TBMI26 – Computer Assignment Reports  
Boosting

Deadline – March 15 2020

Fredrik Johansson

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 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 = 6000

Number of test data = 6 788

Number of Haar-Features = 300

Number of weak classifiers = 80

The blue line illustrates the accuracy of the training data and the red line illustrates the accuracy of the test data.

En bild som visar karta, text

Automatiskt genererad beskrivning

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

I used 80 weak classifiers in the training phase, and I used them all for the final strong classifier for simplicity. However, according to the image above, 16 weak classifiers would be enough to provide a result within the accuracy constraints.

1. **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.**

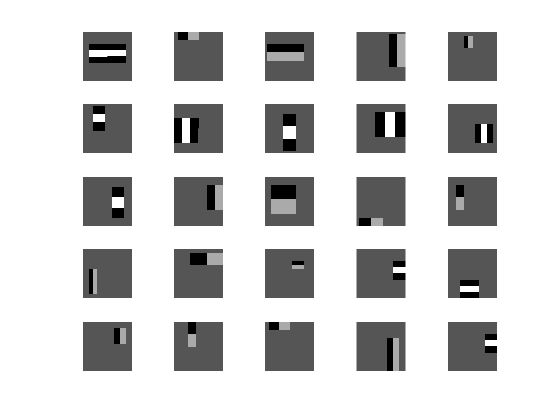
Train accuracy = 97.9%

Test accuracy = 95.6%

I choose the parameters by trail and error using a low amount of training images to speed up the process. By increasing the number of Haar-features the probability to generate good Haar-features increases which makes the weak classifiers better resulting in faster convergence. By increasing the number of weak classifiers, the strong classifier gets more robust as well as it approaches the convergence value.

1. **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?**

They are useful since they represent some distinct feature of our faces. For instance, the Haar-features with a white vertical stripe in the middle represent the highlight of our nose since our faces usually are darker around the eyes compared to the nose. When this Haar-features traverse over the image it will give a strong signal if the image contains a highlighted nose and darkness around the eyes.



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?**

They could be difficult to classify because they have bad contrasts resulting in an even signal when the Haar-features traverse the image. They could also be hard to classify since some of the images contain persons with glasses. Glasses interrupts usual highlights of a person’s face as previously discussed which may produce weak signals from the classifiers. For the non-faces it could be because the images contain patterns which resembles human faces.

**En bild som visar foto, vit

Automatiskt genererad beskrivning**

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

Yes, they are reasonable if the goal is to achieve good accuracy. If the goal is to create a secure face recognition program, then the results are not as reasonable. In that case we would want to ensure that the program have a very low number of false positives at the expense of accuracy.

To improve the result, we could for example use data augmentation on the persons which were hard to recognize. We could also choose Haar-features in a less random way by determining beforehand which would be suitable for faces.

1. **Can we expect perfect results? Motivate your answer.**No we can’t since there will always be edge cases. If we would consider many of these edge cases the algorithm would most likely overfit and detect faces in many non-face images leading to worse results. It is also worth to note that humans are exceptionally good at recognizing faces, it would be a great feat for a computer to exceed this ability.