TBMI26 – Computer Assignment Reports Boosting

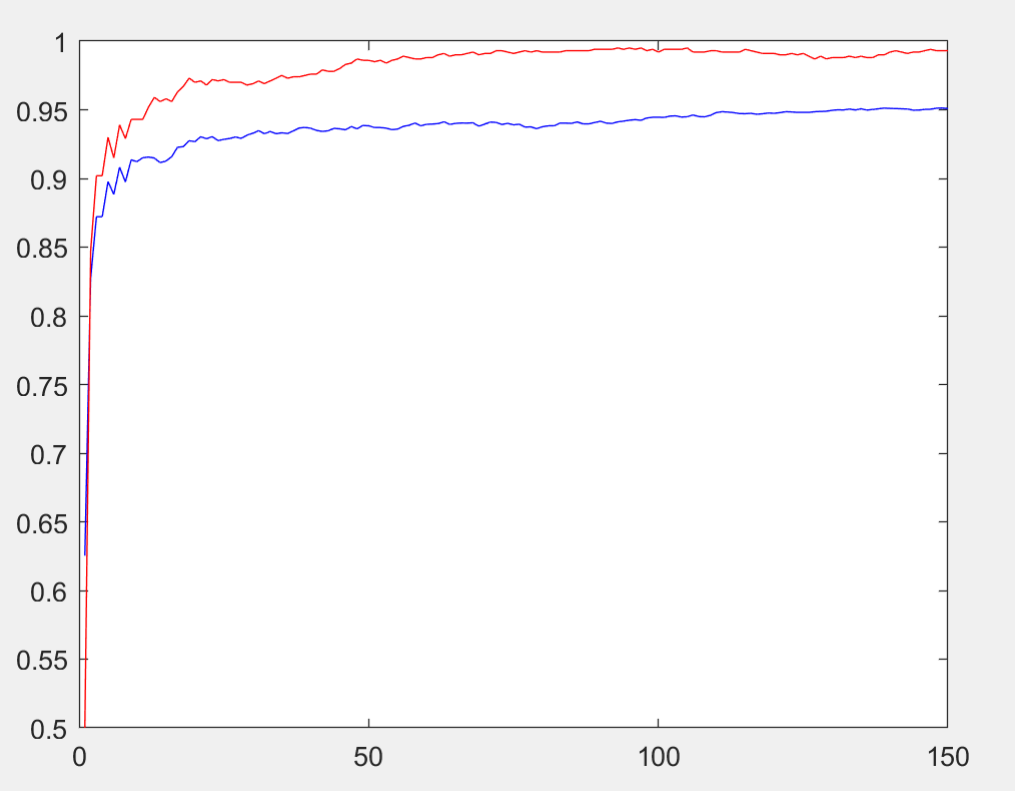
Deadline – March 15 2020

Author/-s:
Gustav Wahlquist & William Stokke
Guswa204, Wilst664

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 randomized Haar-features
nbrHaarFeatures = 300;
% Number of training images, will be evenly split between faces and
% non-faces. (Should be even.)
nbrTrainImages = 1000;
% Number of weak classifiers
nbrWeakClassifiers = 150;
```



Y-axis: Accuracy, X-axis: # of weak classifiers.
Blue line: Accuracy on test data, Red line: Accuracy on training data.

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 choose to use 40 weak classifiers for our strong classifier. This since we can see that the performance of our strong classifier flattens out at about that number of weak classifiers (as seen in the graph above).

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.**

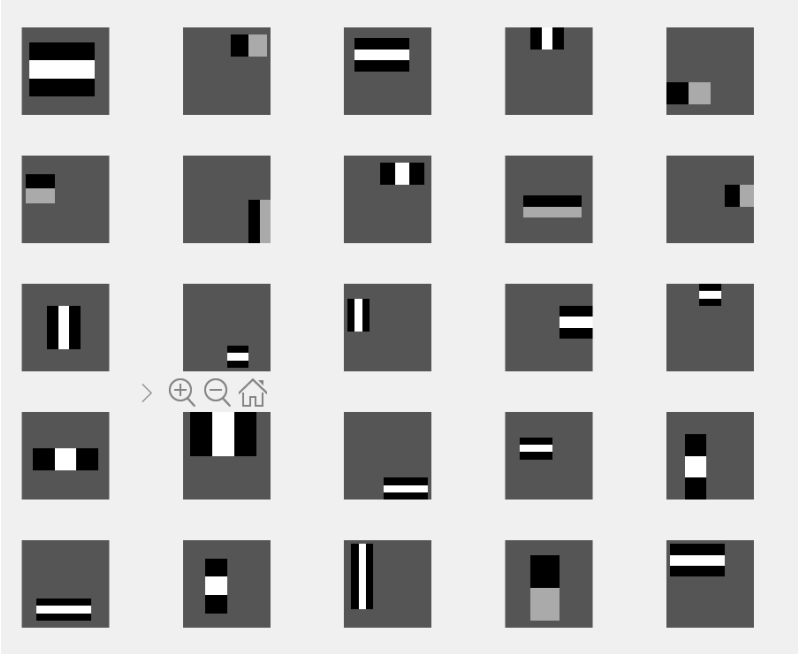
```
trainAcc =

    0.9830


testAcc =

    0.9406
```

The probability of finding features good for classifying the data increases with the number of generated Haar-features. But it will increase the training time.

The number of training samples increases accuracy as well as training time.

The number of weak classifiers used in training increases the training time as well but the increase of accuracy flattens out for large numbers.

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?**



It is hard to be certain but at least some of the Haar-features could be useful to identify distinct features for faces (i.e. eyes, mouth, nose, chin).

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?**

These are wrongly classified as non-faces. Some reasons for this could be:

- Some of the people have glasses
- Some pictures are dark and blurry
- Some pictures have shadows that differs from most of the other pictures

These are wrongly classified as faces. It is hard to find general distinctive features in the pictures that explain this. It could be that these pictures simply differ from the rest of the pictures.

6. **Are your results reasonable? Can you think of any way to improve the results?**
   We could of course train with more data in order to get a more accurate classification, but we believe that the difference isn't substantial enough.

   We don't now the limit for accuracy when using Haar-features, but we may have to change our classifier in other ways than increasing the parameters in order to get better performance.

7. **Can we expect perfect results? Motivate your answer.**
   We haven't inspected the whole data set, but It can be expected that some of the pictures differ too much and because of it making It unreasonable to expect a perfect result.

   In order to be able to reach a very high accuracy we would have to train the model to perform perfect on the training data, this increasing the risk for overfitting the model to the training data and not reach a good result on test data. Its hard to make a perfect and robust classifier.