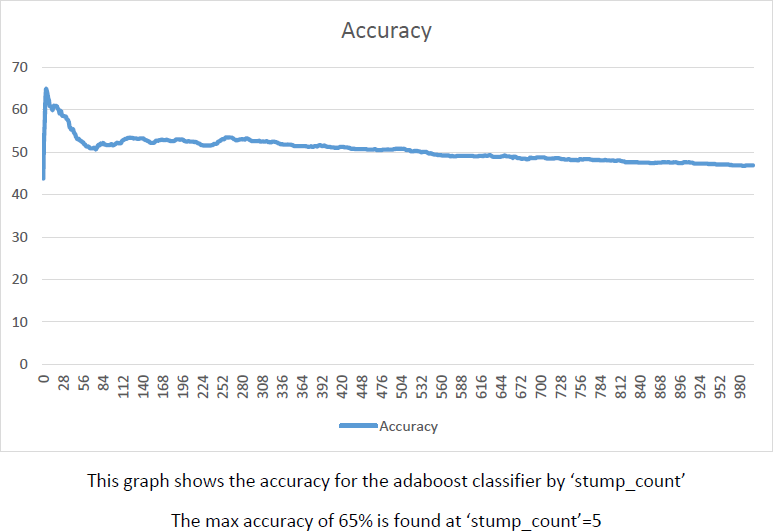
# Adaboost

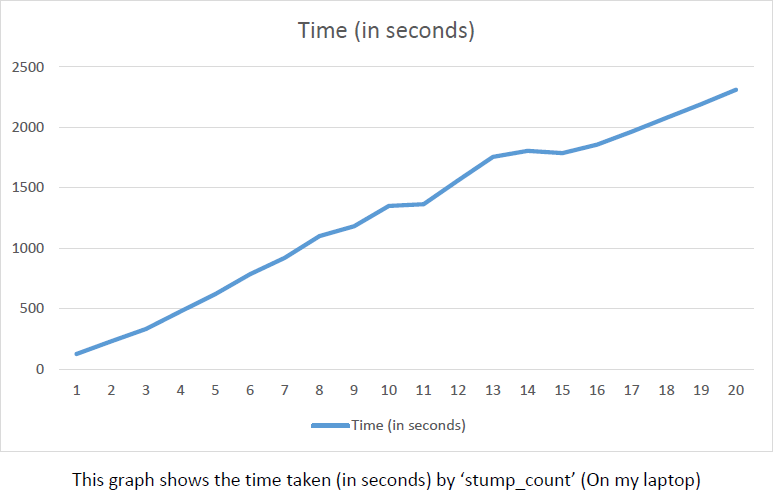
Run python orient.py train train-data.txt adaboost\_model.pickle adaboost [stump\_count]

Run python orient.py train train-data.txt adaboost\_model.pickle adaboost [stump\_count]

We ran the adaboost method with different stump counts and found that initially the classifier is weak, then as we add more stumps the accuracy increases, but after a certain value (in our case 5) it began to decrease as the model began to overfit.

The plot of percentage accuracy vs stump count is shown below:



The plot of Time taken (in seconds) vs stump count is shown below:

# Neural network classification

# 4. Analysis on Partial Data

If we run our models on partial data we see that some of the images are misclassified for example this image below:

the image with id:10351347465 was misclassified by Neural Network in the whole data and also by Adaboost but not by nearest neighbors, but after running on partial data Adaboost and Nearest neighbor could not classify the image correctly but neural network could do it. If we train the models on some other set of data and then again run them we see that this image classification was similar as that on whole. So we can say that changing the training data can lead to misclassification or correct classification of some images, but again if we change the training data it may lead to irratic results.

# 

# 5. Best

Neural network is our best model and our implementation is much efficient and does not take much time to train.

It is working at an accuracy of 73% higher than other classifiers. In general after trying above analysis on different sets of data Neural Network could better generalize the model and performed better than others and after that K nearest neighbor performed better than adaboost, hence we can say that adaboost tend to overfit the data. Hence we see that Neural network was one of our best model.