TBMI26 – Computer Assignment Reports  
Boosting

Deadline – March 15 2019

Authors: Anubhav Dikshit <anudi287>

Hector Plata <hecpl268>

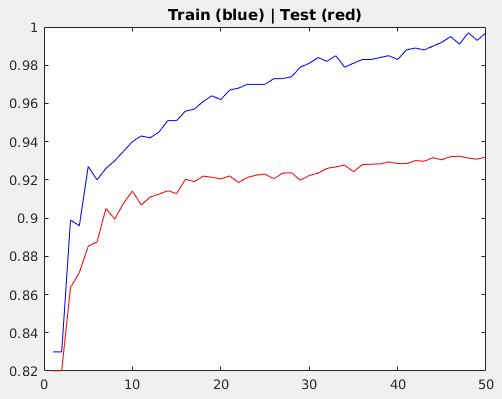
In order to pass the assignment, you will need to answer the following questions and upload the document to LISAM. **You will also need to upload all code in .m-file format**. We will correct the reports continuously so feel free to send them as soon as possible. If you meet the deadline you will have 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 training data and test data 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: 111788

Number of Haar-Features: 100

****

1. **How many weak classifiers did you use when training? How many of them did you use for the final strong classifier? Why?**

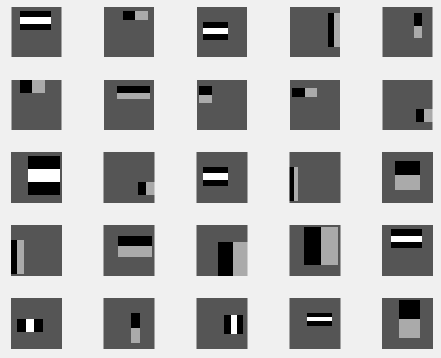
50 weak classifiers were used during training as well as for the final classifier. It was done this way cause this number maximizes the accuracy on the test data.

1. **What is the accuracy on the test data after applying the optimized strong classifier?**

After applying the strong classifier, the accuracy on the test data was 93.2%.

1. **Plot the Haar-features selected by your classifier (one for each weak classifier). If you have many weak classifiers, select some representative subset.**

A subset of 25 (out of 50) Haar-features are shown below.

****

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

Below are shown misclassified faces and non-faces by the algorithm from the test set. These images seem hard to classify thanks to the proximity of the face to the photo and other artifacts like shadows, sun glasses and shadows. As for the non-faces you could see some patterns that the Haar-features might pick up.





1. **Defend your results. Are they reasonable?**

The results are reasonable in the sense that not too many Haar-features were used. For simple images it performs well without a complex model. Another aspect to consider is that the ration of training data with respect to the test data is big, which means that we get pretty good results for just training the model with a thousand images.

1. **Can we expect perfect results? Motivate your answer.**We can’t expect perfect results, at least realistically. The only way to achieve perfect results is by overfitting the test data by doing an automatic grid search over the hyper parameters of the data. Just by change, there is a possibility that the model will get the right number of Haar-features and weak classifiers that enables it to get a perfect accuracy across both data sets. This would require a lot of computational power and time, which in reality no one has. Another fact is that data is always noise, so it’s impossible to get perfect results thanks to the existence of the Bayes error.