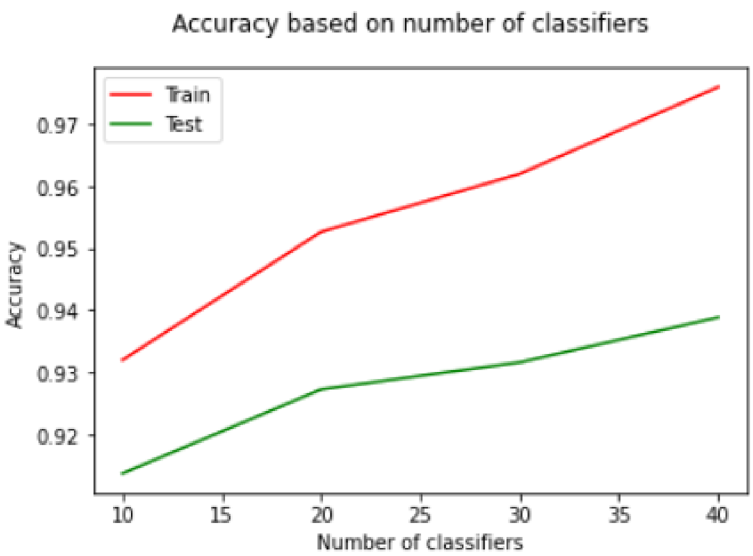
TBMI26 – Computer Assignment Reports
Boosting

Deadline – March 14 2022

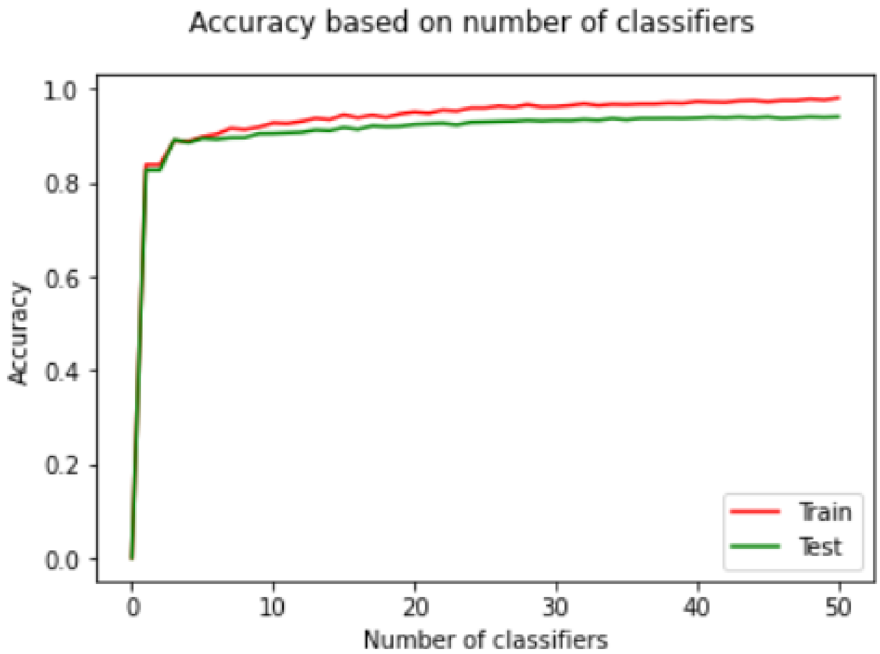
Author/-s: Chayan Shrang Raj (chash345)
Theodor Emanuelsson (theem089)

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.**



We tried different number of classifiers such as 10,20,30,40 and plotted the train and test sets accuracy for each set of classifiers and we see that as the number of weak classifiers increase the accuracy is also increasing. For all the settings above, we had 1500 training data images (first 750 faces + last 750 non-faces) and 11288 test data images (first 4166 faces + last 7122 non-faces). Also, we chose 150 haar features for both settings.



So now, we are not training models for different number of classifiers and have accordingly only measured the accuracy of each classifier. Also, we have increased the number of weak classifiers to 50 this time for better comparison.

