Hybrid Intelligent Systems

Lecture 5. Part 3.

Self-organizing maps of Kohonen

The phrase "Kohonen network" most often refers to one of the following three types of networks:

- VQ: Vector Quantization--competitive networks that can be viewed as unsupervised density estimators or autoassociators (Kohonen, 1995/1997; Hecht-Nielsen 1990), closely related to k-means cluster analysis (MacQueen, 1967; Anderberg, 1973).
- LVQ: Learning Vector Quantization--competitive networks for supervised classification (Kohonen, 1988, 1995; Ripley, 1996). Each codebook vector is assigned to one of the target classes. Each class may have one or more codebook vectors. A case is classified by finding the nearest codebook vector and assigning the case to the class corresponding to the codebook vector. Hence LVQ is a kind of nearest-neighbor rule.
- SOM: Self-Organizing Map--competitive networks that provide a "topological" mapping from the input space to the clusters (Kohonen, 1995). The SOM was inspired by the way in which various human sensory impressions are neurologically mapped into the brain such that spatial or other relations among stimuli correspond to spatial relations among the neurons.

VQ

- Each competitive unit corresponds to a cluster, the center of which is called a "codebook vector". Kohonen's learning law is an on-line algorithm that finds the codebook vector closest to each training case and moves the "winning" codebook vector closer to the training case. The codebook vector is moved a certain proportion of the distance between it and the training case, the proportion being specified by the learning rate, that is:

  new_codebook = old_codebook * (1-learning_rate) + data * learning_rate

VQ (2)

- MacQueen's on-line k-means algorithm is essentially the same as Kohonen's learning law except that the learning rate is the reciprocal of the number of cases that have been assigned to the winnning cluster. Suppose that when processing a given training case, N cases have been previously assigned to the winning codebook vector. Then the codebook vector is updated as:

  new_codebook = old_codebook * N/(N+1) + data * 1/(N+1)

VQ (3)

- This reduction of the learning rate makes each codebook vector the mean of all cases assigned to its cluster and guarantees convergence of the algorithm to an optimum value of the error function (the sum of squared Euclidean distances between cases and codebook vectors) as the number of training cases goes to infinity.
- Kohonen's learning law with a fixed learning rate does not converge. As is well known from stochastic approximation theory, convergence requires the sum of the infinite sequence of learning rates to be infinite, while the sum of squared learning rates must be finite (Kohonen, 1995, p. 34).
- These requirements are satisfied by MacQueen's k-means algorithm.

