Clustering, Self Organizing Maps and Embedding

In this exercise we fit one- and two-dimensional Self-Organizing Maps (SOMs) to toy data.

11.1 1d Self-Organizing Map for 2d data (4 points)

(a) Generate $p = 1000$ data points uniformly distributed in the rectangle $\boldsymbol{x} \in [0, 2] \times [0, 1]$

(b) Implement a one-dimensional self-organizing map (Kohonen network, online algorithm for SOMs) using a Gaussian neighborhood function

$$h_{qp} = e^{-\frac{(q-p)^2}{2\sigma^2}}$$

(c) Fit different maps with $k \in \{4, 8, 16, 32, 64, 128\}$ nodes (prototypes) to the data.

Note: Anneal both the learning rate ε and the neighborhood width σ. The start value σ_0 has to be large enough to unfold the randomly initialized (scrambled) map in the first iterations.

(d) Plot the final map in the data space, i.e. the locations of the prototypes and their connections, for each number of nodes k.

11.2 1d Self-Organizing Maps for 3d data (2 points)

(a) Download and visualize the data contained in the file `spiral.csv`. It contains data described by three coordinates x,y,z.

(b) Adapt your previous SOM with Gaussian neighborhood function to fit one dimensional maps with $k \in \{16, 32, 64, 128\}$ nodes to this dataset.

(c) Initialize your map as a line along the z axis, i.e. with $x = 0$, $y = 0$, and $z = 0, ..., 5$.

(d) Plot the final maps in the data space.

11.3 2d Self-Organizing Maps for 3d data (3 points)

(a) Visualize the 3d-data in the file `bowl.csv`.

(b) Adapt your previous SOM to fit two-dimensional maps with $k \times k$ (cartesian) grid topology for e.g. $k \in \{8, 16, 32\}$ nodes (depending on your computing resources) to this dataset.

(c) Initialize the prototypes randomly or in an informed way (e.g. arranged as small grid centered on the data mean and spread along the first 2 principal directions of the data).

(d) For at least 2 different values of σ, plot the map (protoype locations and their "connections") (i) at t_0, (ii) an intermediate point, and (iii) in the final configuration in the data space.

(e) Suggest how to visualize the data points in map space (e.g. to do dimensionality reduction).

Total points: 10