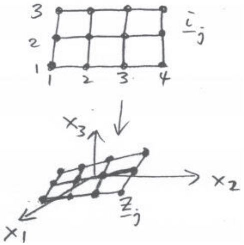
Ch.5 Classification and Clustering

5.6 Self-organizing maps (SOM) [Book, Sect. 10.3]

The *self-organizing map* (SOM) method, introduced by Kohonen (1982, 2001), approximates a dataset in multidimensional space by a flexible grid (typically of 1 or 2 dimensions) of cluster centres.

Many structures in the cortex of the brain are 2-D or 1-D. In contrast, even the perception of colour involves three types of light receptors. Besides colour, human vision also processes information about the shape, size, texture, position and movement of an object. So the question naturally arises on how 2-D networks of neurons can process higher dimensional signals in the brain.



For a 2-dimensional rectangular grid, the grid points or units $\mathbf{i}_j = (l, m)$, where l and m take on integer values, i.e. $l = 1, \ldots, L$, $m = 1, \ldots, M$, and $j = 1, \ldots, LM$. (If a 1-dimensional grid is desired, simply set $M = 1$.)

To initialize the training process, PCA is usually performed on the dataset, and the grid $\mathbf{i}_j$ is mapped to $\mathbf{z}_j(0)$ (in the data space) lying on the plane spanned by the 2 leading PCA eigenvectors. As training proceeded, the initial flat 2D surface of $\mathbf{z}_j(0)$ is bended to fit the data.

The original SOM was trained in a flow-through manner (i.e. observations are presented one at a time during training), though algorithms for batch training is now also available. In flow-through training, there are two steps to be iterated, starting with $n = 1$:

Step (i): At the nth iteration, select an observation $\mathbf{x}(n)$ from the data space, and find among the points $\mathbf{z}_j(n-1)$, the one with the closest (Euclidean) distance to $\mathbf{x}(n)$. Call this closest neighbour $\mathbf{z}_k(n)$, with the corresponding unit $\mathbf{i}_k(n)$ called the *best matching unit* (BMU).

Step (ii): Let

$$\mathbf{z}_j(n) = \mathbf{z}_j(n-1) + \eta\, h(\|\mathbf{i}_j - \mathbf{i}_k(n)\|^2)\,[\mathbf{x}(n) - \mathbf{z}_j(n-1)], \qquad (1)$$

where η is the learning rate parameter and h is a neighbourhood or kernel function.

The neighbourhood function gives more weight to the grid points $\mathbf{i}_j$ near $\mathbf{i}_k(n)$, than those far away, an example being a Gaussian drop-off with distance.

Note that the distances between neighbours are computed for the fixed grid points ($\mathbf{i}_j = (l, m)$), not for their corresponding positions $\mathbf{z}_j(n)$ in the data space.

Typically, as n increases, the learning rate η is decreased gradually from the initial value of 1 towards 0, while the width of the neighbourhood function is also gradually narrowed.

While SOM has been commonly used as a clustering tool, it should be pointed that it may underperform simpler techniques such as K-means clustering. Hence the value of SOM lies in its role as discrete nonlinear PCA (Cherkassky and Mulier, 1998), than as a clustering algorithm.

As an example, consider the famous Lorenz 'butterfly'-shaped attractor from chaos theory (Lorenz, 1963). Describing idealized atmos. convection, the Lorenz system is governed by 3 (nondimensionalized) differential equations:

$$\dot{x} = -ax + ay, \quad \dot{y} = -xz + bx - y, \quad \dot{z} = xy - cz, \qquad (2)$$

where the overhead dot denotes time derivative. A chaotic system is generated by choosing the parameters $a = 10$, $b = 28$, and $c = 8/3$.

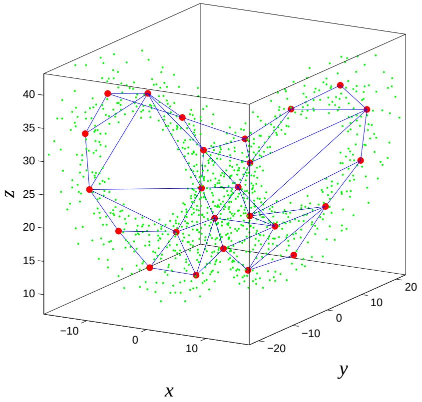


Figure : 2-D SOM with a 5×5 hexagonal mesh fitted to the Lorenz data.

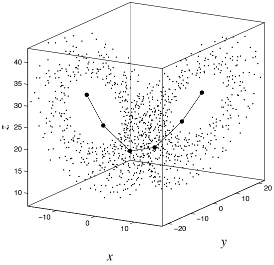


Figure : 1-D SOM with 6 units fitted to the Lorenz data.

How many grid points or units should one use in the SOM?
Underfitting with too few units and overfitting with too many units.

Two quantitative measures of mapping quality are commonly used: average *quantization error* (QE) and *topographic error* (TE) (Kohonen, 2001).

QE is the average distance between each data point $\mathbf{x}$ and $\mathbf{z}_k$ of its BMU.

TE gives the fraction of data points for which the first BMU and the second BMU are not neighbouring units. Smaller QE and TE values indicate better mapping quality.

By increasing the number of units, QE can be further decreased; however, TE will eventually rise, indicating that one is using an excessive number of units.

Functions in Matlab

SOM:

`http://www.mathworks.com/help/nnet/gs/cluster-data-with-a-self-organizing-map.html`

`http://www.mathworks.com/help/toolbox/nnet/ref/selforgmap.html`

net = selforgmap([7 6]);   (uses a 7×6 grid)
net = train(net,x);

References

Cherkassky, V. and Mulier, F. (1998). *Learning from Data*. Wiley, New York.

Kohonen, T. (1982). Self-organzing formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69.

Kohonen, T. (2001). *Self-Organizing Maps*. Springer, 3rd ed. edition.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20:130–141.