**Self Organizing Map (SOM)**

SoxeshShreshta

*Abstract*— *The self-organizing map (SOM) is an automatic data-analysis method. It is widely applied to clustering problems and data exploration in industry, finance, natural sciences, and linguistics. It is one of the most popular artificial neural network (ANN) models that is trained using unsupervised learning to produce a low dimensional (especially two-dimensional), discretized representation of the input space of the training samples which is called a map.*

Keywords— clustering problems, data exploration, unsupervised learning, training samples.

### 

### 1. Introduction

Self-Organizing map is introduced by the Finnish professor Teuvo Kohonen in 1980s is also known as kohonen map or network. With a small number of nodes, SOM behaves in a way that is similar to K-means, larger self-organizing maps rearrange data in a way that is fundamentally topological in character.

SOMs cannot simulate the complex parallel workings of the human nervous system, but they aim to reproduce an observed behavioral principle and by doing so they in fact do turn out to be very powerful tools. The special principle that SOMs aim to imitate is the mapping of high-dimensional input data to a low-dimensional output network in a topology preserving way, i.e. similar inputs are mapped to similar outputs. This feature makes SOMs especially useful for visualization tasks of complex structures that would otherwise be hardly recognizable by humans.

2.Unsupervised Training

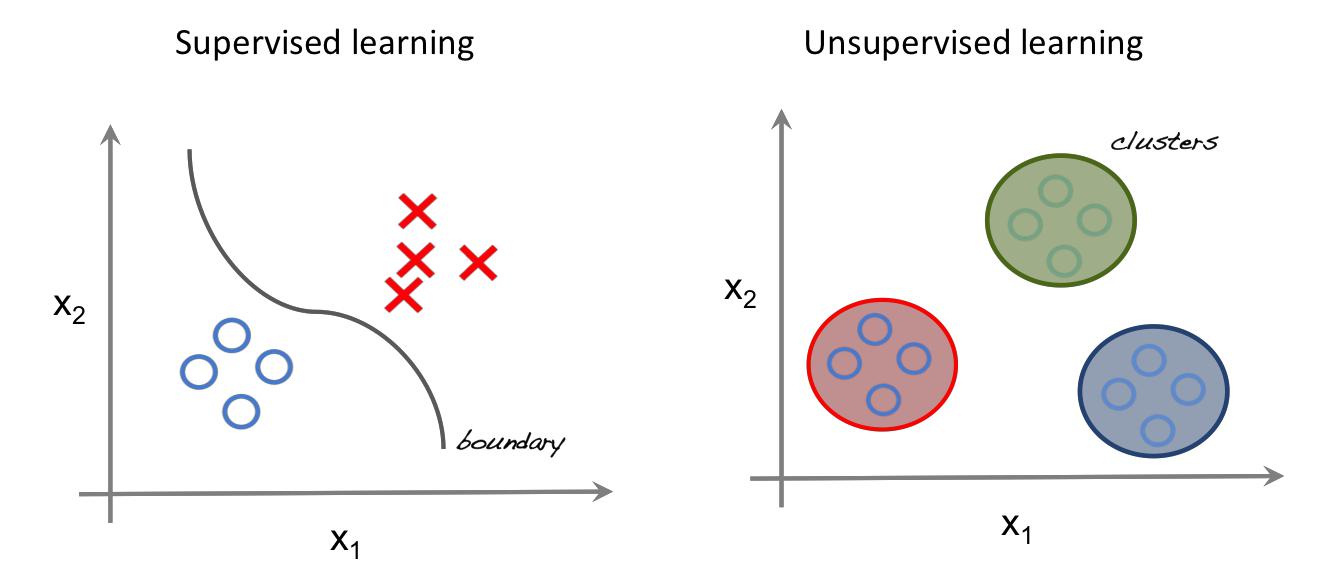
Unsupervised learning is the training of an artificial intelligence (AI) algorithm using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Therefore there is no accuracy calculation of the structure that is output by the relevant algorithm.

* In unsupervised learning, an AI system is presented with unlabeled, uncategorized data and the system’s algorithms act on the data without prior training thus, the output of the system is completely dependent upon the algorithms coded. Unsupervised training is a method to test the AI system.
* Unsupervised learning algorithms can perform more complex processing tasks than supervised learning systems. However, unsupervised learning can be more unpredictable than the alternate model. While an unsupervised learning AI system might, for example, figure out on its own how to sort cats from dogs, it might also add unforeseen and undesired categories to deal with unusual breeds, creating clutter instead of order.

Supervised vs Unsupervised Training

If the training is given for machine learning task for every input with corresponding target, it is called supervised learning, which will be able to provide target for any new input after sufficient training. The learning algorithm seeks a function from inputs to the respective targets. If the targets are expressed in some classes, it is called classification problem. Alternatively, if the target space is continuous, it is called regression problem. Generally, associated targets for every input are used in supervised learning by reducing error to get a potential target.

If the training is given for machine learning task only with a set of inputs, it is called unsupervised learning, which will be able to find the structure or relationships between different inputs. Most important unsupervised learning is clustering, which will create different cluster of inputs and will be able to put any new input in appropriate cluster. Other than clustering, other unsupervised learning techniques are: anomaly detection, Hebbian Learning and learning Latent variable models such as Expectation–maximization algorithm, Method of moments (mean, covariance) and Blind signal separation techniques (Principal component analysis, Independent component analysis, Non-negative matrix factorization, Singular value decomposition). Since the examples given to the unsupervised task have no associated target, there is no credit or blame to be used in learning.



*Fig 1: Supervised learning and unsupervised learning*

3.Working Principle of Unsupervised Training

Unsupervised learning works by analyzing the data without its labels for the hidden structures within it, and through determining the correlations, and for features that actually correlate two data items. It is being used for clustering, dimensionality reduction, feature learning, density estimation, etc.

The hidden structure sometimes called feature vector, represents the input data such a way that if the same feature vector is being use to reconstruct the input, then one can do that with some acceptable loss. The variance in two feature vectors of two inputs directly proportional to the variance in the inputs itself. Thus, this hidden structure or feature vector only represents features in the data that actually give distinction to it.

During the process of unsupervised learning, the system does not have concrete data sets, and the outcomes to most of the problems are largely unknown. In simple terminology, the AI system and the ML objective is blinded when it goes into the operation. The system has its faultless and immense logical operations to guide it along the way, but the lack of proper input and output algorithms makes the process even more challenging. Incredible as the whole process may sound, unsupervised learning has the ability to interpret and find solutions to a limitless amount of data, through the input data and the binary logic mechanism present in all computer systems. The system has no reference data at all.

## 

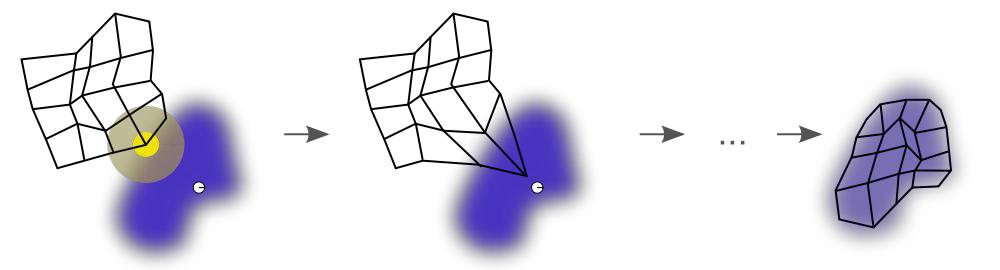
*Fig 2: Working principle Unsupervised training*

## 4.Working of Self Organizing Map

The Self-Organizing Map is one of the most popular neural network models. It belongs to the category of competitive learning networks. The Self-Organizing Map is based on unsupervised learning, which means that no human intervention is needed during the learning and those little needs to be known about the characteristics of the input data. We could, for example, use the SOM for clustering data without knowing the class memberships of the input data. The SOM can be used to detect features inherent to the problem and thus has also been called SOFM, the Self-Organizing Feature Map.

The visible part of a self-organizing map is the map space, it consists of components called nodes or neurons. The map space is defined beforehand, usually as a finite two-dimensional region where nodes are arranged in a regular hexagonal or rectangular grid. [8] Each node is associated with a "weight" vector, which is a position in the input space; that is, it has the same dimension as each input vector. While nodes in the map space stay fixed, training consists in moving weight vectors toward the input data, without spoiling the topology induced from the map space. Thus, the self-organizing map describes a mapping from a higher-dimensional input space to a lower-dimensional map space. Once trained, the map can classify a vector from the data space by finding the node with the closest (smallest distance metric) weight vector to the data space vector.

