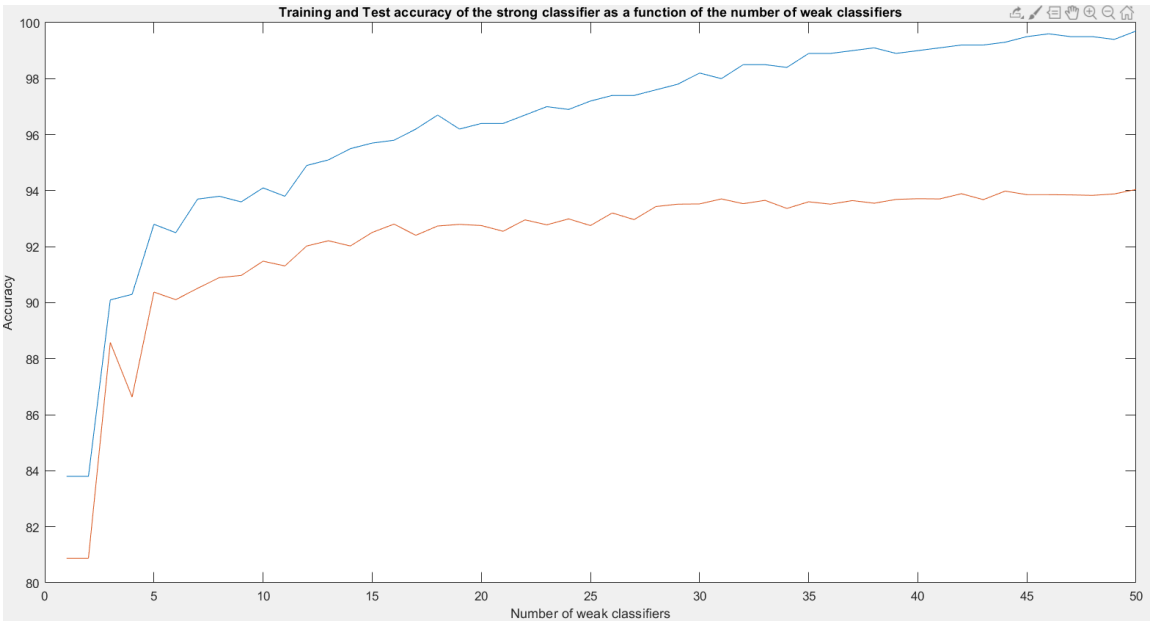
TBMI26 – Computer Assignment Reports Boosting

Deadline – March 15 2020

Author/-s: Vinay<vinbe289>

In order to pass the assignment you will need to answer the following questions and upload the document to LISAM. Please upload the document in PDF format. **You will also need to upload all code in .m-file format**. We will correct the reports continuously so feel free to send them as soon as possible. If you meet the deadline you will have the lab part of the course reported in LADOK together with the exam. If not, you'll get the lab part reported during the re-exam period.

1. **Plot how the classification accuracy on <u>training data and test data</u> depend on the number of weak classifiers (in the same plot). Be sure to include the number of training data (non-faces + faces), test-data (non-faces + faces), and the number of Haar-Features.**



Number of training data: 1000
Number of test data: 11788
Number of Haar-features: 100

2. **How many weak classifiers did you use when training? How many of them did you use for the final strong classifier? Motivate your choices.**

   50 weak classifiers were used for training. The same number i.e. 50 were used for a final strong classifier. The number of weak classifiers could be increased but in this case since there is substantial amount of training data 50 number of weak classifiers would suffice to achieve a good accuracy of the Training and Test data.

3. **What is the accuracy on the training data and test data after applying the optimized strong classifier? Discuss your choice of hyperparameters and how they influence the accuracies.**

