Project 2 – SOM and Clustering   Assigned March 20
DUE March 30, 2017

Team assignment.

Step 1: SOM

As a team, look into your data sets from Homework 2 and decide which data set is best prepared/you would rather work with and utilize it for processing with Self-Organizing Map. The choice of network architecture is yours (1D or 2D – which topology in this case) as are the network parameters: learning rate and neighborhood radius and architecture (only positive pull or having both negative and positive).

Deliver a report and the SOM code or script you have used. The report must contain:

1. The SOM architecture – dimensionality, number of neurons, neighborhood size and configuration and reduction function, learning rate value and reduction function, number of epochs.
   15 points

2. Expectations from SOM – number of clusters expected and actually received, performance analysis (pertaining to different learning rate experiments, varying neighborhood experiments and varying number of neurons experiments).
   35 points

3. Results visualized via silhouettes algorithm and interpreted.
   a. Results from all experiments with varying SOM parameters represented with silhouettes.
   20 points
   b. In addition to silhouettes algorithm, verification of the cluster correction against the class attribute in the data set.
   i. Confusion table is required for each experiment performed.
   20 points
   ii. Analysis of the SOM performance in terms of accuracy of grouping based on the confusion tables.
   30 points

4. Discussion on overall SOM performance, results and conclusion on SOM usability for the task.
   20 points

While concise reports are appreciated, the goal is to include all relevant information, particularly emphasizing your results discussion, interpretation of experiments as well as the final conclusion.