Pattern Recognition & Data classification

- **Pattern** is everything around in this digital world. A pattern can either be seen physically or it can be observed mathematically by applying algorithms.
- **Example:** The colors on the clothes, speech pattern etc. In computer science, a pattern is represented using vector feature values.
- **What is Pattern Recognition?**
- **Pattern recognition** is the process of recognizing patterns by using machine learning algorithm. Pattern recognition can be defined as the classification of data based on knowledge already gained or on statistical information extracted from patterns and/or their representation. One of the important aspects of the pattern recognition is its application potential.
- **Examples:** Speech recognition, speaker identification, multimedia document recognition (MDR), automatic medical diagnosis.
  In a typical pattern recognition application, the raw data is processed and converted into a form that is amenable for a machine to use. Pattern recognition involves classification and cluster of patterns.

- In classification, an appropriate class label is assigned to a pattern based on an abstraction that is generated using a set of training patterns or domain knowledge. Classification is used in supervised learning.
- