Barbara Hammer, Philip Kenneweg Universität Bielefeld

Modern Data Analysis (WiSe 2023/2024)
1. Sheet

Start: Wednesday, 11.10.2023.
End: The worksheets should be solved using Python, in groups of 2-3 people.
Discussion: Monday, 23.10.2023 in den Tutorien.

Information

The worksheets and necessary toolboxes will be made available in the Lernraum "392246 Modern Data Analysis (V)". Worksheets will usually be released every two weeks on wednesday, and discussed during the exercises on monday two weeks later. In order to successfully finish the course, 50% of the available points have to be obtained and each participant has to present his/her results at least once. The monday in between the release and discussion of the sheet will be used to discuss the implementation of the various algorithms presented in the lecture.

Exercise 1:

(*4 Points*)

The Whitening method is a frequently used preprocessing step and plays a major role in the SFA. Use the sheet1_1.py file. Visualize both data sets saved in dataset1.mat, apply the Whitening technique (sklearns PCA has a whitening functionality) and illustrate the result, respectively. What happens? Which kind of information is preserved and which is lost?

Exercise 2:

(*6 Points*)

For the next excersise use the sheet1_2.py file. Here we load a data set consisting of 72 images. These are snapshots of an object which is rotated slowly in front of a camera. The images have a resolution of 128x128 pixels yielding a dimensionality of 128^2 = 16384 if represented as a vector, directly. In order to circumvent this, the images have been preprocessed with the PCA to project them down to 30 dimensions. The low-dimensional data as well as the PCA model are provided in excersise2.pkl.

(a) *(2 Pts.)* Visualize a few of the images by projecting the low-dimensional vectors back to the high-dimensional space (after the projection you can use the command reshape to transform the vector into a matrix and imshow to display the latter). Can you recognize what object it is?

(b) *(4 Pts.)* Apply the (linear) Slow Feature Analysis to the data set (use the sksfa package `https://github.com/wiskott-lab/sklearn-sfa`). Visualize the first two features as a function of the time t. What do these features characterize? What do you see if you plot these features against each other?