LVQ

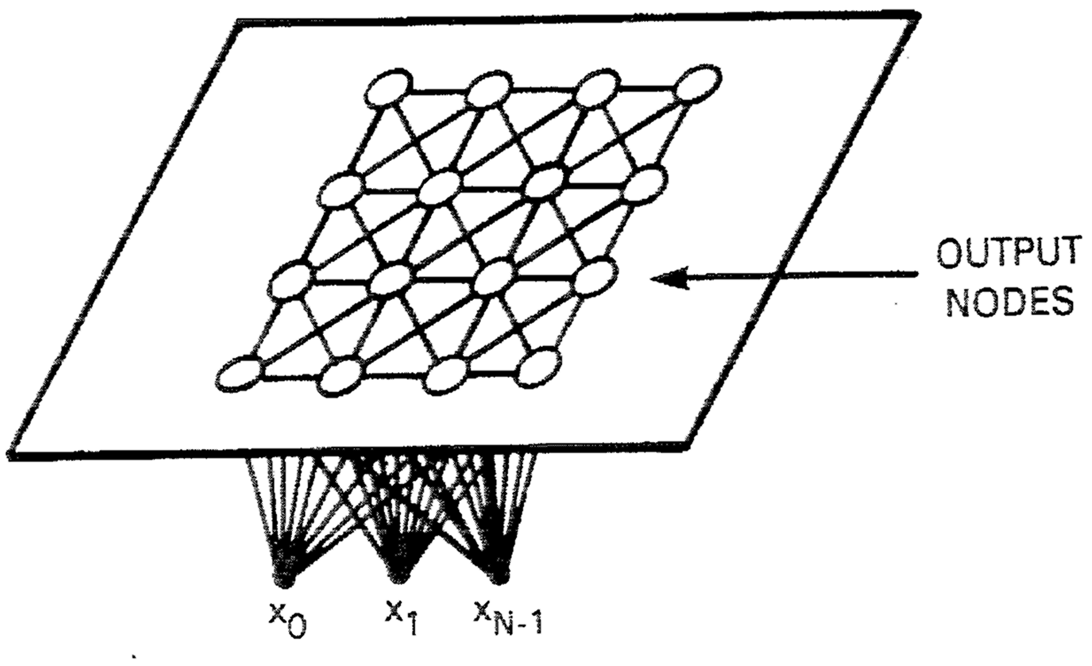
- Ordinary VQ methods, such as Kohonen's VQ or k-means, can easily be used for supervised classification. Simply count the number of training cases from each class assigned to each cluster, and divide by the total number of cases in the cluster to get the posterior probability. For a given case, output the class with the greatest posterior probability--i.e. the class that forms a majority in the nearest cluster. Such methods can provide universally consistent classifiers (Devroye et al., 1996) even when the codebook vectors are obtained by unsupervised algorithms.
- LVQ tries to improve on this approach by adapting the codebook vectors in a supervised way. There are several variants of LVQ--called LVQ1, OLVQ1, LVQ2, and LVQ3--based on heuristics.

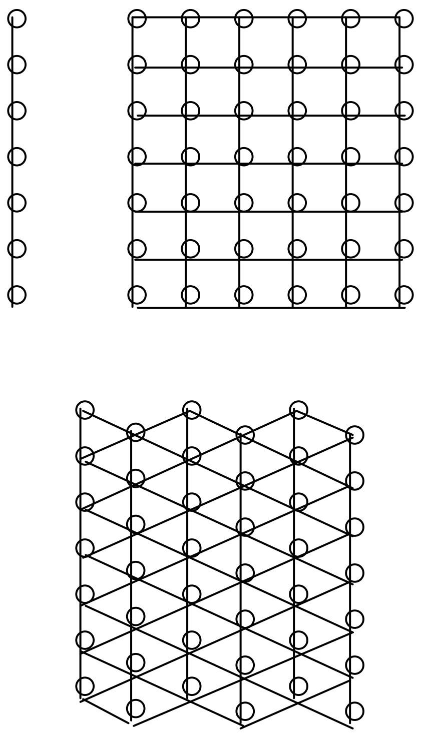
Self Organizing Maps (SOM)

- Based on competitive learning (Unsupervised)
  - Only one output neuron activated at any one time
  - Winner-takes-all neuron or winning neuron
- In a Self-Organizing Map
  - Neurons placed at the nodes of a lattice
    - one or two dimensional (rarely more)
  - Neurons selectively tuned to input patterns
    - by a competitive learning process
  - Locations of neurons so tuned to be ordered
    - formation of topographic map of input patterns
  - Spatial locations of the neurons in the lattice -> intrinsic statistical features contained in the input patterns

Self Organizing Maps

- Topology-preserving transformation





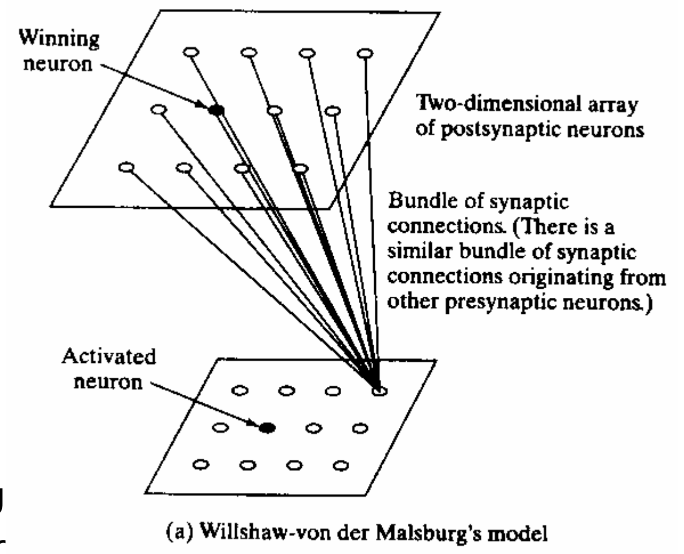
SOM as a Neural Model

- Distinct feature of human brain
  - Organized in such a way that different sensory inputs are represented by topologically ordered computational maps
- Computational map
  - Basic building block in information-processing infrastructure of the nervous system
  - Array of neurons representing slightly differently tuned processors, operate on the sensory information-bearing signals in parallel

Basic Feature-mapping models

- Willshaw-von der Malsburg Model(1976)
  - Biological grounds to explain the problem of retinotopic mapping from the retina to the visual cortex
  - Two 2D lattices : presynaptic, postsynaptic neurons
  - Geometric proximity of presynaptic neurons is coded in the form of correlation, and it is used in postsynaptic lattice
  - Specialized for mapping for same dimension of

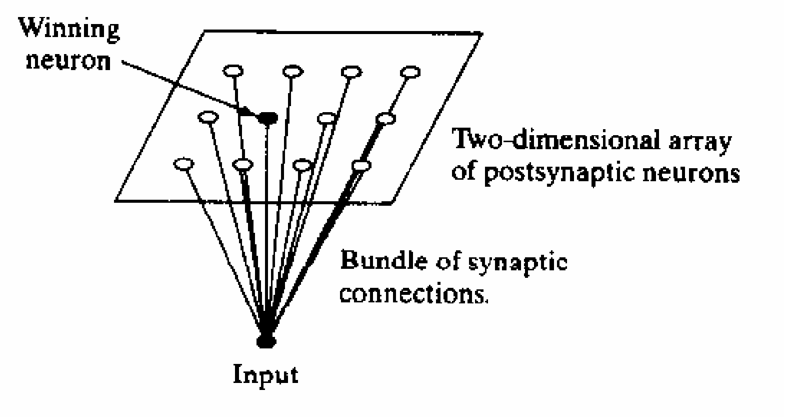
    input and output


Winning neuron
Two-dimensional array of postsynaptic neurons
Bundle of synaptic connections. (There is a similar bundle of synaptic connections originating from other presynaptic neurons.)
Activated neuron

