Machine Learning (lecture) Fall 2014: Exercise sheet 1

Preparation. (I guess that's the hardest part). Get Matlab (or Python) running. Matlab runs on the Teaching Lab Linux computers (tlab036 - tlab071), but you might find it more convenient to install a local copy and start it through the licence server (ask Dr Achim Gelessus, the CLAMV director, for help and a software CD).

Once you have Matlab or Python running, download the data and some (Matlab) convenience scripts from the course homepage (link in the "Schedule" part). Run the `drawDigitsML.m` script – if the digit thumbnail pictures appear on your screen, the hardest part is done.

Exercise. Invent a feature which you think is good at distinguishing "zero" images from "one" images. Use this feature to transform the "zero" and "one" images (200 each) to simple scalars, and use those to classify the "zero" and "one" images with the bin-counting/histogram method presented in the lecture (lecture notes Section 2.3). Train your classifier on a training set composed of the first 100 (out of 200) samples of each class, and test it on the respective other 100 samples. Compute the misclassification percentage on the test set. You may try to optimize your procedure by playing with the bin size and by finding better-working features. Note that a Matlab routine which essentially implements this exercise for a particular feature is contained in the script `FeatureHistogram.m`. You may simply re-use this script, plugging in your newly designed feature map.

Deliverable: a Matlab or Python script which takes as input the original dataset mfeat-pix.txt, returns a misclassification percentage, and draws a histogram picture like Figure 2.2 in the LN. Inside your script write a short comment text explaining the design strategy for your feature in plain English. Send the script to h.jaeger@jacobs-university.de and the course TA Dzmitri Bahdanau d.bahdanau@jacobs-university.de by Friday, September 12, midnight.