SOMs are used for finding other representation of the data which is easy for further analysis by humans. This is a great method of visualizing highly dimensional data, analyzing "what is going on", how are some classes grouped geometrically, etc. The key feature to SOMs is that the topological features of the original input data are preserved on the map i.e. similar input samples (where similarity is defined in terms of the input variables (age, sex, height, weight)) are placed close together on the SOM grid. For example, all 55-year-old females that are approximately 1.6m in height will be mapped to nodes in the same area of the grid. Taller and smaller people will be mapped elsewhere, taking all variables into account. Tall heavy males will be closer on the map to tall heavy females than small light males as they are more “similar”.



*Fig 3: Illustration of Self Organizing Map*

5.Overview of the Project

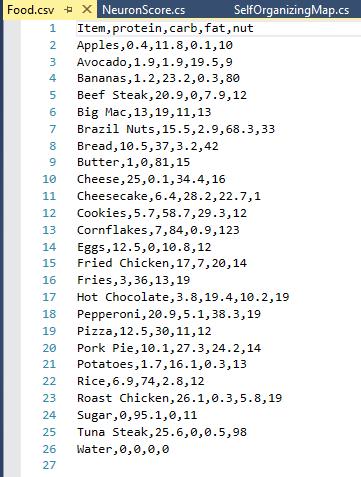
The Self Organizing Map project is based on the unsupervised training where no initial labeled data is provided to the algorithm for training. The reason for choosing the algorithm are:

* Finding Clusters
* Present an input pattern to the network.
* Choose the output neuron with the highest activation (the “winner”)
* Update the weights of neurons that are within the “neighborhood” of the winner, using a relative learning factor.
* Reduce the learning factor monotonically
* Reduce the size of the “neighborhood” monotonically
* Repeat from step two until only small updates are observed.

The spread of the neighborhood function will initially include all neurons on the grid, gradually reducing to include only the winning neuron.

To provide an example of what to expect from a SOM, I have prepared a simple example that will attempt to group twenty-five foods into regions of similarity, based on three parameters, which are protein, carbohydrate and fat.

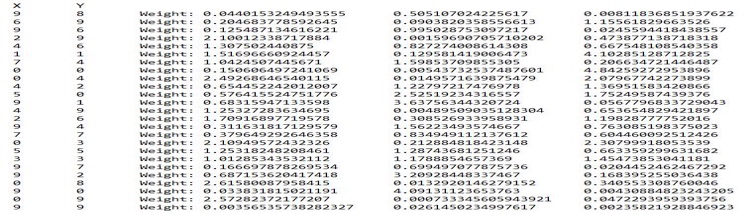
Therefore, the challenge for this SOM is to reduce data containing three dimensions down to two, whilst retaining meaning. It does this by automatically identifying differentiating features that will have the greatest effect.



## 

*Fig 4: Data Sheet for SOM*

In my algorithm, the method initialize () creates a random neuron and assign weight. This is done to select the winner among the neurons. Then it will calculate weights and of its features with the random neurons created. Thus, after the iteration of the algorithm, the neuron that lays closet to the input point will be the winner (euclidean distance). Then the winner neuron is moved closer to the input point compared to other neuron. The adjacent neuron also moves along with the winner neuron and is updated in every iteration which is updated by using the neuron factor in the algorithm. The following output is presented after the completion of the algorithm



*Fig 5: Output of the algorithm*

6.Conclusion

The Self Organizing Map project is based on the unsupervised training where no initial labeled data is provided to the algorithm for training so the weights for the input data was calculated along with the mapping coordinate. Then the inspection is done, and the generated data were arranged accordingly with respect to its similar properties.

### 7.References

[1] http://www.robosoup.com/2008/11/c-self-organising-map-som.html

[2] https://www.shanelynn.ie/self-organising-maps-for-

customer-segmentation-using-r/

[3] https://www.youtube.com/watch?v=H9H6s-x-0YE

[4]https://www.datasciencecentral.com/profiles/blogs/machine-learning-explained-[understanding-supervised-unsupervised](https://www.datasciencecentral.com/profiles/blogs/machine-learning-explained-understanding-supervised-unsupervised)