(a) Willshaw-von der Malsburg's model

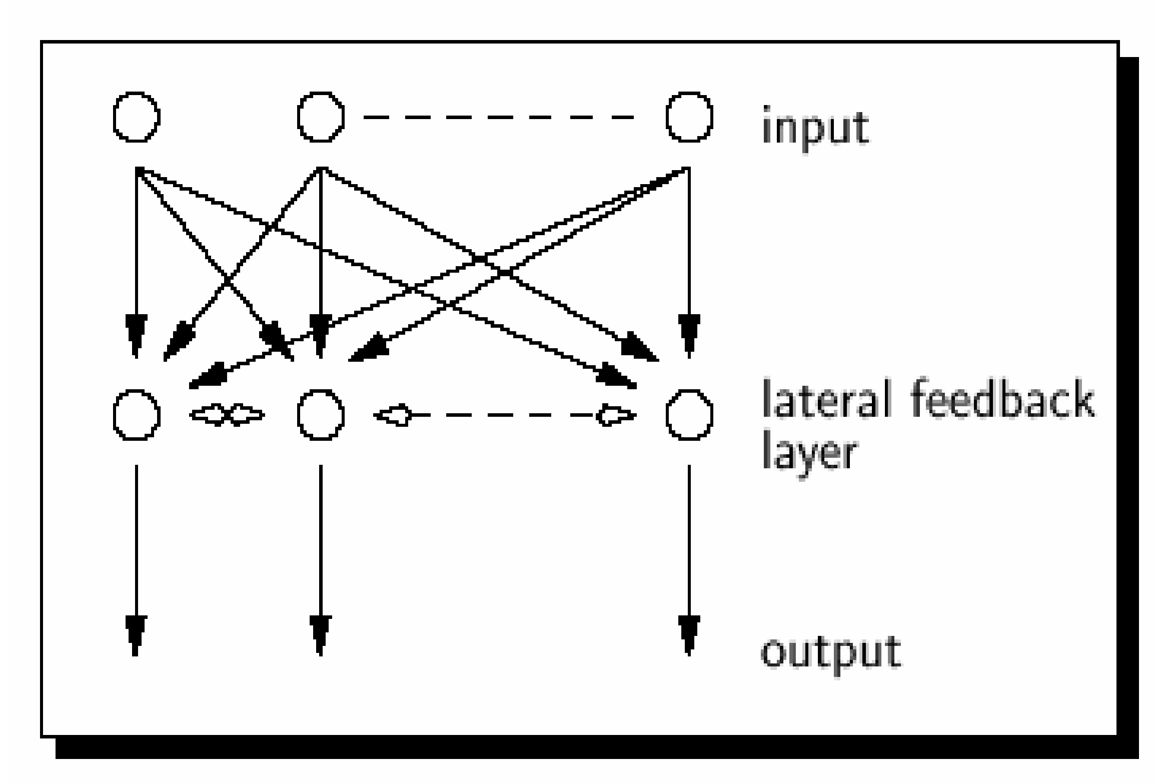
Basic Feature-mapping models

- Kohonen Model(1982)
  - Captures essential features of computational maps in Brain
    - remains computationally tractable
  - More general and more attention than Willshaw-Malsburg model
    - Capable of dimensionality reduction
    - Class of vector coding algorithm


Winning neuron
Two-dimensional array of postsynaptic neurons
Bundle of synaptic connections.
Input

(b) Kohonen model

Structure of Kohonen’s maps



Formation Process of SOM

- After initialization for synaptic weights, there are three essential processes
  - Competition
    - Largest value of discriminant function selected
    - Winner of competition
  - Cooperation
    - Spatial neighbors of winning neuron is selected
  - Synaptic adaptation
    - Excited neurons adjust synaptic weights

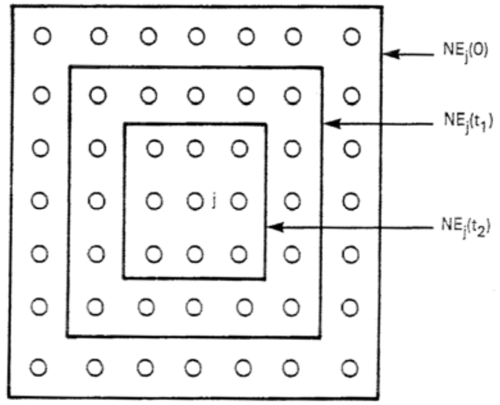
Competitive Process

- Input vector, synaptic weight vector
  - $\boldsymbol{x} = [x_1, x_2, \dots, x_m]^T$
  - $\boldsymbol{w}_j = [w_{j1}, w_{j2}, \dots, w_{jm}]^T, j = 1, 2, 3, l$
- Best matching, winning neuron
  - $i(\boldsymbol{x}) = \arg\min \|\boldsymbol{x} - \boldsymbol{w}_j\|, j = 1, 2, 3, .., l$
- Determine the location where the topological neighborhood of excited neurons is to be centered

