Exercises for
Pattern Analysis Programming
Dalia Rodriguez-Salas
Task 7: June 21, 23, 28 and 30

Dimensionality Reduction

Not only the human intuition breaks down in higher dimensions, but also for some algorithms (e.g. clustering) it is favourable to bring the data to a lower dimension. That means, that we want to find a good representation in a lower dimensional space, that reduces the dimensions of the feature vectors in such a way, that the (possible) data-structure is preserved. In this exercise, we will look at different dimensionality reduction methods and compare them with each other.

Exercise 1

In this exercise we will compare the following algorithms with each other:

(a) Principal Component Analysis (PCA) (Lecture 2h, Bishop Sec. 12.1.1)

(b) Multidimensional Scaling (MDS) (Lecture 2i)

(c) Iso Map (IM) (Lecture 2j)

(d) Laplacian Eigenmaps(LE) (Lecture 2j)

You can implement them by yourself, or just use the corresponding function from the `sklearn` python-library.

To get a *feeling* for the algorithms and be able to visualize the results, we will restrict ourselves in this task to 2D and 3D densities / feature vectors and reduce them by one dimension only. Please test and compare the algorithms qualitatively on the following 2D and 3D data sets:

- One Gaussian distribution (2D or 3D)
- Multiple Gaussian Distributions in 2D
- Multiple Gaussian Distributions in 3D
- Two circles in 2D (`sklearn.datasets.make_circles`)
- Swiss roll in 3D (`sklearn.datasets.make_swiss_roll`)

Figure. 1 gives an example for each data set.

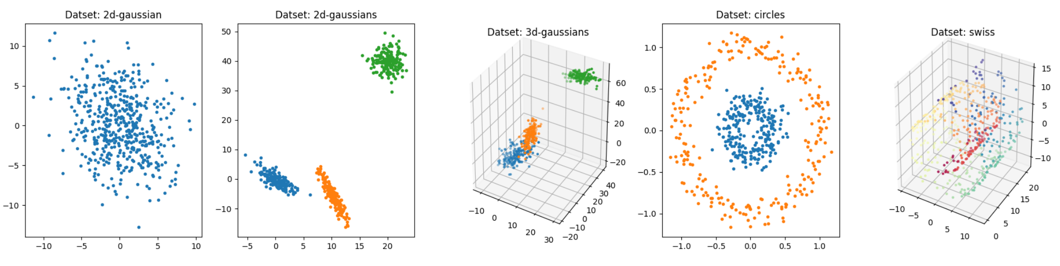

Figure 1: Five different data sets in 2D and 3D.

What results do you get by applying the different algorithms to the different data sets? What result (distribution combined with algorithm) might have surprised you? Play around with the parameters (e.g. number of samples, amount of noise, etc.).

Please post a figure of one of the data sets and the respective lower dimensional representations to the forum and add a short text about it.

Comments:
We ask for only one figure per group. Please also state your group name.