

Machine Learning and Pattern Recognition exercise 1: recognition of human faces

General information

Contact information: Mika Murtojärvi, room 450H (office hours: Wed 14:00-15:00), email mianmu@utu.fi. You can also contact me via Moodle messages.

Grading: 4 points, work can be done alone or in pairs. Both members of a pair have to submit the same document.

Exam qualification: You can take the exam before returning the exercise but the exam will not be evaluated before returning the exercise.

Deadline: Tue May 22nd 2018 at 23:55 (full points), June 19th at 23:55 (exam qualification with no points).

Penalty from missing the deadline: 1 point / week.

Tools: Python or R preferred, Matlab ok, others: please ask.

Goal: Learn to extract features, make a classifier and estimate its quality.

Python for beginners: [gettingStartedWithPython.pdf](#), [pythonInstallInstructions.pdf](#)

Data set: Extended Yale Face Database B (cropped)

- vision.ucsd.edu/~iskwak/ExtYaleDatabase/ExtYaleB.html
 - o When not available, try <https://seafire.utu.fi/f/2d94c45d1c214f4da640/>. If you cannot access the data, please contact the supervisor

Tasks

- 0 points: install the required tools (python, R, Matlab) and download the data.
- **1 point:** Read articles or web pages about face recognition. You may try searching for “face recognition techniques”. Some example papers about the subject: www.mdpi.com/2073-431X/5/4/21/pdf, <https://pdfs.semanticscholar.org/d90d/911f98f40a480074bc23ccd57a01d37c8580.pdf>, https://www.researchgate.net/profile/Hamid_Arabnia/publication/220635738_A_Survey_of_Face_Recognition_Techniques/links/00b49526665d9eab4f000000/A-Survey-of-Face-Recognition-Techniques.pdf. There is also a very easy to read web page <http://colah.github.io/posts/2014-10-Visualizing-MNIST/>, although it is for recognizing hand-written digits.
 - o To get points from this step, you should write 1-3 paragraphs about a freely selected technique. Neural networks are not allowed because they are the subject of exercise 2.
- **1 point:** Feature extraction and preprocessing. Report what you did. You may select the method freely, including using the raw pixel data as feature vectors, scaling to lower resolution, using PCA (eigenfaces) etc. Methods covered in the lectures (or in the course Data Analysis and Knowledge Discovery) are preferred. If you use other methods please include references / links describing the method.
- **1 point:** Classification using methods other than neural networks. Among the possibilities are naïve Bayes, k-NN, logistic regression, SVM, others covered in the lectures.
- **1 point:** Report the overall classification accuracy, confusion matrix and a few example images of failed classifications

- The 38x38 confusion matrix may be too big to display in a numeric form. Consider using a color-coded image instead. There are easy-to-use tools for that in python.
- For classification accuracy, you might also try reporting results separately for different lighting conditions to see whether images taken under certain conditions are harder to recognize than others.

If you have a computer with very little memory or a slow processor, you can do the exercise with only a subset of the images. In this case you should remove pictures randomly. It is forbidden to intentionally select a subset that is easy to recognize. Be prepared to show your selection code if you do not use all images in your work.

Note: Poor results are not an issue in grading but I still encourage you to try several methods in order to learn more about the subject.

Submit the report via Moodle in pdf format. The document should include your name. If done in pairs, both names should be included. **Both members of a pair must return the (same) report separately.**

Points decline after the deadline...

max points	Date
4	May 22
3	May 29
2	June 5
1	June 12
0	June 19
No exam right	...