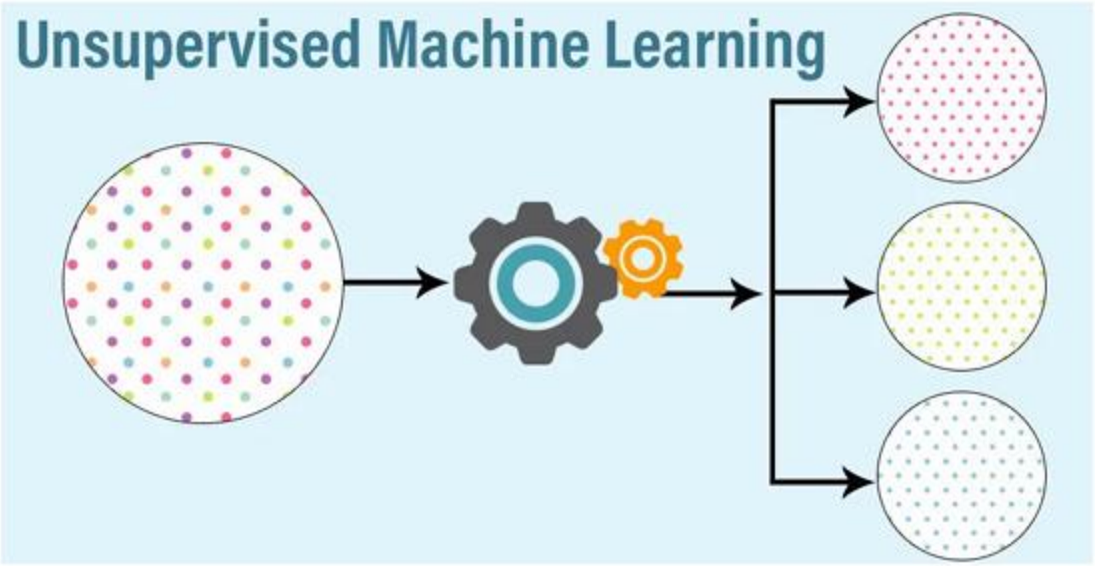
Unsupervised Learning



Type of Artificial Neural Network Inspired by Unsupervised learning

- **Self-Organizing Feature Maps**
- **Adaptive Resonance Theory**

Self-Organizing Feature Maps

Introduction

- Neural Networks use processing, inspired by the human brain, as a basis to develop algorithms that can be used to model and understand complex patterns and prediction problems.
- There are several types of neural networks and each has its own unique use.
- The Self Organizing Map (SOM) is one such variant of the neural network, also known as Kohonen's Map.

Self-Organizing Maps

- A self-organizing map is also known as SOM and it was proposed by Kohonen.
- It is an unsupervised neural network that is trained using unsupervised learning techniques to produce a low dimensional, discretized representation from the input space of the training samples, known as a map and is, therefore, a method to reduce data dimensions.

Self-Organizing Maps…

- Self-Organizing Maps are very different from other artificial neural networks as they apply competitive learning techniques unlike others using error-correction learning methods such as backpropagation with gradient descent, and use a neighborhood function to preserve all the topological properties within the input space.

Self-Organizing Maps…

- Self-Organizing Maps were initially only being used for data visualization, but these days, it has been applied to different problems, including as a solution to the Traveling Salesman Problem as well.
- Map units or neurons usually form a two-dimensional space and hence a mapping from high dimensional space onto a plane is created.

Self-Organizing Maps…

- The map retains the calculated relative distance between the points. Points closer to each other within the input space are mapped to the nearby map units in Self-Organizing Maps.
- SOMs can thus serve as a cluster analyzing tool for high dimensional data.
- SOMs also have the capability to generalize.
- During generalization, the network can recognize or characterize inputs that it has never seen as data before.
- New input is taken up with the map unit and is therefore mapped.

Uses of Self-Organizing Maps…

- SOM provide an advantage in maintaining the structural information from the training data and are not inherently linear.
- Using Principal Component Analysis on high dimensional data may just cause loss of data when the dimension gets reduced into two.
- If the data comprises a lot of dimensions and if every dimension preset is useful, in such cases SOM can be very useful over PCA for dimensionality reduction.

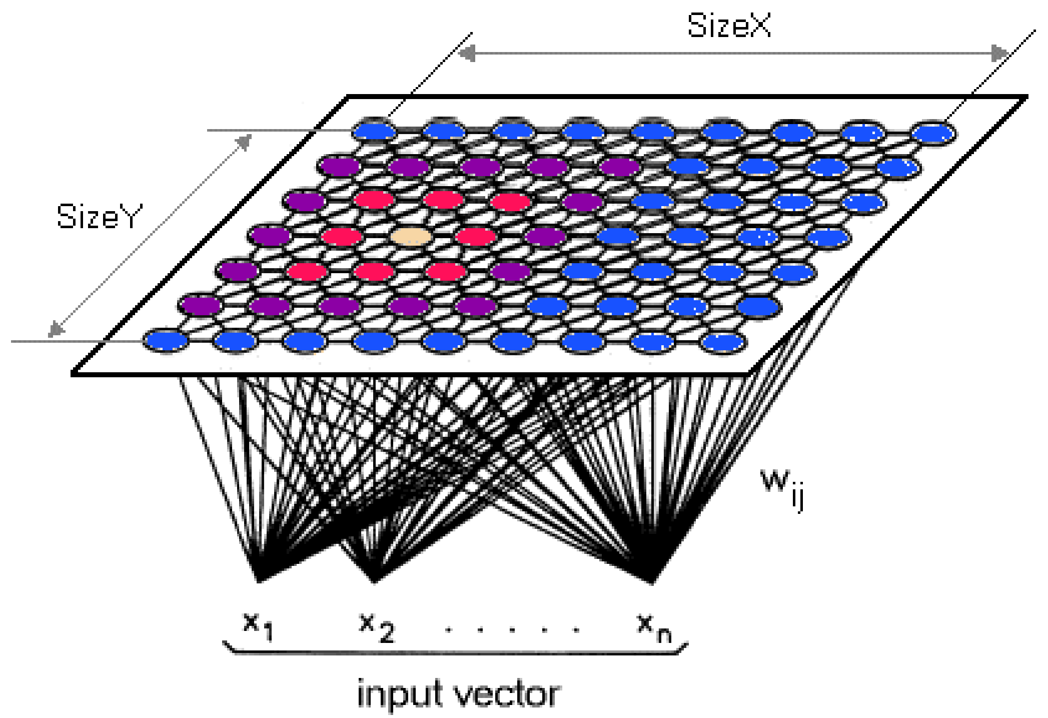
Uses of Self-Organizing Maps...

- Text clustering is another important preprocessing step that can be performed through SOM.
- It is a method that helps to verify how the present text can be converted into a mathematical expression for further analysis and processing.
- Data Exploration and visualization are also the most important applications of SOM.

Introduction to SOM

- SOM is an unsupervised machine learning technique that produces a low-dimensional representation of a higher-dimensional dataset while preserving its topological structure.
- It employs a competitive learning algorithm to train its network, allowing it to organize data into clusters or groups.
- Developed based on the retina-cortex mapping, SOM achieves simplicity in structure and computational form.

