TBMI26 – Computer Assignment Reports
Boosting

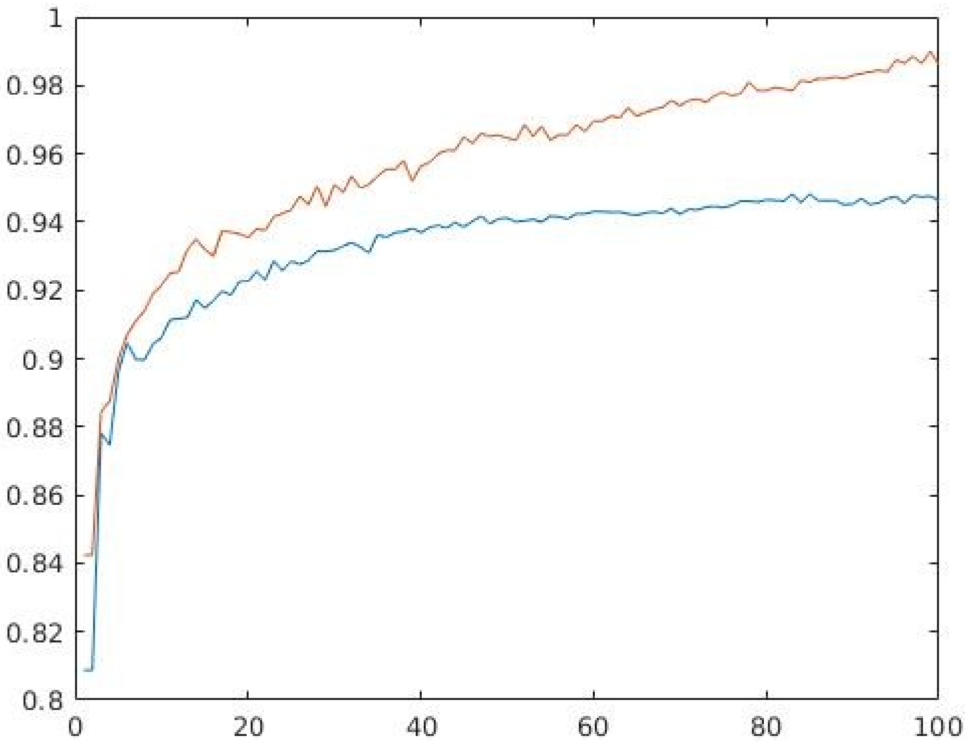
Deadline – March 15 2019

Author/-s: Stefano Toffol (steto820)
Martin Smelik (marsm914)

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.**

   The plot contains the accuracy for both train (red line) and test (blue line) data-set versus the number of weak classifiers. In order to generate this plot we used the code present in the attached scripts.
   In particular, we trained the algorithm on 2000 images, a relatively small number if compared to the total amount of available images (12788). However we noticed that extending the dimension of the training sample increases the computation time without changing much the final estimates.
   The images where equally split between faces and non-faces in both data-sets (so 1000 and 1000 for the training, 5394 and 5394 for the test). The number of Haar features generated was set equal to 100. Therefore the plot shows the accuracy trend using just one until all possible features.



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

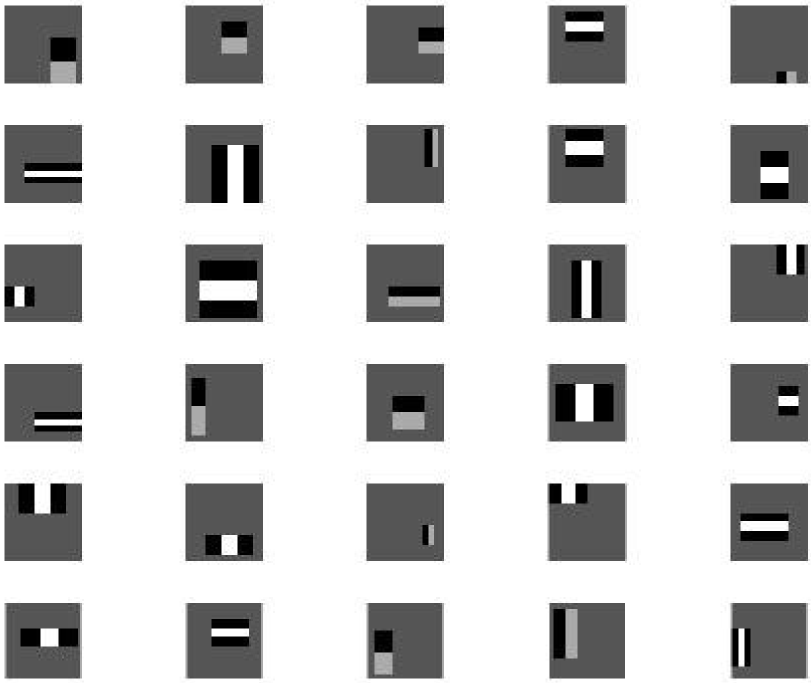
   In order to understand which combination of weak classifiers gave the best result in terms of classification, we trained the model using from one up to 100 weak classifier. We decided to start from just one weak classifier in order to study how the performances of the model change progressively with the increase of the number of weak classifier. We used maximum 100 classifiers (equal to the number of Haar filters created) since, supposing all the individual features are giving interesting contribution to the phenomena, the Adaboost will ideally result in a weighted linear regression with all the variables (saturated model). Including more weak classifier seems excessively redundant.