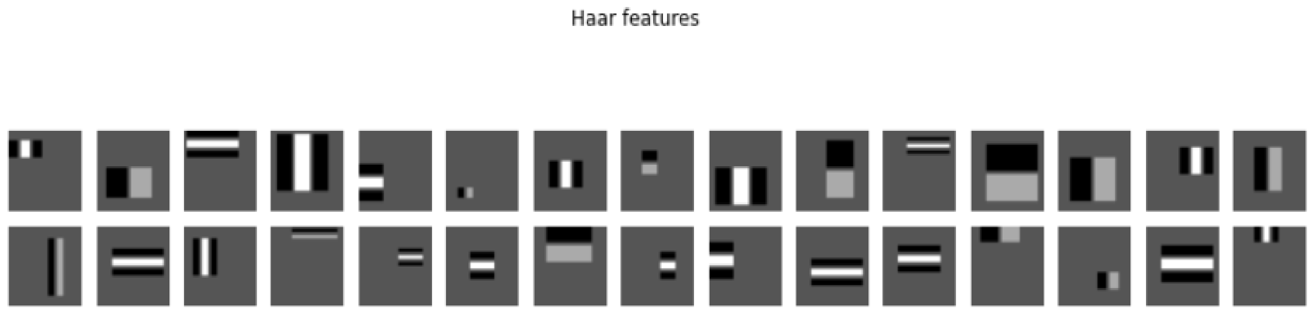
2. **How many weak classifiers did you use when training? How many of them did you use for the final strong classifier? Motivate your choices.**

As, it can be seen from the above graph, even if we use only 35-40 weak classifiers instead of all 50 classifiers, the increase in accuracy is only 0.005 (we got this by looking at the array of test_acc and looking at the increase in accuracy and by looking at the above graph) in the test data and similarly for training. So, keeping in mind the computational complexity, we could do equally well without using all 50 classifiers. As can be seen in the plot, the gap between the training and testing accuracy increases as the number of classifiers increases. A reasonable suggestion would be to use 4-6 weak classifiers as the training and testing accuracy is about the same. One could argue that having the accuracies close to each other is a sign of underfit. However, since no substantial increases in test accuracy occurs by using additional classifiers, this number seems reasonable. But, in our case, it is mentioned that we need to have at least 93% accuracy according to compendium, so we need to use around 30-35 weak classifiers to achieve that test accuracy.

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 accuracies.**

After applying the final strong classifier on training data, we got a training accuracy of 98% and we chose 150 haar features, 50 weak classifiers for training the classifiers for the first time and accordingly for each weak classifier which can be described as a triplet of feature value, threshold for that feature and polarity for that feature. We get testing accuracy of 93.9% accuracy after applying all 50 weak classifiers on test data and would consider as a performance metric as this was evaluated on unseen data not used in training. We can also use less weak classifiers with an inconsequential decrease in accuracy. Increasing the number of training images and consequently keeping less images for test classification reduced some variance but after some point did not matter much. Increasing the number of classifiers also influenced accuracy up to a certain maximum limit but it comes at the cost of computation, so we chose 50 classifiers as it gives decent results. 150 haar features are enough to detect all the faces in the image but also, we would have needed more features if the photo quality were not good enough to separate out faces and non faces.

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?**



This is a subset of the total number of haar features that we have used in our classification model. We think they are enough because if we look closely at them, we can see they cover almost all the representations of various parts of a face and each feature is working to learn one distinguishing feature of a face. We may or may not need more features which depend on the quality/noise of the images.

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?**

We think they were difficult to classify as we can see there is a lot of noise in these images, and they also have some type of noisy screen in front of them like rotated face, glasses or blurry lines.

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

Yes, we think our results are reasonable. As we have used around 150 haar-features out of some 150,000 haar features. And around 1500 training images where half are faces and the other half are non-faces. We could slightly improve the results by choosing more training data or increasing the number of weak classifiers although the training error is already almost null. Each feature is a single value obtained by subtracting the sum of pixels under the white rectangle from the sum of pixels under the black rectangle. Now these features are scalable meaning they can made of different sizes inside a 24x24 (the image could size be different and accordingly different number of features) image in our case. These features can be scaled to any size except a 1x1 pixel. This is because each cell of the image array represents a single pixel, which can only contain the value for a single color. The other reason is because the entire function of these frames is to compare the whiteness and blackness of neighboring pixels and, therefore, cannot be scaled down to a single pixel. Hence, if we look at all the possible combinations of these scalable and placed where these features can fit inside an image, the total number crosses 150,000 features for a 24x24 image. And, it will differ based on the dimensions of the images.

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

Although we can expect the error rate for training data to reach zero, it is difficult to expect perfect results on new unseen data because of several reasons. This algorithm is highly influenced by noisy data because it tries to fit in every data point. Also, it depends on the choice of weak classifiers that are being used for training. This algorithm tends to produce an increasingly large number of false positives and can be reduced by choosing lots and lots of high quality-less noisy images of both faces and non-faces.