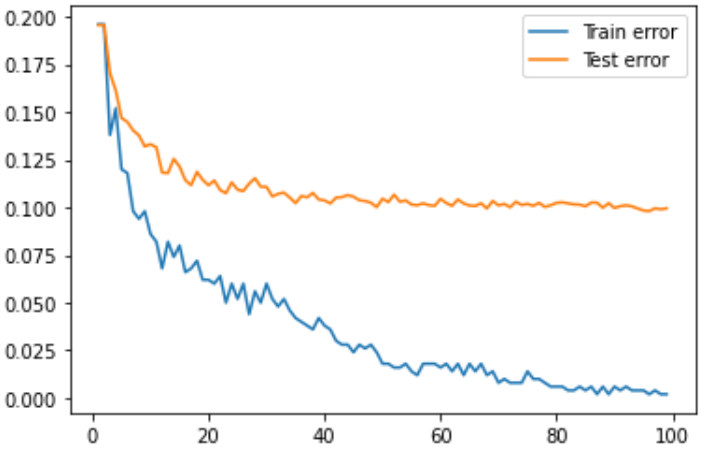
TBMI26 – Computer Assignment Reports Boosting

Deadline – March 14 2021

Author/-s: Mohammed Ali & Simon Alsén

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.**



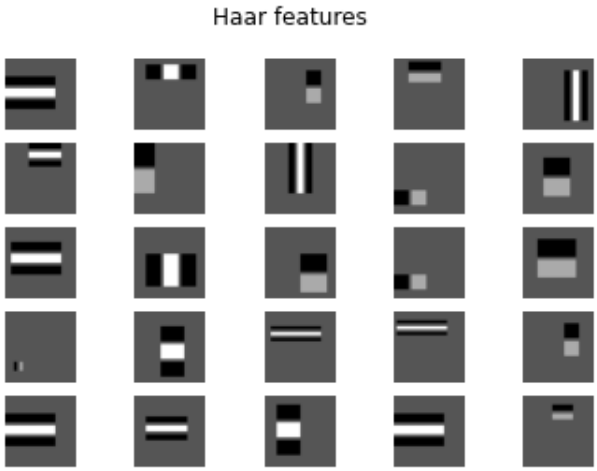
The number of training data is 2000, and the number of test data is 10788. 80 Haar-features have been used. As the number of weak classifiers increases, the error decreases for both the training and the test data. However, there is very little improvement in test accuracy when 60 or more weak classifiers are used.

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.**
   We used 60 weak classifiers for our final strong classifier. The plot above shows that there is very little improvement in test accuracy when more than about 60 weak classifiers are being used. Using a large number of weak classifiers increases the risk for overfitting.

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.**
   The accuracy on the training data is 0.989, and the accuracy on the test data is 0.944. We have used 2000 training images, 80 Haar-features, and 60 weak classifiers. Increasing the number of training images will increase the accuracy, but it will also increase the time it takes to train the model. Likewise, increasing the number of Haar-features might increase

accuracy, but also the time it takes to train the model. Increasing the number of weak classifiers increases the accuracy, but after about 60 weak classifiers the gain in accuracy is very small.

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?**



Most of the Haar-features above seems to be located where distinct facial features such as eyes, nose etc. are usually located in images. Hence, the Haar-features can be useful for classifying faces.

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?**

Potential reasons to why the above faces are hard to classify correctly could be that the images are very bright, or very dark, that the contrast is low, strong shadows, or the people wearing glasses, having facial hair etc.

It is hard to tell why the non-face images above have been wrongly classified. Some of them seem to contain circles that potentially could be confused for heads.

6. **Are your results reasonable? Can you think of any way to improve the results?**
   Yes. The model manages to classify more than 9 out of 10 images correctly, despite a relatively low number of training images. The accuracy could eventually be improved by using a larger amount of training data, or more Haar-features,

7. **Can we expect perfect results? Motivate your answer.**
   Perfect results are very unlikely, even with large amounts of training data. Images can differ in many different ways – for example facial features, contrast, brightness, shadows etc. can differ a lot. Object in non-face images can also be aligned in such a way that they almost look identical to a face.