Supervised vs.Unsupervised Learning

- In supervised learning we train algorithms with predefined concepts and functions based on labeled data
  $D = \{ (x, y) \mid x \in X, y \in \{yes,no\}.$
- In unsupervised learning we are given a set of instances *X* and we let the algorithms discover interesting properties of this set.
- Most unsupervised learning algorithms are based on the idea of discovering *similarities* between elements in the set *X.*

The k-Means Algorithm

The goal is to find k clusters of similar elements in the instance set X.

Suppose that we have n instances $x_0, x_1, ..., x_{n-1} \in X$; let $k < n$; let m_i be the mean of the instances in cluster i. Then we say that:

x is in cluster i if $min_i \| x - m_i \|$, where $0 \le i < k$ and $x \in X$.

Here $\| x - m_i \|$ is the Euclidean distance between x and m_i.

This suggests the following procedure for finding the k means:

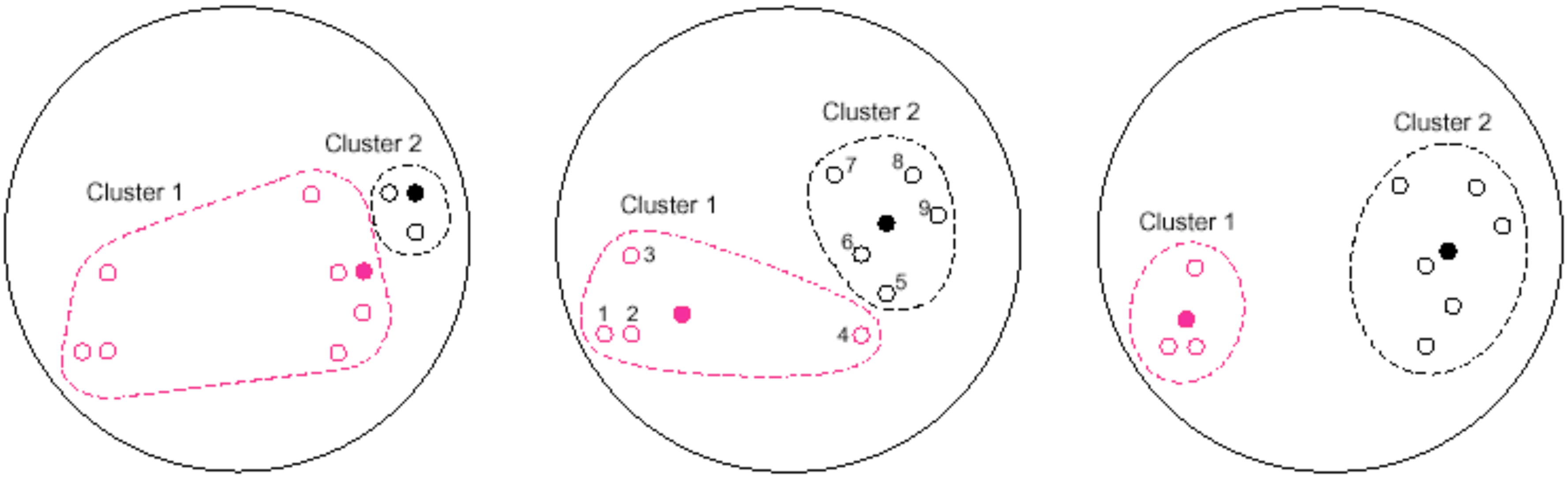
```
Make initial guesses for the means m_1, ..., m_k
Until there are no changes in any mean
    Use the estimated means to classify the examples into k clusters
    For i from 1 to k
      m_i ← mean of all of the examples for cluster i
```

The k-Means Algorithm

(a) Setup:
Reference point 1 (filled red circle) and reference point 2 (filled black circle) are chosen arbitrarily. All data points (open circles) are then partitioned into two clusters: each data point is assigned to cluster 1 or cluster 2, depending on whether the data point is closer to reference point 1 or 2, respectively.