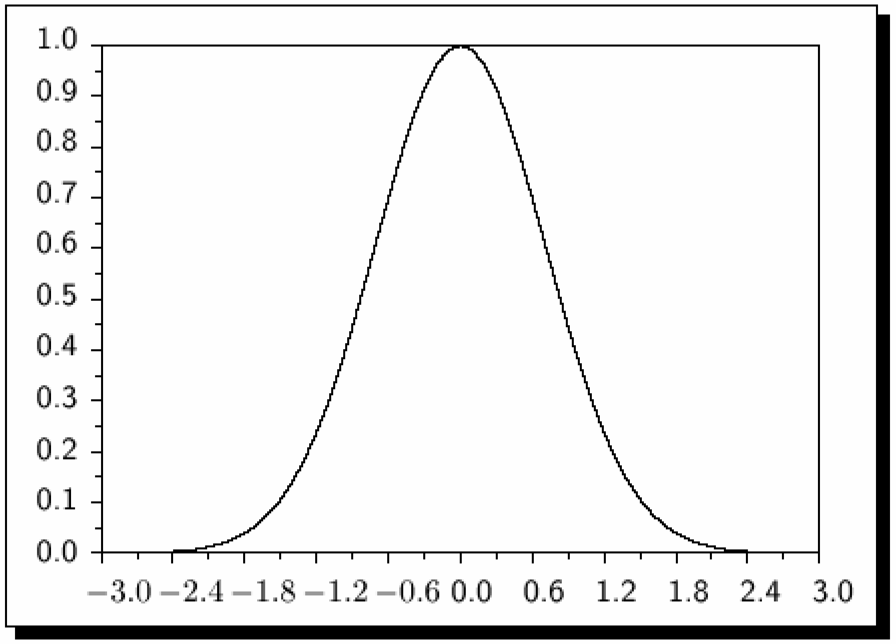
Cooperative Process

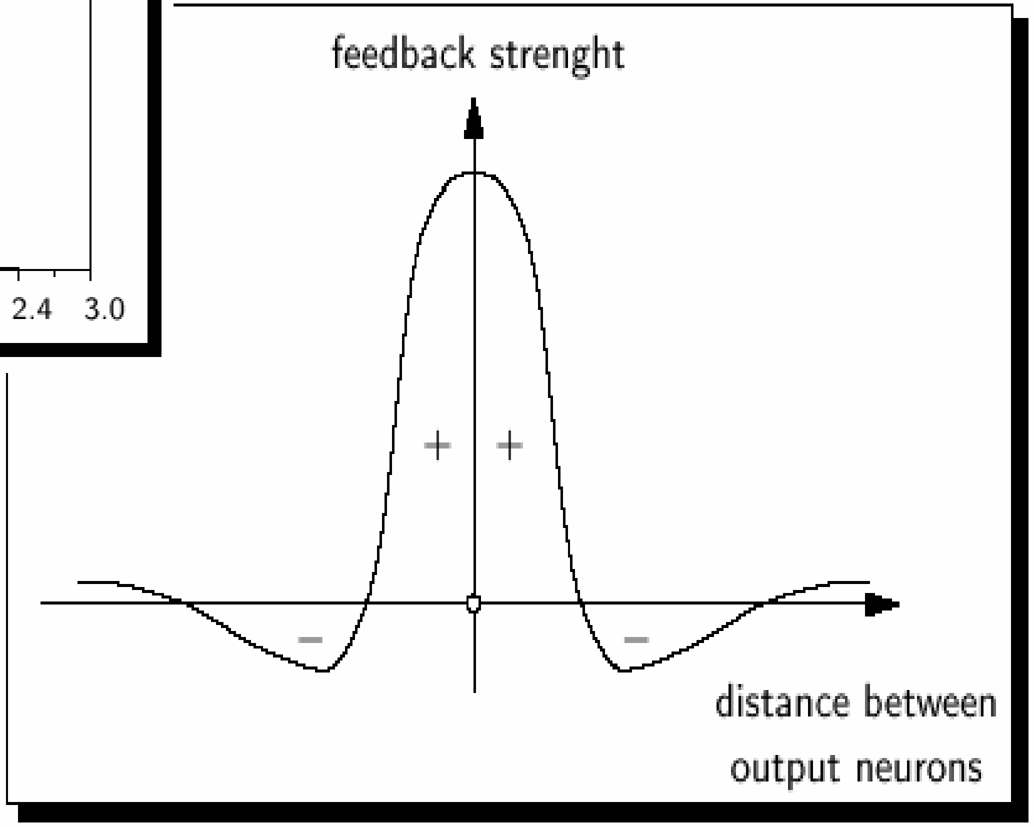
- For a winning neuron, the neurons in its immediate neighborhood excite more than those farther away
- topological neighborhood decay smoothly with lateral distance
  - Symmetric about maximum point defined by $d_{ij} = 0$
  - Monotonically decreasing to zero for $d_{ij} \rightarrow \infty$
  - Neighborhood function: Gaussian case $$h_{j,i(x)} = \exp\left(-\frac{d_{j,i}^2}{2\sigma^2}\right)$$
- Size of neighborhood shrinks with time

