

Machine Learning, Spring 2017: Exercise Sheet 3

Please send your type-set solutions by email to our two TA's Xu He ("Owen") x.he@jacobs-university.de and Felix Schmoll f.schmoll@jacobs-university.de. Join into groups of two or three and submit a single solution sheet per group, indicating the group members' names on the sheet.

Deadline for submission is Friday Mar 17, 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 1 (programming). 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> .

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, but can be more) 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.