   Training data accuracy: 99.7 %
   Test data accuracy: 94.0448 %

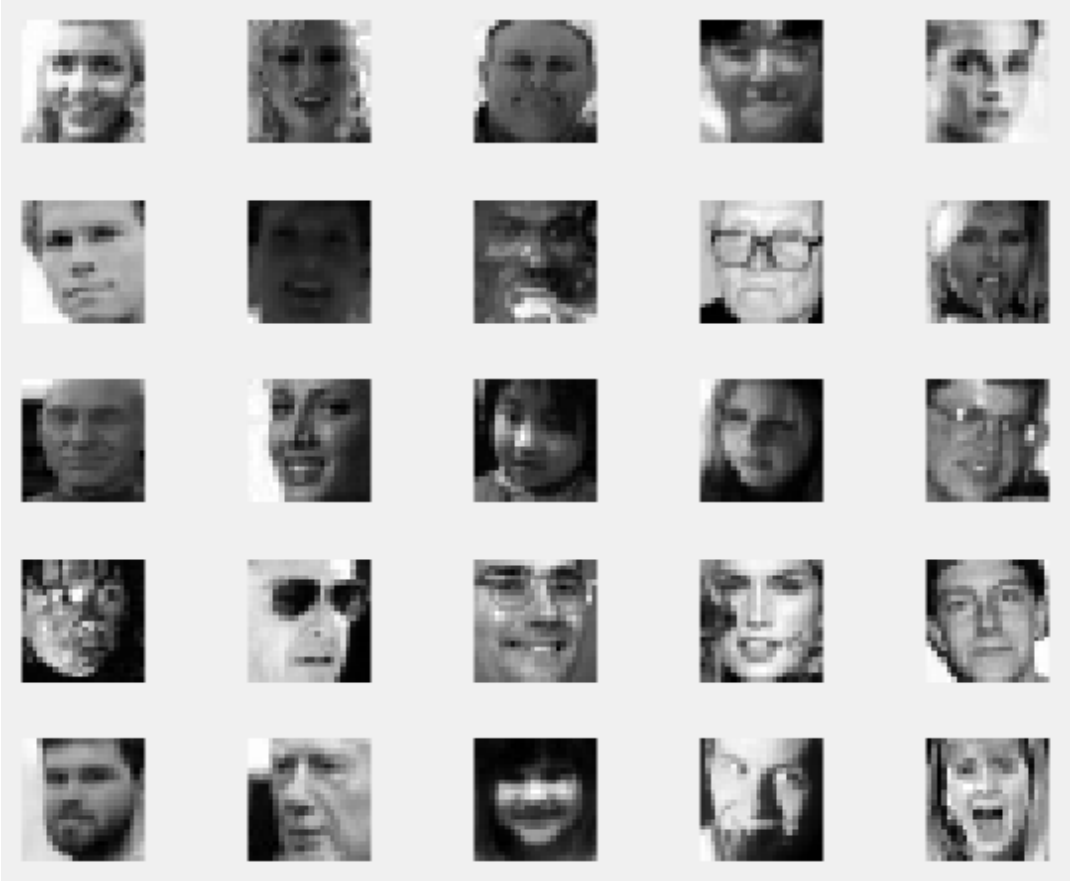
   Following are the chosen values for hyper-paramaters
   Number of Haar-Features: 100
   Number of Training images: 1000
   Number of Weak classifiers: 50

   Theoretically increasing any of the above hyper-parameters would increase the accuracy of the classifier. Number of Haar-features were chosen to be 100 at random since it seemed like a good choice. Number of training images were chosen to be 1000 since it would roughly be 10% of the total data and half of it of one class type and the other to be the second class type. Number of weak classifiers could typically be higher but as mentioned above since the data used for training is of a considerable amount we could keep the number of weak classifiers as low as 50.

4. **Plot the Haar-features selected by your classifier (one for each weak classifier). If you have many weak classifiers, select some representative subset. Can you think of why they would be useful for classifying faces?**

   Classifying faces and no-faces is recognizing if there is a face in the given image. The key aspect in recognizing faces is detecting relevant features in human face like eyes, eyebrows, nose, lips. The Haar features displayed above are square shape function that can be applied to all relevant parts of the image to check if a face exists in the image. The edge and line features displayed above are the kind of features required to check if the image consists of a face or not.

5. **Plot some of the misclassified faces and non-faces that seem hard to classify correctly. Why do you think they are difficult to classify?**



The above image consists of a number of misclassified faces. The reason for these faces being misclassified as non-faces maybe the fact that the faces are in different forms, a few faces are tilted at an angle, a few have a very bad image quality, a few faces have glasses on them and several other factors that might not help the Haar features in distinguishing faces from non faces but it is different in the case of distinguishing the images by naked eyes.

The above image consists of a number of misclassified non-faces. The reason for these non-faces being misclassified as faces maybe the fact that each of these above images has curved features in it which might be mistaken to be facial features such as eyes, nose etc. by the model. The quality of images is very bad similar to the previous case which might also trick the Haar features into picking up wrong signals from the images and classifying them as faces.

6. **Are your results reasonable? Can you think of any way to improve the results?**

   The results are fairly reasonable considering the amount of training samples, the number of features used and the training time. The result accuracy might be improved by using more number of Haar features, probably more sophisticated features, and also by increasing the amount of training samples and the number of weak classifiers.

7. **Can we expect perfect results? Motivate your answer.**

   We cannot expect perfect results i.e. 100% accuracy considering the fact that not all the images are the same, there might always be a few images which deviates from how the other images of the same class behaves. As mentioned above faces with bad image quality, faces with glasses on them, faces tilted at an angle are anomalies that is hard to capture and classify distinctly unless very sophisticated Haar features are used for training the model.