Report Assignment 4: Learning

Team Members :
Aravind Bharatha (abharath)
Nawaz Hussain Khazielakha(nawazkh)
Rahul Pochampally(rpochamp)

We have addressed the following queries in our report.

- *In your report, present neatly-organized tables or graphs showing classification accuracies and running times as a function of the parameters you choose.*
- *Which classifiers and which parameters would you recommend to a potential client?*
- *How does performance vary depending on the training dataset size, i.e. if you use just a fraction of the training data?*
- *Show a few sample images that were classifiers correctly and incorrectly? Do you see any patterns to the errors?*

Tables of classification, accuracies and running times

- For the implementation of KNN (or K-nearest) classification, we have logged the accuracy and Run Time for each value of K.

Value of **K**	Accuracy(in %)	Run Time(mins)	Training Data Size(Images)	Test Data Size(Images)
5	69.14	5.16	36,976	943
10	70.12	5.21	36,976	943
11	71.04	5.39	36,976	943
15	70.2	6.11	36,976	943
20	68.57	8.21	36,976	943

- For the implementation of Adaptive Boost classification, we have logged the accuracy and Run Time for each number of decision stumps used.

No of **Decision Stumps**	Accuracy(in %)	Run Time	Training Data(Images)	Test Data(Images)
100	28	3(secs)	36,976	943
500	32	2(mins)	36,976	943
1000	37.04	35.39(mins)	36,976	943
10000	35.04	1.8(hrs)	36,976	943
$191 * 192$	43.57	2.26(hrs)	36,976	943

- For the implementation of neural-nets classification, we have logged the accuracy and Run Time for each number of iterations used.

Training Data(Images)= 36,976 **;** Test Data(Images) = 943			
No of **Iterations**	Learning Rate	Run Time	Accuracy(in %)
100	0.5	30mins	71.37
500	0.25	1hr50mins	70.55
1000	0.01	1hr50mins	70.45

What to recommend to a potential client?

As per our testing, we concluded that Neural Nets provide best accuracy. Hence, to a potential customer, we would recommend to go with Neural-Net classifier.
Recommended parameters would be :

- Number of Iterations: 100
- Learning Rate : 0.5

Adaboost algorithm also provides *good accuracy*, however the if the client has **lot of time** at his expense, then **Neural-Nets** is a suggestible option.

However, if the client has **very limited time**, then **K-Nearest** is suggestible classifier as the *accuracy of KNN is comparable* if not the best among all.

Selection of classifier depends on the amount of time client is willing to spend on receiving the output.

Performance Vs Training Data-set size

Having a richer set of training dataset is always beneficial in generating the best model. In our case,

- For KNN classifier:
  - Bigger the training data set, more time consumed to compare and predict the solution.
  - However, if the training data set is reduced then there are high chances of not able to predict the solution with high accuracy. It is because we are essentially comparing the input image with all the existing data.
  - Hence, If performance is in terms of accuracy, then bigger the Training data set, the better the performance.
- For adaptive boosting classifier
  - Bigger the training data set, more time consumed to tuning the model.
  - If the data is really big, then there are chances of over-fitting the model and loosing our accuracy.
  - Performance is dependent on number of decision stumps being generated too. Therefore, if the input image(here) has more attributes(pixels) to be tuned for, then adaboost would suffer with major time lag.
  - Finally, if the input training data set is really big, then there are chances of over-fitting the model and comprising our accuracy.
- For Neural Nets
  - If the number of iterations are increased, then the model is trained for all the iterations. Hence, this classifier would also suffer with over-fitting the model hence compromising the accuracy.
  - Training data set has slight effect on the performance of the classifier, rather the number of iterations and learning rate play a major role in predicting the value of an input.
  - Therefore, we have to balance the number of iterations with the size of the input data to achieve optimal accuracy in small amount(practical) amount of time.

Pattern in classification

- We observed that images which were bright such that the source of light was right before the subject gave a hard time to the classifiers in predicting their orientation.
- Moreover the images which had a gradient in their brightness were predicted correctly (most of them) by Adaboost and Neural Nets as both the approaches(Adaboost and Neural Nets) train themselves to detect the gradient(shift in image brightness).
- Images with complex background were tough to predict with any of the given classifiers. (Complex background implies images having similar pixel values all around)