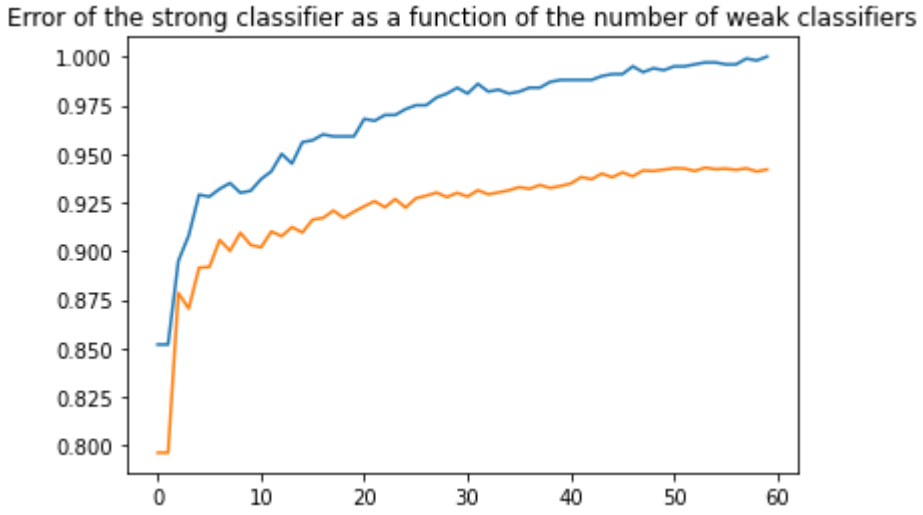
TBMI26 - Assignment 3

Ludvig Knast - ludkn080
Erik Jareman - erija971

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.

Number of Haar-features: 150
Number of train images: 1000
Number of test images: 11788
Number of classifiers: 60

Final Train acc: 100%
Final Test acc: 94%



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 started out with 30 classifiers and noticed that we couldn't reach over 93% accuracy, even though we increased the number of Haar-features as well as the number of training images. Therefore we slowly increased the number of classifiers until we always got a 93%+ result on every run, and ended up with 60 classifiers. This turned out well for us as it always performed well without having to train for too long of a time, (i.e always passed 93% with margin but still trained fast which we were happy with).

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.

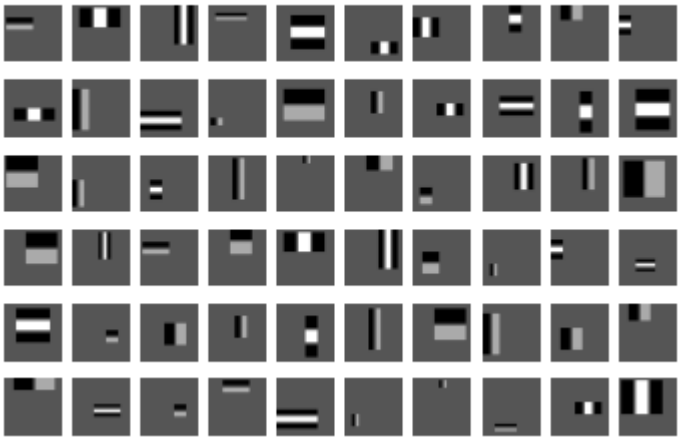
Final Train acc: 100%
Final Test acc: 94%

The result is kind of expected as the optimized strong classifier is biased towards the training data, however it generalized quite well and managed to perform well on the test set.
As mentioned before we tried to hold the number of classifiers as low as possible while still managing to perform well on every run, which ended up with being 60. More classifiers would most likely make the optimized classifier better on the test data (until starting to overfit) as it would have more features to use but would lead to longer training time. We chose to have about 10 % of the data as training and the rest as test (with around 13000 total images). This made sure we had enough images to not overfit towards some small set of images which might not be a good representation of the data, while at the time making sure to not use too many as it could lead to longer training time and not leaving enough images left for testing which could make the scores not accurate/good representation. The number of Haar-features is one we thought a lot about. By increasing the number of features one can expect to accomplish better weak classifiers, thus better accuracy and generalization, however it increases training time as we are looping through all the haar-features for every weak-classifier construction. We wanted at least double the amount of features than classifiers so we made sure that the strong classifier did not have to include “bad” features that could be created as they were randomly generated. We ended up with 150 Haar-features which made sure the weak classifiers had enough variety to choose from as well as keeping training time within reasonable length.

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?

Some features seem to be good for classifying the vertical side of a face where the side of the face is more bright than the background besides it, for example feature 2,2. Similarly features like 5,2 could be used to see the contrast between the mouth and eyes where the area around the mouth is often darker. Overall it seems to be a theme in the features where brighter parts are in the middle of the image and have a contrast with the background. This could be useful as faces often are in the middle of the images and appear brighter than the backgrounds.

Choosen Haar features



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?

There can be many reasons why these faces got wrongly classified. Image 1 doesn't have eyes due to a mix of noise and glasses which could have been problematic. Half of image 2´s face is gone due to angle and shadow and image 3 is just super pixelated. These are just examples of what could have been the problem, however it is hard to say what "went wrong".

These faces got classefied as non-faces

The miss-classed faces are a bit more difficult to grasp but what they all have in common is that they show some clear contrast between a bright object and a dark background. This combined with "lucky" matches to the chosen haar-features (used in the weak classifiers) could be the reason for misclassification.

These non-faces got classefied as faces

Q6. Are your results reasonable? Can you think of any way to improve the results?
We think the results are reasonable since the classifier is good, but not perfect, and perfect results are not to be expected (Q7). One way to get a better result could be to use more data to train the strong classifier, to make sure of a broader variety of images where trained upon as well as not overfitting to the training data. More classifiers and haar-features to choose

from would also most likely improve the results but would be more computationally heavy both during execution but also during the use of the final strong classifier.

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

With enough training data and weak classifiers we can get very good results. But expecting a perfect result is not reasonable since it is possible that some face-images (not used in training) do not really follow the “pattern” of the training data, and these images are going to be hard to classify no matter how good our classifier is. Simply, the immense diversity in all possible face-images are just too big to expect perfect results.