Uses of Self-Organizing Maps…



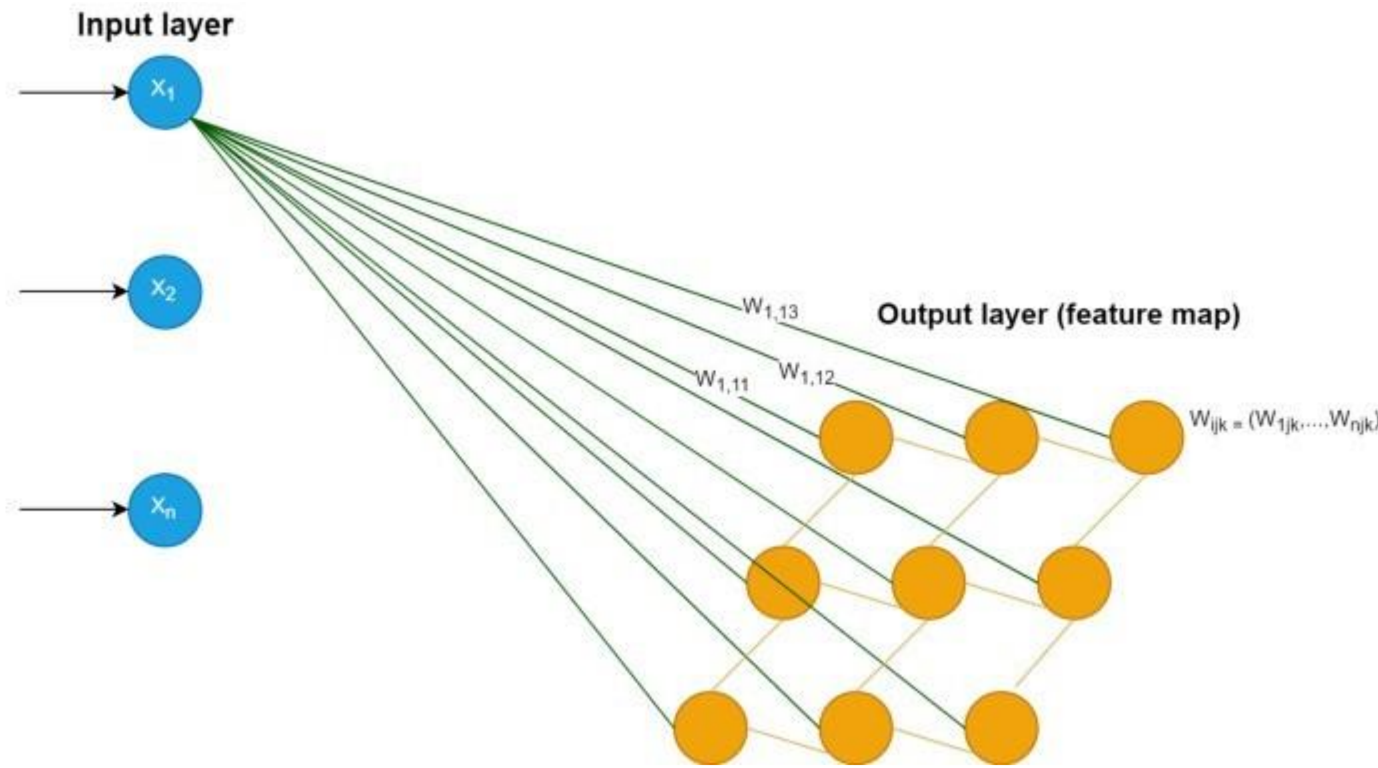
SOM Training

- SOM doesn’t use backpropagation with SGD to update weights, this type of unsupervised artificial neural network uses competitive learning to update its weights.
- Competitive learning is based on three processes :
  - Competition
  - Cooperation
  - Adaptation

Competition

- As we said before each neuron in a SOM is assigned a weight vector with the same dimensionality as the input space.
- In the example below, in each neuron of the output layer we will have a vector with dimension n.
- We compute distance between each neuron (neuron from the output layer) and the input data, and the neuron with the lowest distance will be the winner of the competition.
- The Euclidean metric is commonly used to measure distance.

Competition

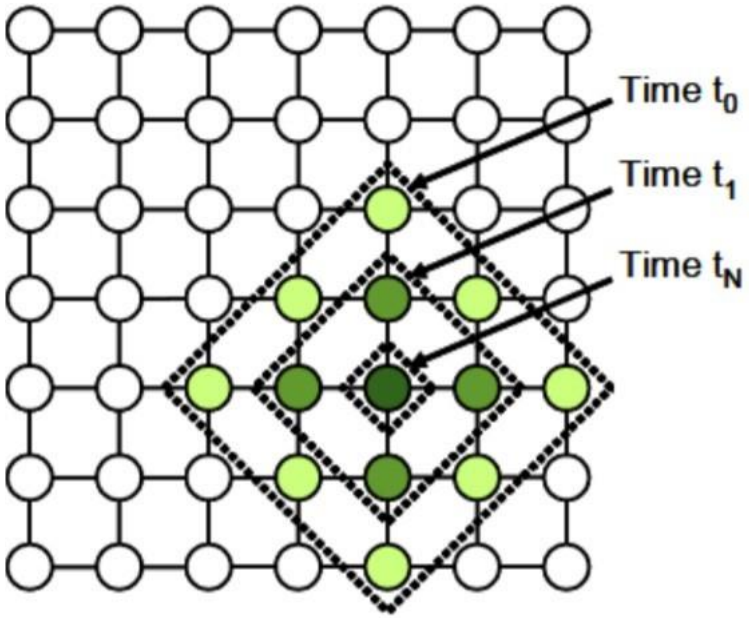


Cooperation

- We will update the vector of the winner neuron in the final process (adaptation) but it is not the only one, also it’s neighbor will be updated.
- How do we choose the neighbors ?
- To choose neighbors we use neighborhood kernel function, this function depends on two factor : time (time incremented each new input data) and distance between the winner neuron and the other neuron (How far is the neuron from the winner neuron).

Cooperation…

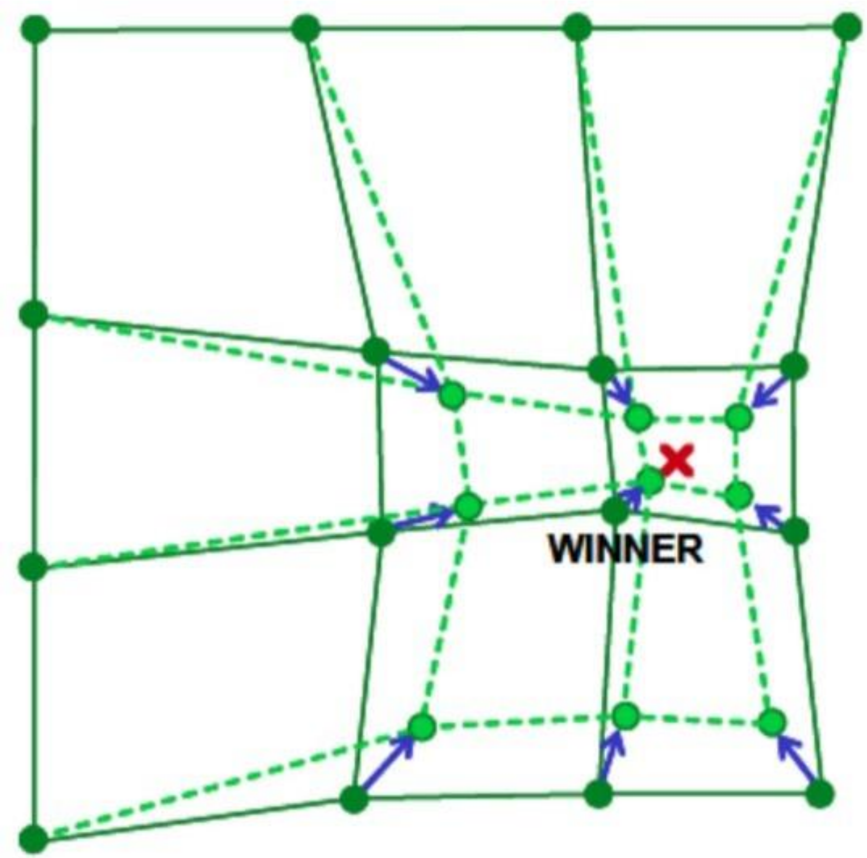
- The image below show us how the winner neuron's (The most green one in the center) neighbors are chosen depending on distance and time factors.



Adaptation

- After choosing the winner neuron and it’s neighbors we compute neurons update.
- Those chosen neurons will be updated but not the same update, more the distance between neuron and the input data grow less we adjust it like shown in the image below :

Adaptation



Adaptation…

- The neighborhood kernel depends on the distance between winner neuron and the other neuron (they are proportionally reversed : d increase make h(t) decrease) and the neighborhood size which itself depends on time (decrease while time incrementing) and this make neighborhood kernel function decrease also.

Here's a high-level summary of the SOM algorithm:

1. **Initialization**:
    - Start by initializing the weight vectors of the SOM's nodes (neurons) randomly or based on some prior knowledge.
2. **Competition:**
    - For each input vector, determine the Best Matching Unit (BMU) by finding the node with the closest weight vector to the input vector.
3. **Cooperation:**
    - Define a neighbourhood around the BMU, and adjust the weights of the nodes within this neighbourhood to make them more like the input vector.
4. **Adaptation:**
    - Update the weight vectors of the neurons within the neighbourhood of the BMU to move them closer to the input vector.
5. **Continuation:**
    - Repeat steps 2-4 for each input vector and for a number of iterations, gradually decreasing the neighbourhood size and the learning rate over time.
6. **Convergence:**
    - The algorithm continues until the map starts to stabilize and the changes to the neurons' weights are minimal.

Architecture

- Self-Organizing Maps consist of two important layers, the first one is the input layer, and the second one is the output layer, which is also known as a feature map.
- Each data point in the dataset recognizes itself by competing for a representation.
- The SOMs' mapping steps start from initializing the weight to vectors.