$$\sigma(n) = \sigma_0 \exp\left(-\frac{n}{\tau_1}\right), \quad n = 0,1,2,3$$



Examples of neighborhood function





Adaptive process

- Synaptic weight vector is changed in relation with input vector

  $\boldsymbol{w}_j(n+1)= \boldsymbol{w}_j(n) + \eta(n)\, h_{j,i(x)}(n)\, (\boldsymbol{x} - \boldsymbol{w}_j(n))$
- Applied to all neurons inside the neighborhood of winning neuron *i*
- Upon repeated presentation of the training data, weight tend to follow the distribution
- Learning rate $\eta(n)$: decay with time
- May decompose two phases
  - Self-organizing or ordering phase : topological ordering of the weight vectors
  - Convergence phase : after ordering, for accurate statistical quantification of the input space

Summary of SOM

- Continuous input space of activation patterns that are generated in accordance with a certain probability distribution
- Topology of the network in the form of a lattice of neurons, which defines a discrete output space
- Time-varying neighborhood function defined around winning neuron
- Learning rate decrease gradually with time, but never go to zero

Summary of SOM(2)

- Learning Algorithm
  - 1. Initialize *w*'s
  - 2. Present input vector
  - 3. Find nearest cell

$$i(\underline{x}) = argmin_j \, || \, \underline{x}(n) - \underline{w}_j(n) \, ||$$

  - 4. Update weights of neighbors

$$\underline{w}_j(n+1) = \underline{w}_j(n) + \eta \, (n) \, h_{j,i(x)}(n) \, [\, \underline{x}(n) - \underline{w}_j(n) \,]$$