(b) Results of first iteration:
Next each reference point is moved to the centroid of its cluster. Then each data point is considered in the sequence shown. If the reference point closest to the data point belongs to the other cluster, the data point is reassigned to that other cluster, and both cluster centroids are recomputed.

(c) Results of second iteration:
During the second iteration, the process in Figure 3(b) is performed again for every data point. The partition shown above is stable; it will not change for any further iteration.



Iterations of the k-means algorithm with *k = 2*.

The k-Means Algorithm

- Pros:
  - Fast convergence
  - Conceptually simple
- Cons:
  - Very sensitive to the choice of numbers of clusters *k*

⇒ Self-Organizing Maps

Self-Organizing Map: SOM

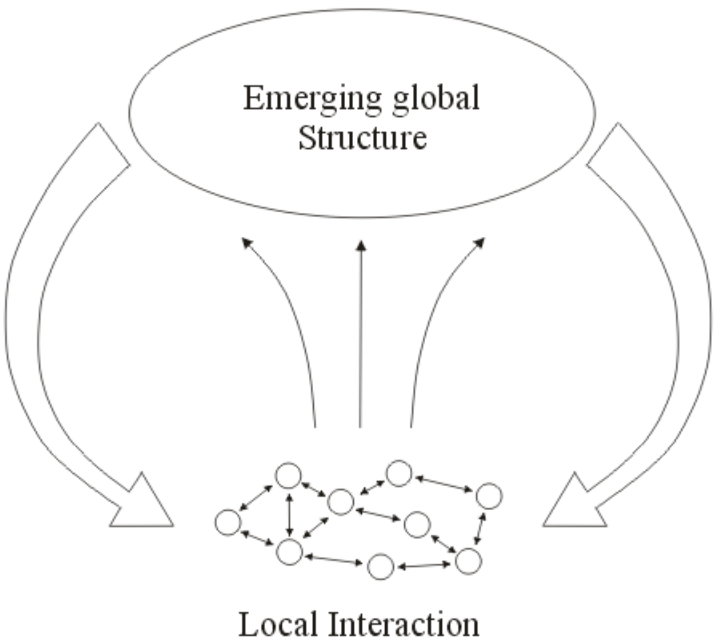
- A feed-forward neural network architecture based on competitive learning invented by Teuvo Kohonen in 1981.
- Does not depend on *a priori* selection of number of clusters to search for – will find the appropriate number of clusters for given the set of instances.
- Learning is slower than in k-means clustering - *competitive learning*

Self-Organization and Learning

- **Self-organization** refers to a process in which the internal organization of a system increases automatically without being guided or managed by an outside source.
- This process is due to local interaction with simple rules.
- Local interaction gives rise to global structure.

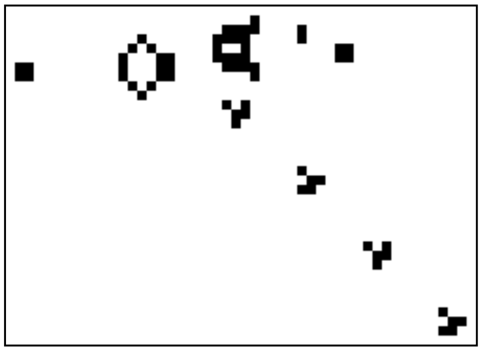
☞ We can interpret emerging global structures as learned structures.

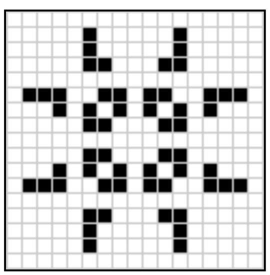
☞ Learned structures appear as clusters of similar objects.



