

## Machine Learning, Spring 2018: Exercise Sheet 5

*This is a programming exercise. It will be graded and the grade counts toward the course grade. **Join into groups of two** and submit a single solution per group, indicating the group members' names on the report sheet. You can use Python or Matlab.*

*Please send your type-set solutions by email to our two TA's Tianlin Liu ([t.liu@jacobs-university.de](mailto:t.liu@jacobs-university.de)) and Tayyab Mateen ([t.mateen@jacobs-university.de](mailto:t.mateen@jacobs-university.de))*

*Deadline for submission is Wednesday March 21, 23:59 hrs (email sending timestamp). Submissions arriving later (even a second after midnight) will be corrected but not counted for the course grade.*

**Task.** *This task again is on the digits dataset that you can find at*

<http://minds.jacobs-university.de/sites/default/files/uploads/teaching/share/DigitsBasicRoutines.zip> .

The objective of this task is to compress and uncompress images with PCA.

Pick one of the digits (for instance, pick the "3" digits if you like 3's better than other digits) and continue the rest of this task only with the  $N = 200$  images of this digit. Compute a PCA of this size-200 dataset. Do not use a ready-made PCA routine but program your PCA function from scratch, building on an SVD. Repeat the reconstruction described in the lecture notes in Section 5.4 for the following percentages of variances preserved in the reconstruction: 50%, 80%, 95%, 99%, 100%. Document the number  $k$  of PC's needed in each case and generate images of reconstructions for the first 5 digits in your dataset, for each of the percentages. Explain the number  $k$  that you get for the 100% reconstruction. Deliverable: a typeset discussion (say, half a page of text excluding graphics, but can be more text) with nice graphics, and the code that you produced. The code must be minimally documented inline such that the TA's can quickly grasp what you are computing where in your code.