Universität Hamburg
Department Informatik, ABe NATS, WTM & TAMS
Wolfgang Menzel, Stefan Wermter, Jianwei Zhang

Contact: Cornelius Weber
Office F-233
Tel 42883-2537

(for groups who focus on the SOM)

Exercises to the module: Algorithmisches Lernen
SS 2012 Sheet 7 (for groups who focus on the SOM)

Due: 23.05.2012

In this exercise you will gather experience with the Self-Organizing Map (Kohonen map) and some data, as well as with setting the network parameters. For your convenience, there is a Python implementation of the SOM available on the MinCommSy: `SOM.py`. Furthermore, some data sets with mostly 10000 data points. The program used to create the data, `data_create.py`, is also available, for example to let you create the data in a format that can be easily read by some other SOM software.

One possible way of visualizing the SOM – in case of a 2D or 3D input space – is by displaying the weights of the neurons as a dot in that input space, and connecting the neighboring neurons by lines. This can be done with the program Gnuplot. After training the SOM, the program `SOM.py` exports two files: `out.dat` constins the points of the neurons in the input space; `out.gnu` contains Gnuplot script commands to draw the connecting lines.

Task 7.1 SOM on a 2D manifold in 3D

Train the SOM with the data in `data_sphere.dat`. A useful map size may be a 10x10 lattice, i.e. in `SOM.py` parameters are set to `dim_a = 10` and `dim_b = 10`.

In the 3D space, display the data as dots and the weights as crosses. In the Gnuplot shell this is done by typing: `splot “data_sphere.dat“ with dots, “out.dat“ with points`

Thereafter, plot the connecting lines between neighboring neurons via:
`load “out.gnu“`

Vary the parameters, such as the map size, learning step size, or the reduction factor with which the neighborhood interaction size `sigma` is being reduced at every learning iteration. Characterize the data. Is there overfitting?

Task 7.2 Convolution of the map in space

Observe how the map folds in space. Train a map with a topology of 1x50 neurons on the data in `data_snake.dat`, which lie in 2D. Also, train a map with a topology of 15x25 neurons on the data in `data_waves.dat`, which lie in 3D.

What can you say about the data if you only look at the behaviour of the map?

Note: to display 2D data, type `plot` in Gnuplot, instead of `splot`, which is for 3D. Display all data for a check.

Task 7.3 Travelling salesperson problem

The file `data_capitals.dat` contains the 2D coordinates of 44 European capitals. The task of the travelling salesperson is to visit all cities once, along the shortest possible route. A possible methodology to solve this task is to train a SOM of topology 1x44 neurons on these data. Do this! Note: in the program you may better increase the variable `batch` to learn longer.

To make the trip a round trip, modify the code so that the network topology is periodic, i.e. the last neuron is neighbour of the first neuron.

Has the problem been solved perfectly? If not, how can you improve the solution?

— see reverse —

Wolfgang Menzel
(menzel@informatik. . .)
Stefan Wermter
(wermter@informatik. . .)
Jianwei Zhang
(zhang@informatik. . .)

Vorlesung:
Mi 8:15-11:45, F-009
Übungen:
Mi 12:15-13:45, F-009

Dienstgebäude
Universität Hamburg
Vogt-Kölln-Str. 30
22527 Hamburg

Öffentliche Verkehrsmittel
Buslinien 181, 281
Haltestelle: Informatikum

Task 7.4 k-Means

Modify the SOM to become the k-means algorithm (in on-line, not batch, version). A representative exercise is to train a network with k=4 neurons on the data in `data_clusters.dat`. Are the data always being clustered correctly, if you train the network several times? If not, what went wrong. Would the problem be solved if using a SOM?

Note for users of the operating system Windows

The homepage of Gnuplot is `www.gnuplot.info`. From there you can get via SourceForge the file `gp460win32.zip`. Unzip this file, and in the sub-directory `gnuplot\binary` you find the executable `wgnuplot.exe`. Execute this and you have it running. In Gnuplot you can use `pwd`, and you can change to the directory of your data like in the following example:
`cd “C:/Users/weber/Desktop/AL/SOM_data/“`