

Exercise 09

Principal Component Analysis

Deadline: Please, hand in your protocol in pdf format to saleksic@zedat.fu-berlin.de by **Thursday, 14th July, 10.15 a.m.**

9.1 Principle Component Analysis (100 P)

- (a) Generate a scatter plot of the trajectory available in Files section. (10 P)
- (b) Calculate the mean values $\langle x \rangle$, $\langle y \rangle$ and mark on the axes of the plot as the constants. (10 P)
- (c) Calculate the covariance matrix $C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle$ with $i, j = 1, 2$. In Python, you can compute the covariance matrix with the function `C=np.cov(x,y)`. (15 P)
- (d) Calculate the eigenvalues and eigenvectors of the covariance matrix. (10 P)
- (e) Plot the eigenvectors into the scatter plot centered at $(\langle x \rangle, \langle y \rangle)$. (15 P)
- (f) Project the trajectory onto the eigenvectors following this procedure: (30 P)
 - (a) Transpose the eigenvector matrix.
 - (b) Subtract the mean values $\langle x \rangle$, $\langle y \rangle$ from the trajectory and transpose the new dataset.
 - (c) Multiply the eigenvector matrix by the adjust dataset to get the new dataset respect to the principal components.
 - (d) Plot the new trajectory (PC1 vs. PC2).
- (g) Find the time points in which the projected trajectories have the largest and the lowest value. Mark the corresponding data points in the scatter plot. (10 P)

9.2 Files

Trajectory

<https://www.dropbox.com/s/9g2oh10h31xuyb3/trajectory.txt?dl=0>

PCA tutorial

http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf