Self-organizing Maps (SOM)

Machine Learning (CS 306)

Instructor: Dr. Moumita Roy

Teaching Assistants: Indrajit Kalita, Veronica Naosekpam

Email-ids:

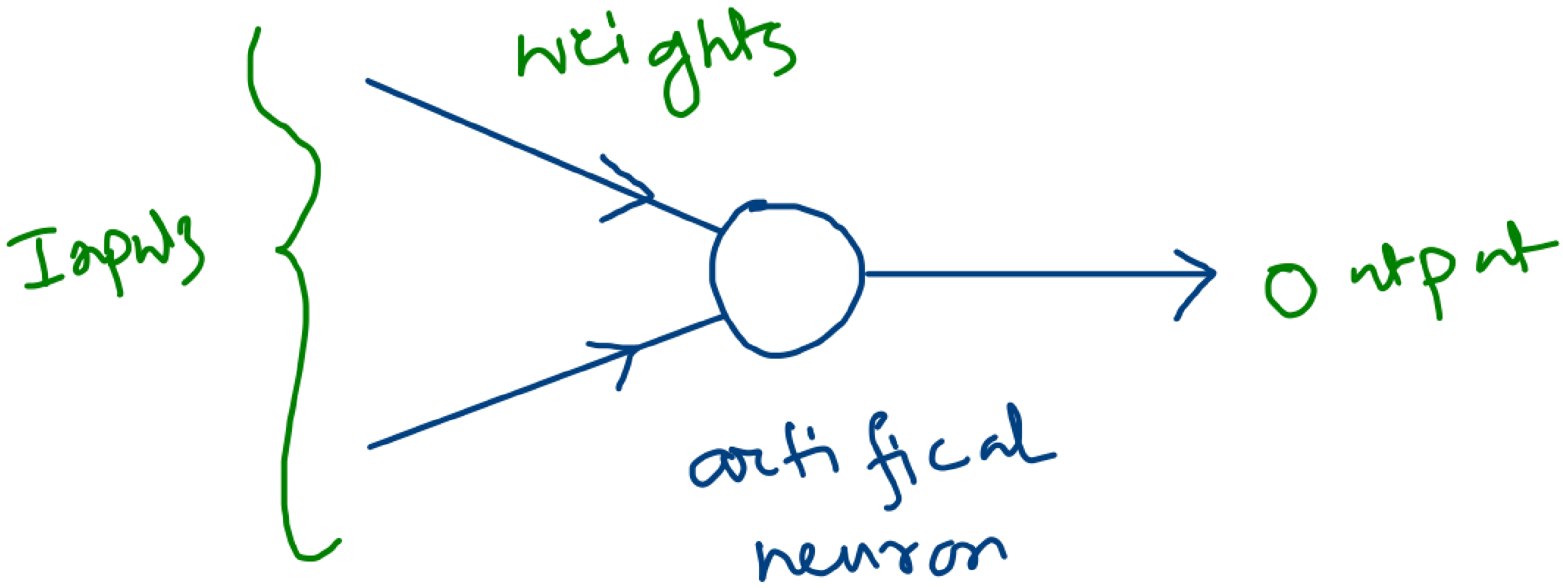
moumita@iiitg.ac.in

veronica.naosekpam@iiitg.ac.in

indrajit.kalita@iiitg.ac.in

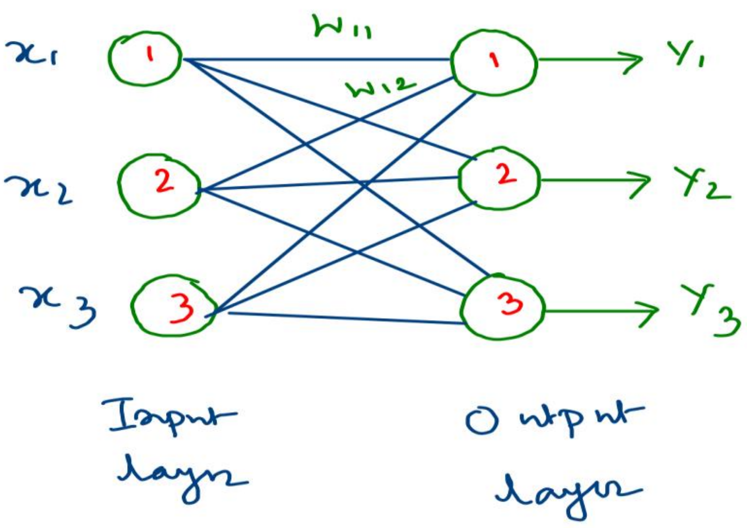
Mobile No: +91-8420489325 (only for emergency quires)

Artificial neurons for clustering



Observations

- In supervised mode, learning algorithm is conducted on error-correction method (presence of training samples).
- However, in unsupervised learning, learning process is conducted considering similarity between the patterns belongs to the same clusters.
- From these observation, we may use the following architecture for clustering.



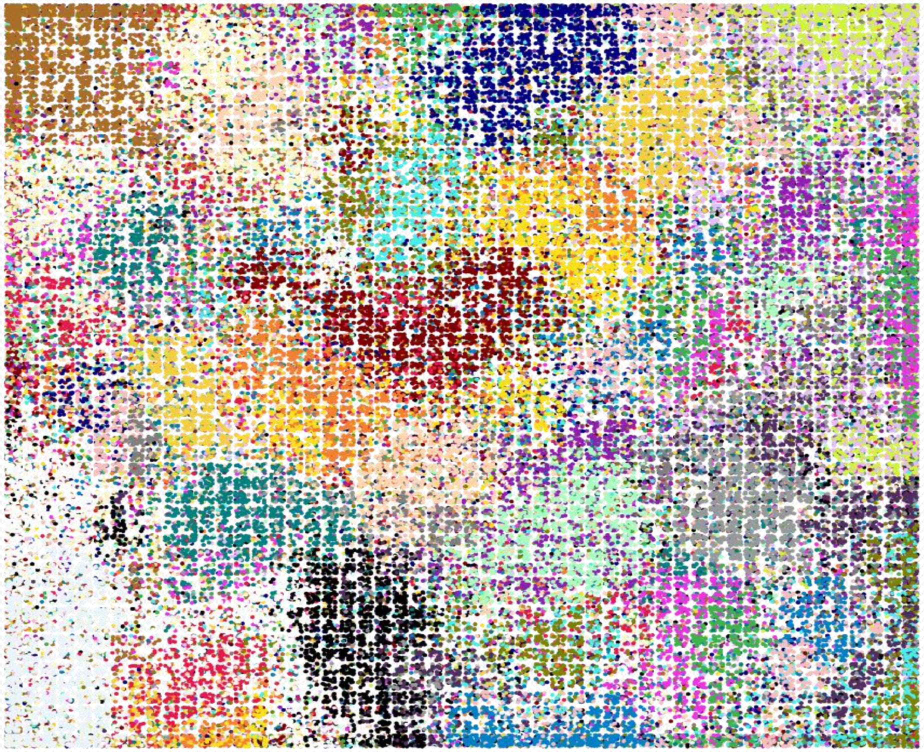
1. 3 features, 3 clusters
2. Each output neuron represents one cluster
3. Each pattern maps to any one output neuron
4. After training, 3 clusters will be formed

Competition

Winner-take-all

Self-Organizing Maps (SOM)

- Introduce by Prof. Teuvo Kohonen in 1982
- Clustering tools for high-dimensional complex data (unsupervised learning framework)



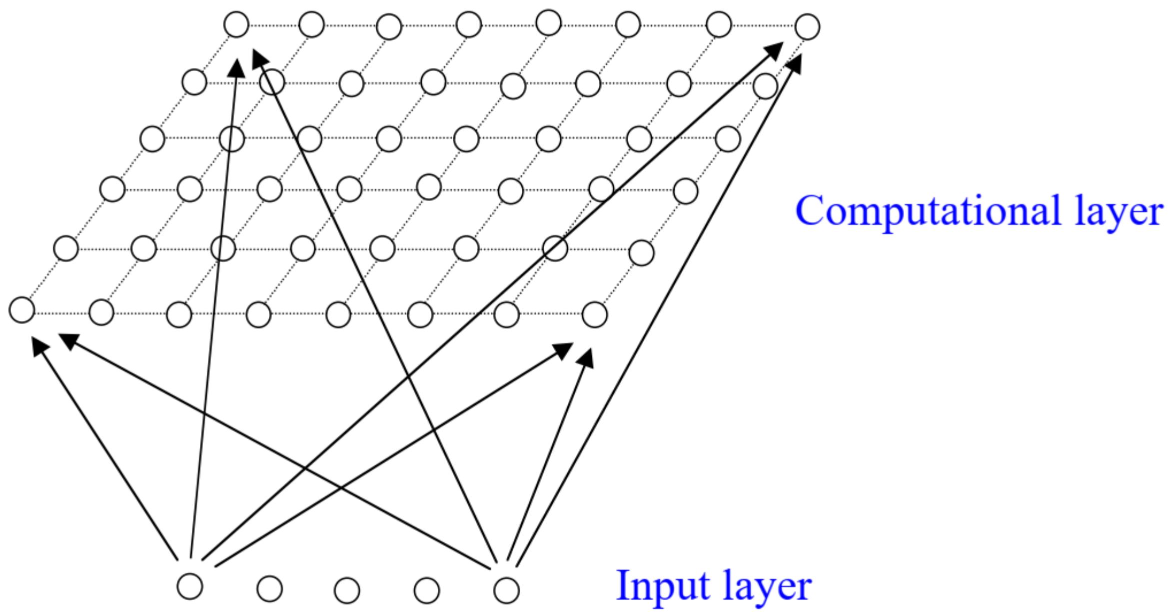
SOM

- This is partly motivated by how visual, auditory or other sensory information is handled in separate parts of the cerebral cortex in the human brain.

- Each sensory input is mapped into a corresponding area of cerebral cortex. The cortex is a self-organizing computational map in the human brain.

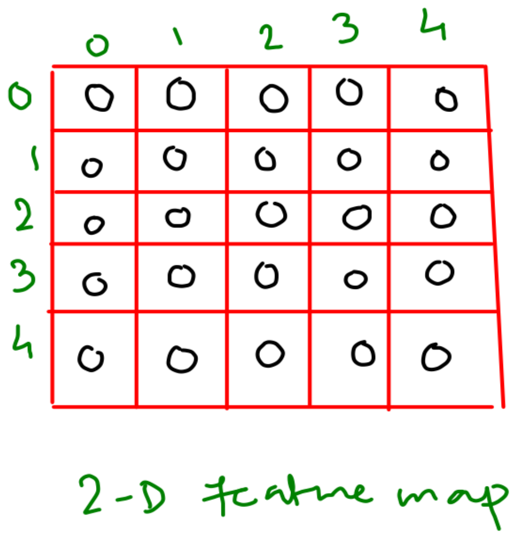
- SOM can also used for dimensionality reduction, as it maps high-dimension input patterns to a low (typically 2D/1D) dimensional space.

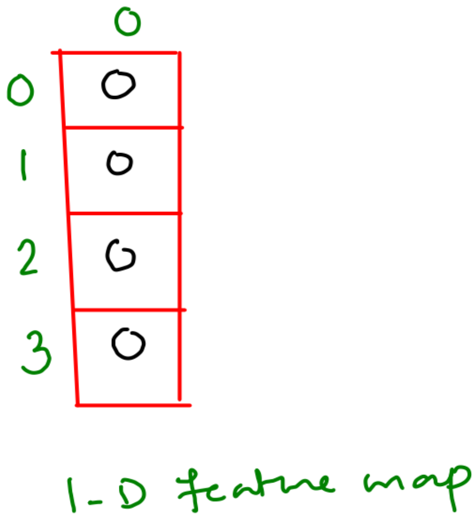
Architecture of SOM



Architecture of SOM

- It has two layers: input layer and output layer or computational layer or feature map
- Output layer: typically represented by 1-D or 2-D grid; each grid point represents a output node (has own coordinate).
- Input layer: number of node in the input layer is equal to the number of input features
- Feed-forward network; each neuron/node in input layer is connected to each neuron/node in the output layer





Topographic Maps

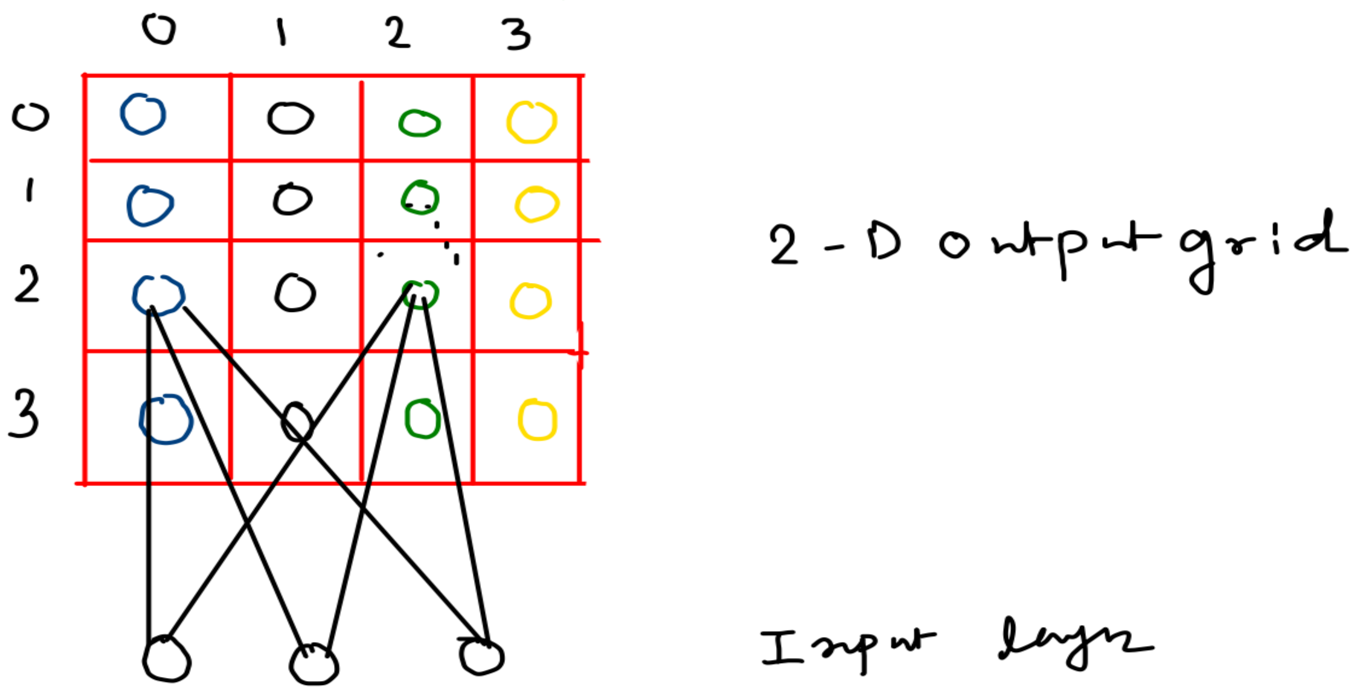
- SOM is maintaining a topology between input and output spaces
- Input patterns that are close/ similar in high dimensional space are also mapped to nearby nodes in the 2D/1D output space (conserves the underlying structure of its input space).
- SOM can also used for dimensionality reduction, as it maps high-dimension input patterns to a low (typically 2D/1D) dimensional space (in grid representation).
- The learning process is relying on self-organizing behaviors of the neurons

Cluster 1 : Images of apple

Cluster 2 : Images of Orange

Cluster 3 : Images of potato

Cluster 4 : Images of Carrot



Question

- Suppose, you need to form 3 clusters considering the students of your class and after clustering representatives should stand in a topographical order.

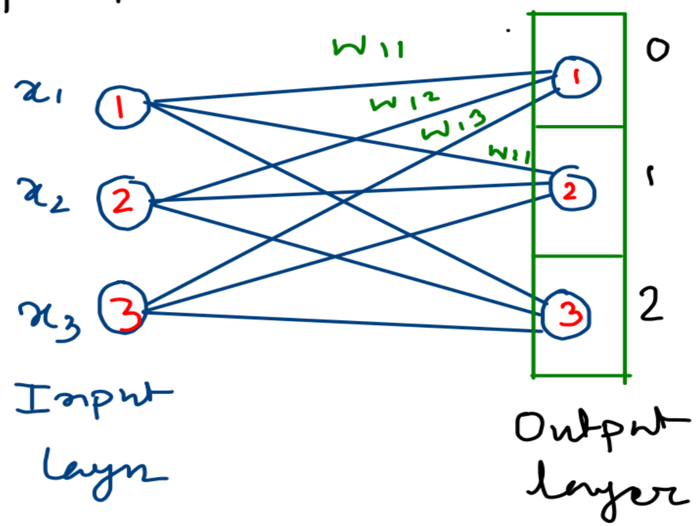
Learning steps of SOM (Competitive learning)

- Competition
- Cooperation
- Weight updating

Task: Design a SOM for 3 features and 3 cluster problem

Architecture of SOM

Input pattern: $X = [x_1 \; x_2 \; x_3]$



Placed three representative output neuron in 1-D grid structure

Three weight representative corresponding to three output neurons

$W_1 = [w_{11} \; w_{12} \; w_{13}]$

$W_2 = [w_{21} \; w_{22} \; w_{23}]$

$W_3 = [w_{31} \; w_{32} \; w_{33}]$

Competition Step

- The output neuron whose weight vector comes closest to the input pattern (most similar to it) is declared the winner or best matching unit (BMU).

- In this way, the continuous input space can be mapped to the discrete output space of neurons by a simple process of competition between the neurons.

Competition

$X = [x_1 \ x_2 \ \dots \ x_n] \rightarrow$ n-dimensional input pattern

$w_j = [w_{j1} \ w_{j2} \ \dots \ w_{jn}]$

$j \rightarrow \{1, 2, \dots N\}$

$N \rightarrow$ total number of output neurons

$d_j = \| x - w_j \|$

$I(x) = \underset{j}{\operatorname{argmin}} \ d_j$

winner output neuron for X input pattern

Competition

Calculate distance measure between input pattern and weight representative of each output neuron.

$D_1 = ||X - W_1||$

$D_2 = ||X - W_2||$

$D_3 = ||X - W_3||$

Choose the output neuron/node corresponding to minimum distance

Suppose, D_2 is minimum / output neuron 2 → winner

input pattern X is assigned to cluster 2.

Cooperation Process

- In neurobiological studies, there is lateral interaction within a set of excited neurons. When one neuron fires, its closest neighbours tend to get excited more than those further away. There is a topological neighbourhood that decays with distance.

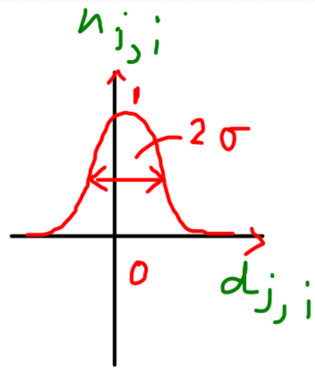
- In SOM, the similar concept of topological neighbourhood function is used to compute for amount of cooperation between winner neuron and its neighbouring neurons.

- This process is helpful to map similar patterns to the nearby neurons in the output space (generation of topographic map).

- Different types of neighbourhood functions are available: Gaussian neighbourhood function, Rectangular neighbourhood function etc.

Cooperation process

Gaussian neighbourhood function



$$\underbrace{h_{j,I(x)}} = \exp\left(-d^{2}_{j,I(x)} / 2\sigma^{2}\right)$$

maximum at winning neuron / decreases with increasing lateral distance

$d_{j,i} \rightarrow$ lateral distance between output neuron i and j

$I(x) \rightarrow$ winner neuron

$j \rightarrow \{1, 2, \ldots N\}$ $N \rightarrow$ number of output neurons

$\sigma \rightarrow$ effective width of topological neighbourhood

Suppose, $I(x) = 2$

$$d_{1,2} = \|[0\ \ 0] - [1\ \ 0]\| = 1$$

$$d_{2,2} = 0$$

Let, $\sigma = 1$

$$d_{3,2} = 1$$

$$h_{1,2}(x) = \exp\left(-\frac{d_{1,2}^{2}}{2 \times 1^2}\right) = \exp\left(-\frac{1^2}{2}\right) = ?$$

$$h_{2,2}(x) = \exp\left(-\frac{0^2}{2 \times 1^2}\right) = 1$$

$$h_{3,2}(x) = ??$$

Updating of weights

- SOM has some learning process by which the outputs become self-organised and the feature map between inputs and outputs is formed.
- The training process of SOM is based on competitive learning (not error-correction based learning).
- The weight representations of winner neuron as well as its neighbours are updated considering the topological neighbourhood function.
- Concept: Weight representative of a cluster (output node) comes closer to the input patterns which are assigned to that cluster

Weight updation

Let,

$x = [5] \rightarrow$ one feature input pattern

$w = [2] \rightarrow$ any weight representative

if x is assigned to corresponding cluster of w.

then updation should bring w closer to x

$$w^{itr+1} = w^{itr} + (x - w)$$

$$= 2 + (5 - 2)$$

$$= 5$$

Rule

$$w_j^{itr+1} = w_j^{itr} + \alpha * h_{j,I(x)}(x) * (x - w_j^{itr})$$

$\alpha \rightarrow$ learning rate

updation is applicable for all output neurons
for each winning neuron

$$W_{11}^{itr+1} = W_{11}^{itr} + \alpha * \underbrace{h_{1,2}}_{\text{winner}}(x) * (x_1 - W_{11})$$

update $W_1, W_2, W_3 \rightarrow$??

In SOM, α, σ are typically decreases with time (epoch)

Learning process of SOM

Step 1: Randomly initialized all weight representatives with small values

Step 2 (competition): For each input pattern, select the winner neuron and assign it to a particular cluster

Step 3 (cooperation process): The amount of cooperation between the winner neuron and its neighbors is calculated using topological neighborhood function

Step 4 (updating): Update the weight representatives of winner neuron and its neighbor

Step 5: Repeat the Steps 2-4 for all input patterns (epoch) until convergence (no change in cluster assignment)

Question

Perform the clustering using SOM considering these patterns:
P1:[2 4], P2:[10 2], P3:[4 2]

Demonstrate the learning process using the followings:

- Input pattern: P2
- Number of clusters:3
- Learning rate: 1
- Gaussian neighborhood function with width =1
- Initialization of weight vectors:[1 1], [2 2], [3 3]