Architecture

- After this, a random vector as the sample is selected and the mapped vectors are searched to find which weight best represents the chosen sample.
- Each weighted vector has neighboring weights present that are close to it. The chosen weight is then rewarded by being able to become a random sample vector.
- This helps the map to grow and form different shapes. Most generally, they form square or hexagonal shapes in a 2D feature space.
- This whole process is repeatedly performed a large number of times and more than 1000 times.

Architecture

- To simply explain, learning occurs in the following ways:
  - Every node is examined to calculate which suitable weights are similar to the input vector. The suitable node is commonly known as the Best Matching Unit.
  - The neighborhood value of the Best Matching Unit is then calculated. The number of neighbors tends to decrease over time.

Architecture

- The suitable weight is further rewarded with transitioning into more like the sample vector. The neighbours transition like the sample vector chosen.
  - The closer a node is to the Best Matching Unit, the more its weights get altered and the farther away the neighbour is from the node, the less it learns.
- Repeat the second step for N iterations.

Pros

- Data can be easily interpreted and understood with the help of techniques like reduction of dimensionality and grid clustering.
- Self-Organizing Maps are capable of handling several types of classification problems while providing a useful, and intelligent summary from the data at the same time.

Cons

- It does not create a generative model for the data and therefore the model does not understand how data is being created.
- Self-Organizing Maps do not perform well while working with categorical data and even worse for mixed types of data.
- The model preparation time is comparatively very slow and hard to train against the slowly evolving data.

Adaptive Resonance Theory

Adaptive Resonance Theory

- The Adaptive Resonance Theory (ART) was incorporated as a hypothesis for human cognitive data handling.
- The hypothesis has prompted neural models for pattern recognition and unsupervised learning.
- ART system has been utilized to clarify different types of cognitive and brain data.

Adaptive Resonance Theory…

- The Adaptive Resonance Theory addresses the **stability-plasticity**(stability can be defined as the nature of memorizing the learning and plasticity refers to the fact that they are flexible to gain new information) dilemma of a system that asks how learning can proceed in response to huge input patterns and simultaneously not to lose the stability for irrelevant patterns.

Adaptive Resonance Theory…

- Other than that, the stability-elasticity dilemma is concerned about how a system can adapt new data while keeping what was learned before.
- For such a task, a feedback mechanism is included among the ART neural network layers.
- In this neural network, the data in the form of processing elements output reflects back and ahead among layers.
- If an appropriate pattern is build-up, the resonance is reached, then adaption can occur during this period.

Adaptive Resonance Theory…

- It can be defined as the formal analysis of how to overcome the learning instability accomplished by a competitive learning model, let to the presentation of an expended hypothesis, called adaptive resonance theory (ART).
- This formal investigation indicated that a specific type of **top-down learned feedback** and matching mechanism could significantly overcome the instability issue.

Adaptive Resonance Theory…

- It was understood that top-down attentional mechanisms, which had prior been found through an investigation of connections among cognitive and reinforcement mechanisms, had similar characteristics as these code-stabilizing mechanisms.
- In other words, once it was perceived how to solve the instability issue formally, it also turned out to be certain that one did not need to develop any quantitatively new mechanism to do so.

Adaptive Resonance Theory…

- One only needed to make sure to incorporate previously discovered attentional mechanisms.
- These additional mechanisms empower code learning to self- stabilize in response to an essentially arbitrary input system.
- Grossberg presented the basic principles of the adaptive resonance theory. A category of ART called ART1 has been described as an arrangement of ordinary differential equations by carpenter and Grossberg.
- These theorems can predict both the order of search as the function of the learning history of the system and the input patterns.

ART Learning

- ART learning is based on the concept of "resonance", where the network enters a stable state when the input matches the learned category.
- The learning process involves adjusting the weights and category representations based on the degree of match between the input and the existing categories.

ART Variants

- ART1 is an unsupervised learning model primarily designed for recognizing binary patterns.
- ART2: Unsupervised learning for analog input patterns
- Fuzzy ART: Handles fuzzy input patterns
- ARTMAP: Supervised learning for classification tasks

