TBMI26 – Computer Assignment Report
Boosting

Deadline – March 16 2020

Authors: Karolin Stork & Viktor Gustafsson
Revision 1

Q1. 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.

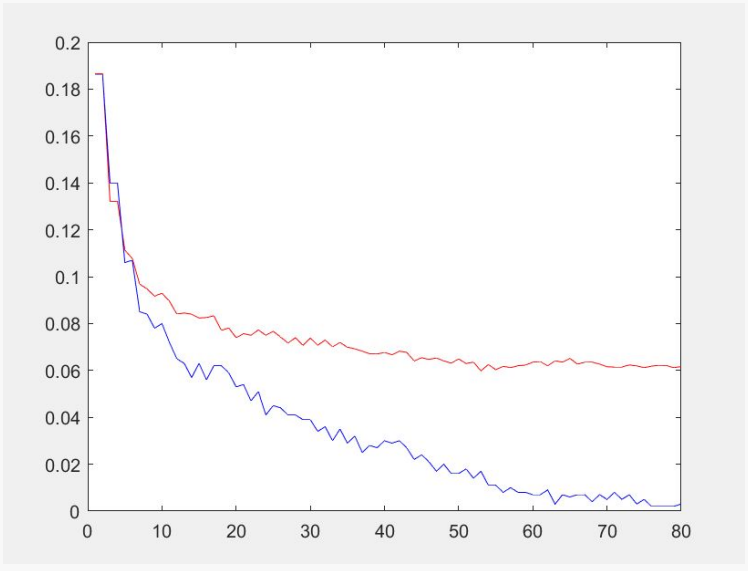

Fig 1. Train error (blue) and test error (red) in relation to the number of weak classifiers included.

Haar features: 100
Training set: 1000 samples
Testing set: 11788 samples

The test error decreases with the number of weak classifiers included in the strong classifier. As we can see, the error is initially decreasing fast but the decrease starts to slow down substantially after about 20 weak classifiers. The error can increase locally (slight oscillation seen in the graph) but the error decreases over many classifiers.

Q2. 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 chose 80 for training, as we noticed that the error didn't decrease much past ~30 classifiers. The final optimal model was chosen by finding the amount of classifiers where the test error was at its minimum during training.
optAmmount = find(errorsTest==min(errorsTest));

This turned out to be **53**.

Q3. 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.

We increased the number of train images from 500 to 1000 to achieve a smaller testerror so that the model has more data to train on. We could increase train set size even further but it would significantly slow down our training time so we kept it at 1000. An increase in size would probably result in a lower error.

We also increased the amount of haar features from 40 to 100 to achieve a smaller error. This will yield a higher accuracy because there is a larger pool of features to choose from in the weak classifiers each iteration of AdaBoost.

For 53 weak classifiers with 100 haarFeatures we obtained:

Test Error	5.98% => ~94% accuracy
Train Error	1.70%

Q4. 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?

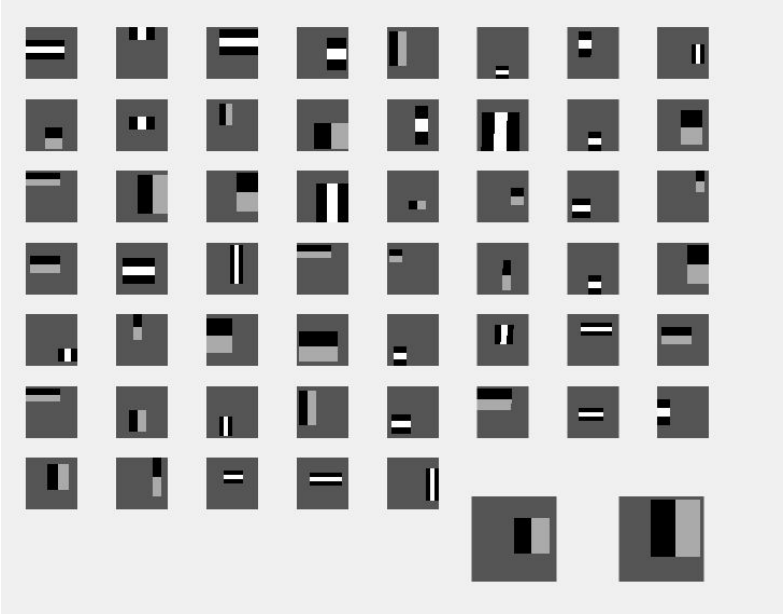

Fig 2. The optimal haar feature for all the weak classifiers used.

Many of the haar features seem to be aligned along possible edges of faces. We believe this is because the most easily recognizable shape of the face is the overall oval structure.

Q5. Plot some of the misclassified faces and non-faces that seem hard to classify correctly. Why do you think they are difficult to classify?

Fig 3. The faces were the algorithm failed to classify them correctly.

Fig 4. The non-faces were the algorithm failed to classify them correctly.

There are similarities between many of these pictures which seems to imply that there are recurrent issues with identifying these as faces. Some of the possible issues we found were glasses, different angles of the face, low brightness and abnormal contrasts.

All of these contribute to a wrong classification probably because the values in the pictures are outliers. For instance, most pictures in the training set are probably well lit, so the classifiers are not fitted to low brightness photos. The same applies to all qualities that are not well represented in the dataset.

Regarding the case of photos classified as faces were there were none it seems like they match the characteristics of the haar features in some ways e.g. they have strongly contrasted edges along places where faces often are positioned in the training data.

Q6. Are your results reasonable? Can you think of any way to improve the results?

Considering the proportions of train data to test data (1000 training samples vs 12000 test samples) the resulting accuracy 94% is reasonable. As said previously we could increase the amount of training data to possibly achieve a higher accuracy.

Q7. Can we expect perfect results? Motivate your answer.

As long as there are circumstances where the qualities of the photo in the test set is not represented in the training set, we cannot expect perfect results. There are theoretically infinite ways to vary parameters in a photo such as angle, brightness, motive, the person itself etc. A way to counteract this problem would be to preprocess the data by adding duplicates images that are scaled, rotated and more. There is also the matter of who is represented in the faces training set; if we don't include e.g. different skin colors here it might affect the performance of the algorithm when testing.