   Finally we decided to keep 60 weak classifiers to get our final strong classifier. The reason behind it is the plot shown in the previous task: as we can observe both accuracy (train and test data-set) after including 40 weak classifiers increases much more slowly and in particular the one from the test data-set (our main focus in evaluating the performances of a model) seems to not increase substantially. In particular, it takes approximately 20 more classifiers (from 40 to 60) to increase the accuracy of 0.005 (from 0.9306 to 0.9354) but more than double to increase the performances again of 0.005 (from 60 to 100 weak classifiers the accuracy raises from 0.9356 to 0.9397). We therefore decided to stop at 60 because the increase in complexity does not seem to be worth the increase in performances.

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

   The accuracy on the test data-set using 60 weak classifiers to get the final strong classifier is equal to 0.9354.

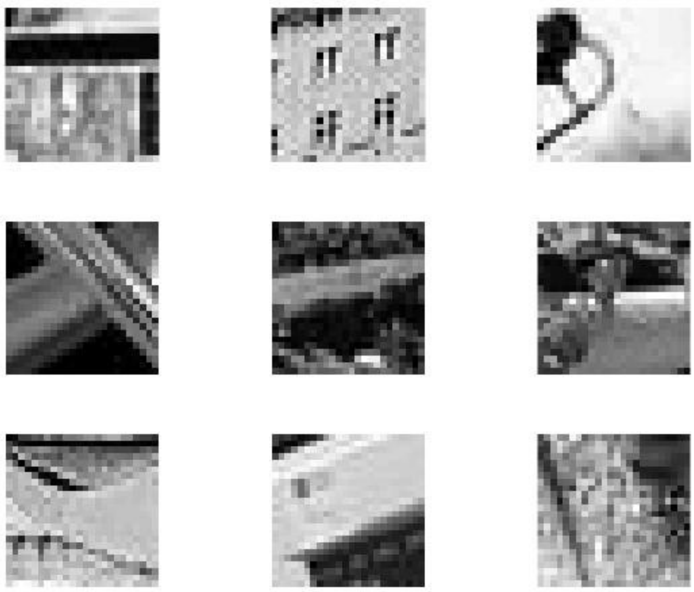
4. **Plot the Haar-features selected by your classifier (one for each weak classifier). If you have many weak classifiers, select some representative subset.**



   As mentioned above, the number of weak classifiers selected was actually 60, but they could not fit into the one image. We therefore decided to plot the first half of them (30), which are the ones contributing the most to the predictions.

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?**

   We can see misclassified faces and non-faces below. Reasons of the misclassification could be for example having sunglasses (position 2,1), very pale face (3,3) or having quite a dark photo (1,2) and some peculiar types of beard cuts, hair cuts or accessories. On the other hand it is slightly more difficult to notice reasons of misclassification on non-faces images. We think that the windows in the image (1,2) might look a bit like eyes. The other images probably have a combination of pixels which result in some ways similar to the human traits (in (2,3) the dark spots may be interpreted as eyes). It is actually difficult even for human eyes to understand what is shown in the images.



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

   With our code, we managed to get accuracy over 93% on test data and we also can see that the accuracy of the train data grows almost to 100%. On the pictures above we can see that most of the images that were misclassified were difficult to judge also by a human eye. In the test and train error graph we can see some small oscillations, but this is caused by combining different weak classifiers. However we can see that the trend of the misclassification rate is decreasing.

7. **Can we expect perfect results? Motivate your answer.**

   We can of course expect better results to what we got simply by increasing the number of train images, the number of Haar Features and most importantly the number of weak classifiers. However getting a perfect results in the test data is almost impossible because of the big complexity and varieties of the images. Moreover we could be able to get 100% accuracy from training data but that would lead to overfitting in the test data.