  - 5. Reduce neighbors and η
  - 6. Go to 2

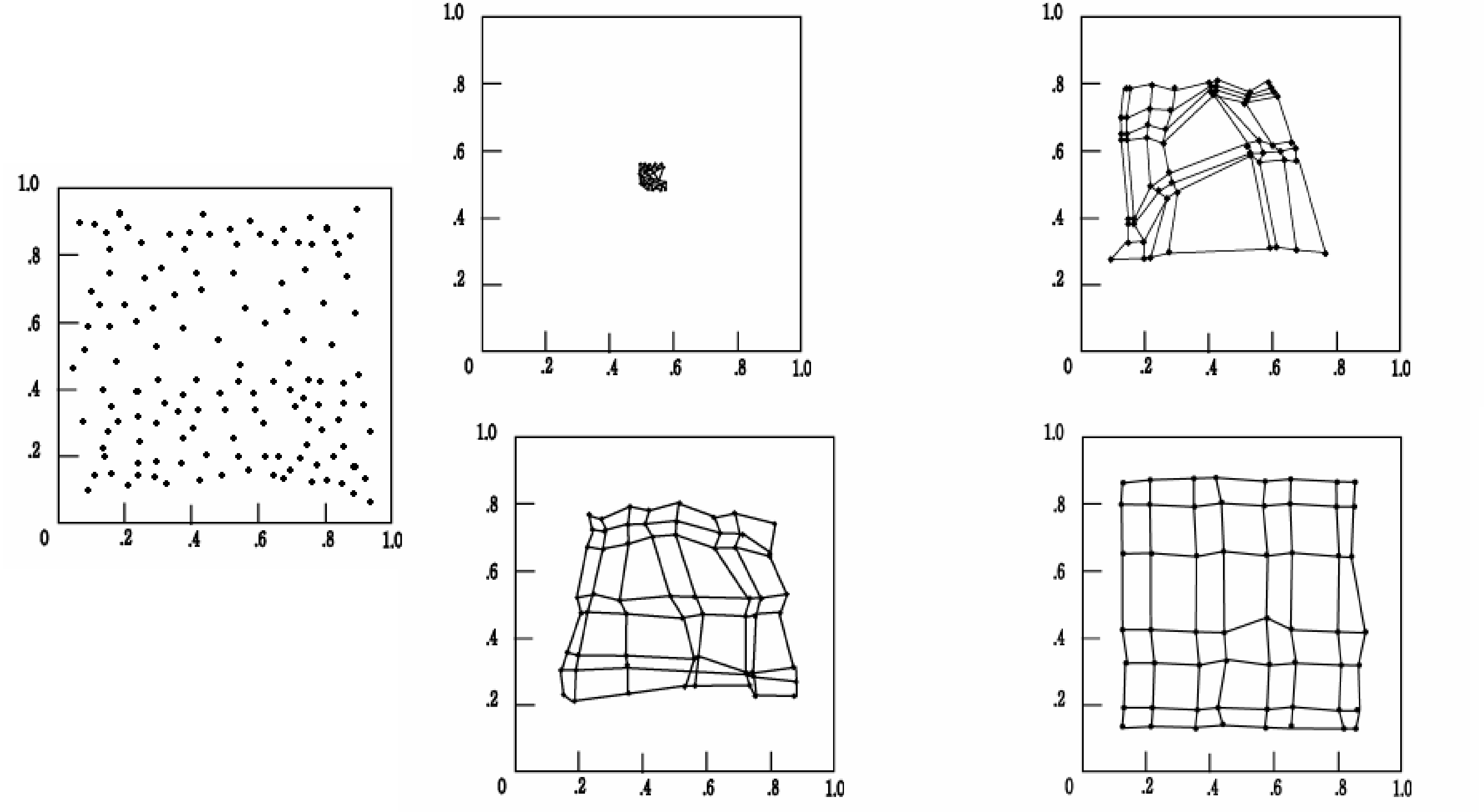
$h_{j,i(x)}(n)$ - the neighborhood function that has value 1 when i=k and falls off with the distance $| r_{i(x)} - r_i |$ between units j and i(x) in the output array.

Example of the neighborhood function is:

$$\aleph(i, k) = e^{\left(-\left|r_k - r_i\right|^2\right)/(2\sigma^2)}$$

where σ^2 is the width parameter that can gradually be decreased over time.

SOFM Example
2-D Lattice by 2-D distribution



Application of SOM in business

- Industrial process control
  - quality control and classification
  - process tracking and analysis
- Customer data analysis
  - segmentation of current customers
  - tailored products
  - targeted information distribution
  - better customer profitability
  - profiling of lost customers
- Identifying new customer groups
  - groups, that resemble your best customers
  - profitable groups, that are not currently being served
  - profitable groups, that have not been identified without eSom
- Business data analysis
  - grouping companies by financial key figures
  - identifying profitable and non-profitable companies
  - discovering possible growth opportunities
  - predicting bad credit and bankruptcies

Other applications

- Classification of documents for searching of its by content
- Compression of data
- Visualization of data in decision support systems
- Camera-robot coordination in robot-manipulator with vision

Difference between SOM and ART

- SOM
  - Fixed number of clusters
  - Orientation on classification and visualization

- ART
  - Variable number of clusters
  - Orientation on classification and categorization