Exercise 11

Cluster algorithms

Deadline: Please hand in your protocol in pdf format by Thursday, the 11th of July 2019, 10 am to jan.joswig@fu-berlin.de and marco.manni@fu-berlin.de.

An example 2D-data set for this exercise can be found under
login.bcp.fu-berlin.de:/home/janjoswig/MD19/Ex11/p12.npy .

You can download the cnn python module from github
https://github.com/janjoswig/CNN.git
or you copy the cnn directory from
login.bcp.fu-berlin.de:/home/janjoswig/MD19/Ex11/ .

11.1 *k*-means

1. Implement a k-means procedure according to the standard algorithm. Limit the maximum number of iteration steps und choose a convergence criterion.

2. Let the algorithm cluster a 2D-data set into 2, 3, 4 ... clusters and plot the result of the partitioning, so that you can see the location of the cluster centers and which point belongs to which cluster. Is the result reproducible?

11.2 Common nearest neighbours

1. “Install” the cnn python module by copying the folder cnn/ including cnn.py and __init__.py to your computer (e.g. under ∼/CNN/). The directory containing the cnn module directory needs to be in your PYTHONPATH (export PYTHONPATH=$HOME/CNN:$PYTHONPATH). In your python session you should than be able to import cnn.

2. Cluster a 2D-data set using the cnn algorithm and find the radius R and common nearest neighbour count N that gives the best result. Plot the result of the partitioning.