You should not use any other data other than those that we provide you. You are also not allowed to hand-label the given data. You can make at most 200 submissions on the validation dataset and 10 on the test dataset.

1 Introduction

In this project you will classify images into 10 different classes. We have processed the images and extracted a total of 2048 features, which you will use to build your classifiers.

2 Input and output specification

The data is provided in the HDF5 format. For reading this format please a suitable library, e.g. H5py for Python, or the h5read function in MATLAB. The features are stored under the dataset «data» as a matrix with one data point per row. The training file has an additional dataset «label» with the corresponding labels, which are given as integers 0–9. We provide you with the following files.

- train.h5 — The features and labels for training.

- validate.h5 — The features of the validation data.

- test.h5 — The features of the testing data.

3 Evaluation and Grading

You have to provide two files of predictions — one for the validation dataset, and one for the testing dataset. You should produce files that contain one integer from 0–9 per line for the corresponding row. If the true labels are $\mathbf{y}$ and your predictions are $\mathbf{y}'$, the submissions are evaluated using the classification error

$$\ell(\mathbf{y}, \mathbf{y}') = \frac{1}{n} \sum_{i=1}^{n} [y_i \neq y_i'].$$

We will compare the score of your submission to two baseline solutions: a weak one (called "baseline easy") and a strong one (called "baseline hard"). The grade is computed as the *maximum* of the following two percentages.

- Perc$_A$ — Equal to 50% if you are performing at least as good as the easy baseline on the *validation set* and 0% otherwise. Hence, by looking at the ranking you can immediately know if you will receive at least 50% of the grade.

- Perc$_B$ — Let the scores of the easy baseline and the hard baseline on the *test set* be BE and BH respectively. If we denote the score that you reach on the *test set* as E, then you will obtain a score of

$$\text{Perc}_B = \left(1 - \frac{E - BH}{BE - BH}\right) \times 50\% + 50\%.$$

If you perform better than the hard baseline, you will receive Perc$_B = 100\%$.

3.1 Report

You are requested to upload a ZIP archive containing the team report *and* the code. We have included a template for LATEX in the file `report.tex`. Please keep the reports brief (under 2 pages). If you do not want to use LATEX, please use the same sections as in `report.tex`. Reports are uploaded on the same page as the test set submissions.

3.2 Deadline

The submission system will be open until **Sunday, 26.04.2015, 23:59:59**.