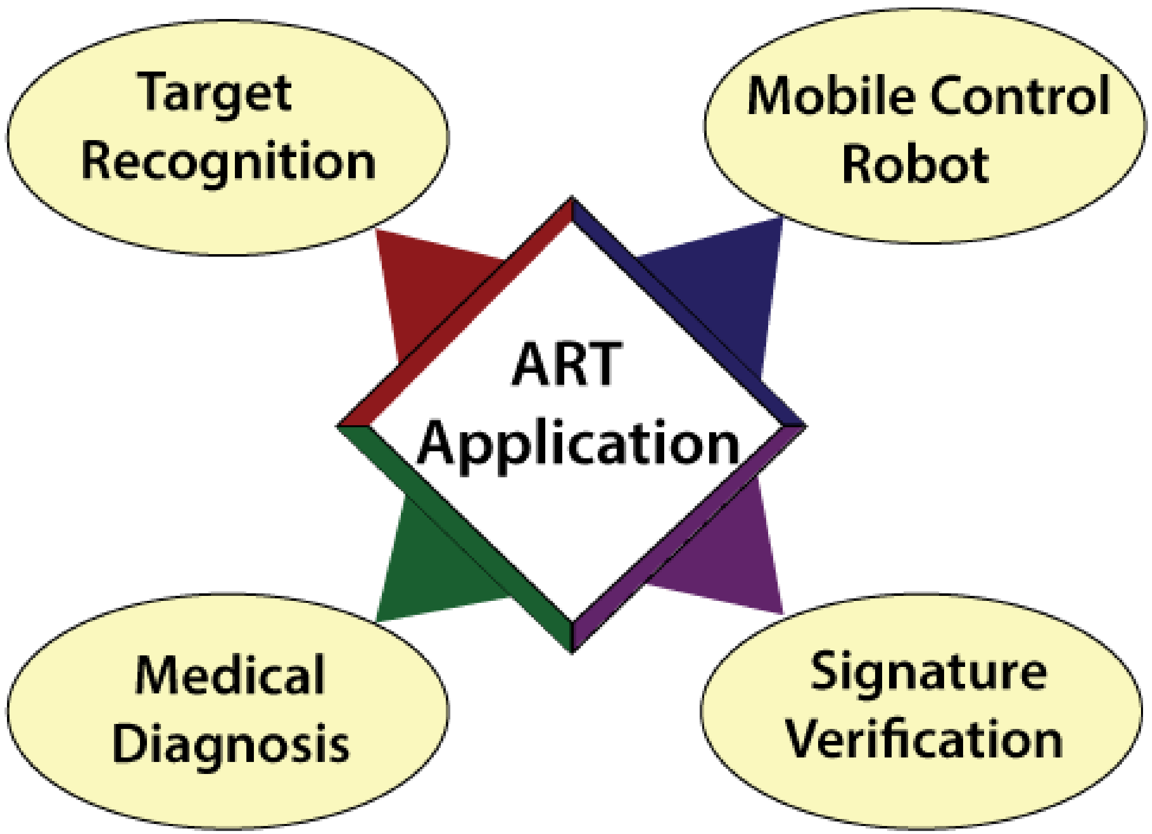
Advantage of adaptive learning theory(ART)

- It can be coordinated and utilized with different techniques to give more precise outcomes.
- It doesn't ensure stability in forming clusters.
- It can be used in different fields such as face recognition, embedded system, and robotics, target recognition, medical diagnosis, signature verification, etc.
- It shows stability and is not disturbed by a wide range of inputs provided to inputs.
- It has got benefits over competitive learning. The competitive learning cant include new clusters when considered necessary.

Applications

- ART neural networks used for fast, stable learning and prediction have been applied in different areas.
- The application incorporates target recognition, face recognition, medical diagnosis, signature verification, mobile control robot.

Auto Associative Memory



Target recognition

- Fuzzy ARTMAP neural network can be used for automatic classification of targets depend on their radar range profiles.
- Tests on synthetic data show the fuzzy ARTMAP can result in substantial savings in memory requirements when related to k nearest neighbor(kNN) classifiers.
- The utilization of multiwavelength profiles mainly improves the performance of both kinds of classifiers.

Medical diagnosis

- Medical databases present huge numbers of challenges found in general information management settings where speed, use, efficiency, and accuracy are the prime concerns.
- A direct objective of improved computer-assisted medicine is to help to deliver intensive care in situations that may be less than ideal.
- Working with these issues has stimulated several ART architecture developments, including ARTMAP-IC.

Signature verification

- Automatic signature verification is a well known and active area of research with various applications such as bank check confirmation, ATM access, etc. the training of the network is finished using ART1 that uses global features as input vector and the verification and recognition phase uses a two-step process.
- In the initial step, the input vector is coordinated with the stored reference vector, which was used as a training set, and in the second step, cluster formation takes place.

Limitations

- Some ART networks are contradictory as they rely on the order of the training data, or upon the learning rate.