Statistical Machine Learning 2021

Final Project Specification

In the final project, you are supposed to select two classification algorithms and apply it to the two digit recognition problems:

- Data from all persons with a training set that includes data from all persons ("all persons in")
- Data from all persons where the training set does not include data from the persons in the test set ("disjunct")

For both problems all persons shall be included in either training or test set.

Of course you are also allowed to combine algorithms to boost performance.

Take the following issues into account:

1) Give a brief intuitive summary of the two chosen classification algorithms. Identify the critical parameters.
2) Describe the data you chose.
3) Describe your preprocessing (dpi, PCA, centering, smoothing, normalization). Also show images.
4) Optimize critical parameters (discussed under 1) on a smaller set and document this optimization process.
5) Do a proper cross validation and indicate also mean and variances for all problems. Describe results on test and trainings set and reflect on overfitting.
6) Give information about the computational time required.
7) Analyze the results and give proper explanations.
8) Give an indication what could be further improved.

It is very likely that we will start with discussing your report in the exam. Hence you will probably have a good starting position, when the report is done properly.

You can make group reports. If you want that the written report will be part of the evaluation, you need to indicate who has written what part. Otherwise only the oral exam can be taken into account.