Your task for the "clustering part" of the practical exercise is to carry out a data analysis exercise, using different technologies/methods presented and discussed in class, i.e.:

1. PCA
2. Classical MDS
3. Sammon's Mapping
4. SOM
5. GHSOM
6. LLE
7. Isomap
8. SNE
9. t-SNE

To this end, you will find attached various data sets to experiment with:

ds1_features.txt	1000 items,	192 dimensions
ds2_features.txt	969 items,	49820 dimensions
ds3_features.txt	224 items,	1244 dimensions
ds4_features.txt	201 items,	2730 dimensions

Each line represents one data item. The different values for the variables in each dimension are separated by a double space.

In **groups of three people**, please perform the following tasks:

Familiarize yourself with the topics discussed in lecture (this is also a good exercise for the final exam!). You may (and should) use different sources, such as the lecture slides, papers you find on the topics, Web pages, etc.

Look for implementations of the chosen techniques. For each of the methods, there is a bunch of implementations available online (e.g., in C++, Java, Python, Matlab, R), likely also in your preferred programming language. Have a look at them and experiment with the data sets. As a bonus exercise you may also re-implement one or another technique by yourself.
For the parametric models, try out different settings (depending on the technique, for example, initialization method, number of neighbors to consider, stress function, grid size, learning rate, distance measure, etc).

Hint: If you run into computational problems, you may also think of making use of the data sets "intrinsic dimensionality".

Your analysis should be guided by the following questions:

- **Can you find interesting structure in the data sets?**
- **Can you determine groups of similar items (clusters)? Why are they similar (i.e., which dimensions tend to be responsible for their similarity)?**
- **Can you find correlations between data dimensions?**

After having answered these questions, you may want to take a look at the provided class labels (ds[1-4]_classes.txt), which indicate for each of the data instances in ds[1-4]_features.txt the corresponding (numeric) class.

Investigate whether the groups of similar items you identified beforehand correspond to the different classes. **Were you able to determine specific clusters? Did you experience any surprises when analyzing the class labels?**

Guided by the tasks detailed above, prepare a **presentation of no more than 15 minutes**, including sample plots of your analyses and thoughts/answers to all raised questions.