Time	Group	Submission in Moodle; Mails with subject: [SMD2023]
Th. 12:00–13:00	A	lukas.beiske@udo.edu and tristan.gradetzke@udo.edu
Fr. 08:45–09:45	B	jonas.hackfeld@ruhr-uni-bochum.de and ludwig.neste@udo.edu
Fr. 10:00–11:00	C	stefan.froese@udo.edu and vincent.latko@udo.edu

Exercise 14 *Principal Component Analysis (PCA) by hand*   **5 p.**

(a) Briefly describe how the Principal Component Analysis works. In words, state the calculations necessary in correct order to perform the principal component analysis.

(b) Calculate the individual steps by hand on the data set:

$$x_1 : [1, 3, 1, 2, 3, 2]$$
$$x_2 : [1, 0, 3, 0, 1, 1].$$

Exercise 15 *Principal Component Analysis (PCA)*   **5 p.**

(a) Generate a dataset with the function `sklearn.datasets.make_blobs`. Use the following settings: n_samples=1000, centers=2, n_features=4, random_state=0. Now plot any two dimensons of the dataset in a scatterplot.

(b) Now apply the Principal Component Analysis (PCA) to the dataset generated in a). Use the package `sklearn.decomposition.PCA` for this purpose. What are the eigenvalues of the covariance matrix? How do you interpret the eigenvalues?

(c) Now histogram x' in each dimension and plot x'_1 and x'_2 in a scatterplot.