*Institut für Mathematik*
*Abteilung Numerik*
*Jana de Wiljes*

# 9. Problem sheet for **Statistical Data Analysis**

**Exercise 1** (6 Points)
Determine the eigenvalues and corresponding eigenvectors of the following matrices

$$a) \quad \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \quad b) \quad \begin{pmatrix} 1 & 2 & 3 \\ 2 & -4 & -2 \\ 3 & -2 & 1 \end{pmatrix}, \quad c) \quad \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & -1 & 2 \end{pmatrix} \tag{1}$$

Which of the matrices can be diagonalized?

**Exercise 2** (10 Points)

- Download the Iris data set from the following webpage:

  $$\text{http://archive.ics.uci.edu/ml/datasets/Iris} \tag{2}$$

  The data contains 3 classes of 50 instances each, where each class refers to a type of iris plant. There are 4 different attributes describing the data.
- Use principal component analysis to transform the data to a lower dimensional space in order to visualize it. Find an appropriate subspace by computing eigenvalues and eigenvectors
- Project the data onto its first two principal components and plot the results. Use the class label to color the instances when plotting it in the 2-dimensional space.