1 General remarks

Today's exercise is similar to last week's. BUT: another problem has been added to the data that you need to fix. So: focus on fixing the problems in the data today, and train a DNN!

2 Data loading and preparation

Download the data using the link on the e-learning site of the course. On LinuxLab computers, pass the provided link to the wget command to download the file. Unzip the file, it contains a directory called DATA.

3 Data to numpy

The DATA directory contains a lot of files, all with filenames of the form "X_Y.png", where X is the class and Y is a global file number. From these files, create numpy arrays X and T. Hints (google these functions!):

- use the method os.listdir() in the *os* module to get a list of file names in a directory!
- use the string method ""split" to process the filenames
- use the function "imread" from the *imageio* module to read a png file into a numpy array (you may need to install that first: *python3 -m pip install imageio*).

4 Exploring the data

Check whether ranges are the same for all image pixels, whether data re block-sorted, and whether all classes are equally represented. Propose solutions if any of these problems occur and implement them if possible!

5 Training a DNN

Use the DNN training sample code and apply it to the corrected and uncorrected data. Implement a confusion matrix computation for the trained DNN. What do you observe for corrected and uncorrected data?