Complexity : Life at the Edge of Chaos, Roger Lewin, University Of Chicago Press; 2nd edition, 2000

Game of Life

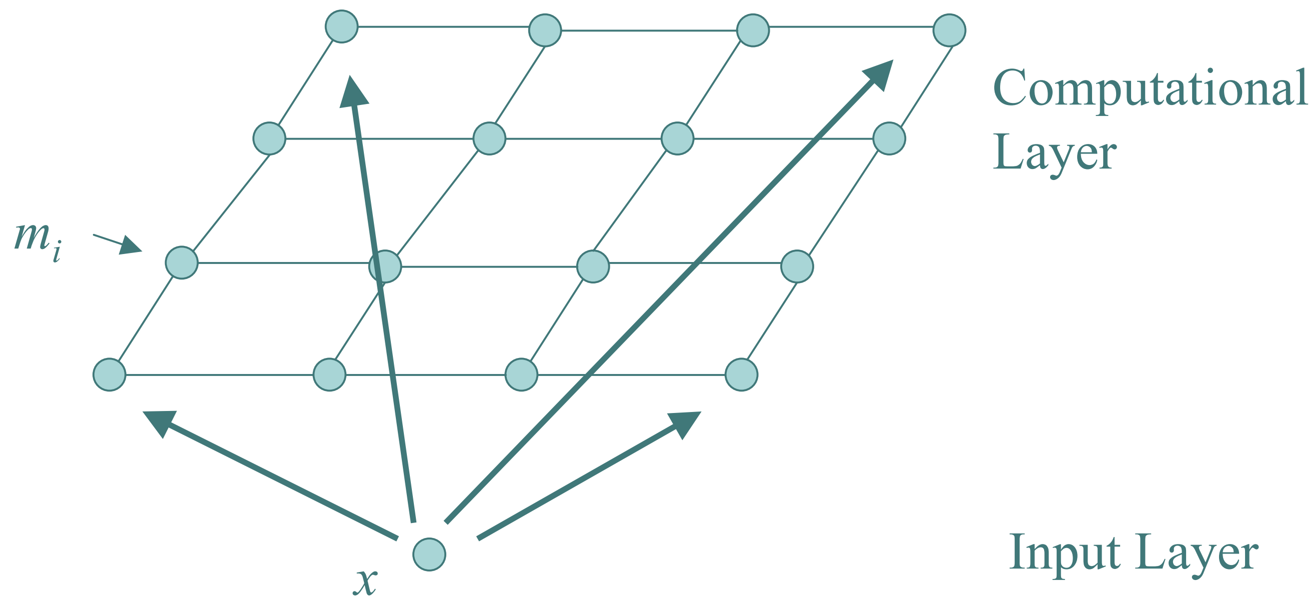




- Most famous example of self-organization - Game of Life
- Simple local rules:
  - Any live cell with fewer than two live neighbours dies, as if caused by under-population.
  - Any live cell with two or three live neighbours lives on to the next generation.
  - Any live cell with more than three live neighbors dies, as if by overcrowding.
  - Any dead cell with exactly three live neighbors becomes a live cell, as if by reproduction.

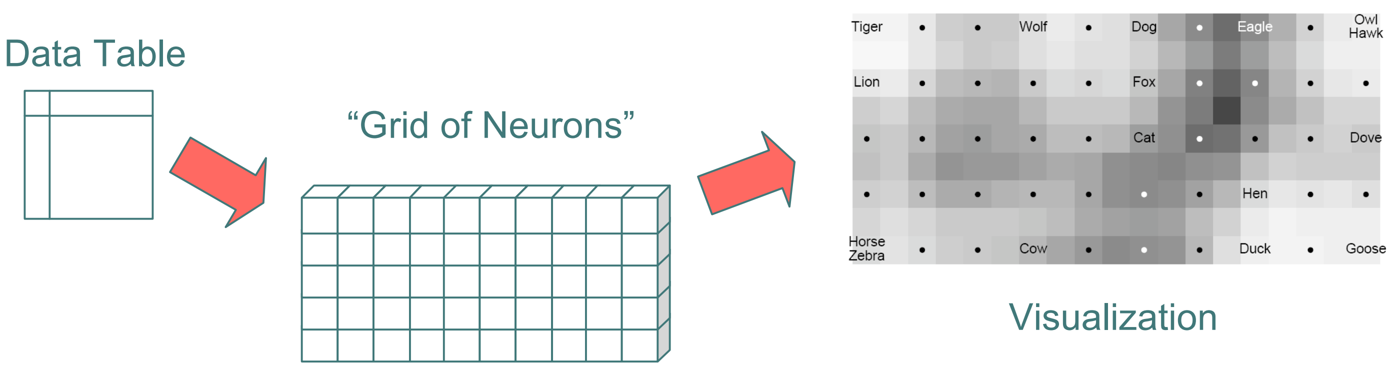
Source: http://en.wikipedia.org/wiki/Conway's_Game_of_Life

