Machine Learning for Applications in Computer Vision: Week 3

Date: Friday, May 8, 2015

Please hand in your **printed(!!)** reports next week (15.05.2015), before the lecture.
Please don't forget to put your names and the GitHub link on the report.
Exercises with an asterisk (*) require written answers. You don't have to write down anything about the other exercises.

Coding

Please code in MATLAB. We will use the same data as used in the earlier exercise (MNIST).

Exercise 1: Boosting

1. Make yourself familiar with the Boosting framework in MATLAB, in particular the command `fitensemble`
2. Load the training data (images and labels) into your MATLAB program. You might have to unpack the data set and convert it to a reasonable format.
3. Train the default AdaBoost on the first 1,000 samples.
4. Check the model you trained against the test set (NEVER do training with the test set!).
5. Play with the number of learning rounds. For visualisation you can plot the training error for each given number of learning rounds, e.g. for 100, 200, ... until 1000.
6. Now run your training on the entire dataset and check it against the test set.
7. Report the methods used and your results.
8. Compare your result with other available Bossting methods, eg. LPBoost

Exercise 2: Gaussian Process Classification

1. Download and install the code from
   `http://www.gaussianprocess.org/gpml/code/matlab/doc/index.html`
2. Run the demo script
   `http://www.gaussianprocess.org/gpml/code/matlab/doc/demoClassification.m`

3. Play around with the parameters. Change the kernel, the inference method and the likelihood function. Evaluate the classifiers on a new test set that is produced in a similar way than the training data, i.e. Gaussians with similar mean and covariance matrices.

4. Also evaluate the classifiers with respect to their uncertainty estimation. For that, split the test data after evaluation in two subsets: one where the classifier was correct and one where the classifier was incorrect. Then, for each subset plot a histogram of uncertainties with 10 bins. The uncertainties can be computed from the predictive probabilities p using the *entropy*:

$$h = -(p * \log(p) + (1 - p) * \log(1 - p))$$

5. Do the same for the Boosting classifier and compare the histogram plots.