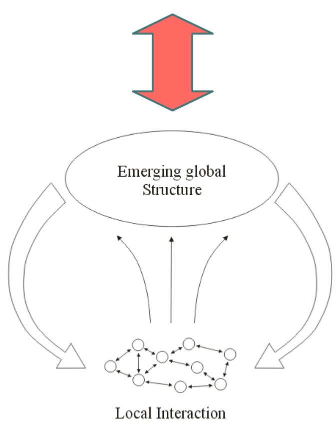
SOM Architecture



- This SOM has a feed-forward structure with a single computational layer arranged in rows and columns.
- Each neuron is fully connected to the input node in the input layer.
- The goal is to organize the neurons in the computational layer into clusters/regions associated with patterns in the instance set *X.*

Self-Organizing Maps





≡

Algorithm:

```
Repeat until Done
  For each row in Data Table Do
    Find the neuron that best describes the row.
    Make that neuron look more like the row.
    Smooth the immediate neighborhood of that neuron.
  End For
End Repeat
```

SOMs Sample the Data Space

- Given some distribution in the data space, SOM will try to construct a sample that looks like it was drawn from the same distribution.

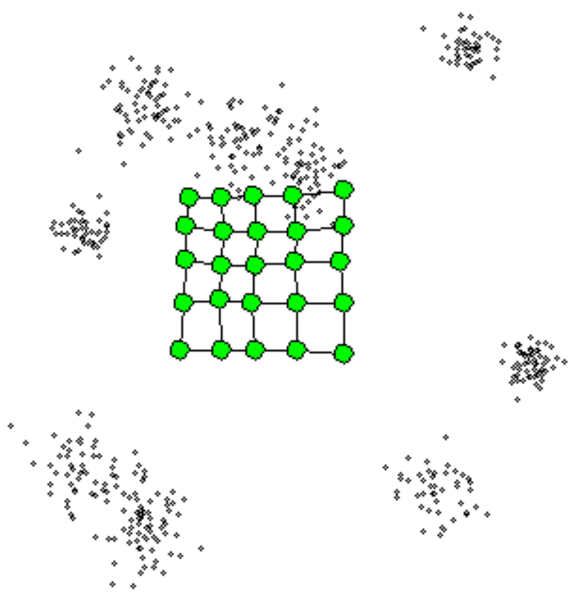
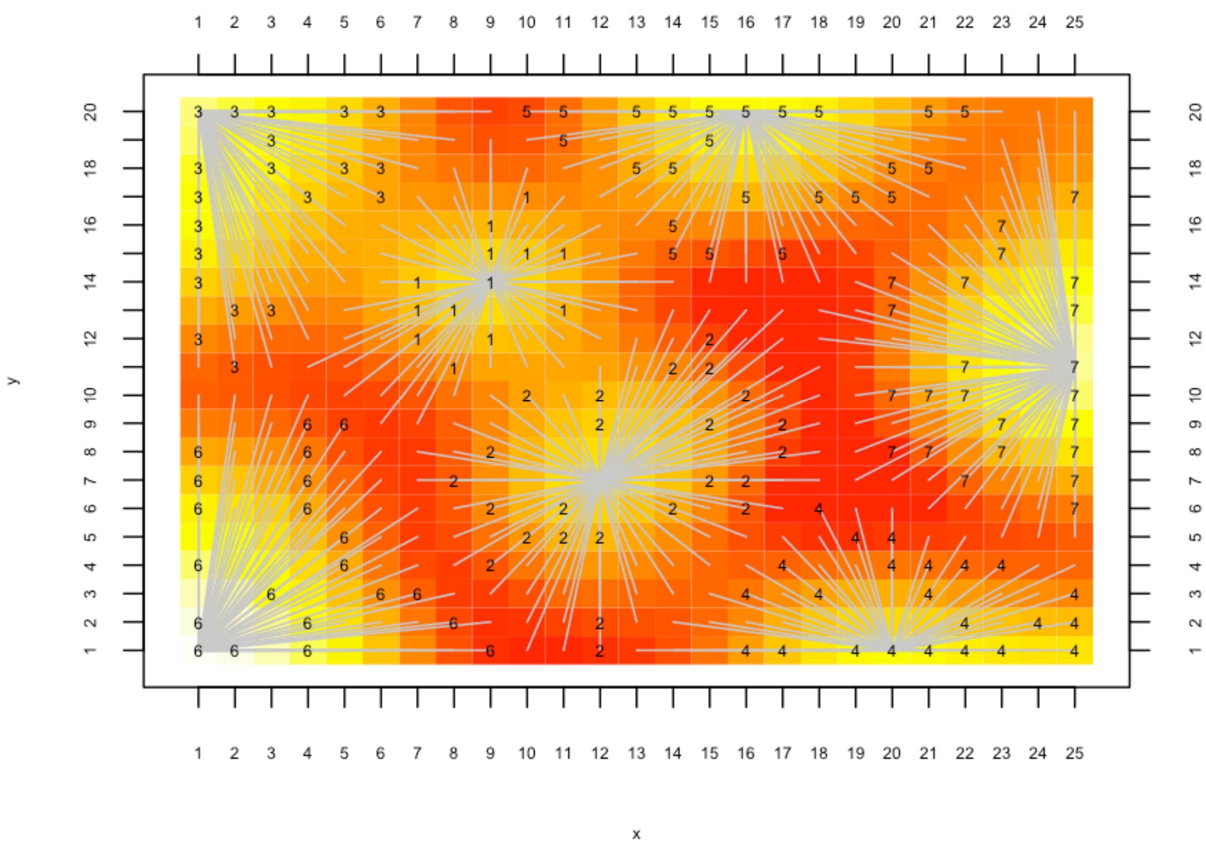


Image source: www.peltarion.com

Algorithm:

```
Repeat until Done
  For each row in Data Table Do
    Find the neuron that best describes the row.
    Make that neuron look more like the row.
    Smooth the immediate neighborhood of that neuron.
  End For
End Repeat
```

SOM Visualization



Visualization of Seven clusters using SOM

Training, Evaluation and Prediction

Unsupervised Learning

- Since in unsupervised learning we don't have labeled data, we cannot proceed as in the case of supervised learning.
- However, we can take a look at the quality of the clusters in terms of cluster density.
- Also, we can use trained clusters to assign unseen instances to clusters.

Training & Evaluation

- The variance of the distance of each point in a cluster from the centroid or mean is defined as follows:

$$E_k = \sum_{i=0}^{n_k - 1} \| x_{ki} - m_k \|^2$$

  where n_k is the number of points in cluster k, x_{ki} is the ith point in cluster k, and m_k is the centroid or mean of cluster k.
- Dense clusters have small variances, therefore, the magnitude of the variance is an indication of the quality of a cluster.
- This is also often called the *quantization error*
- SOM tries to reduce this quantization error as much as possible.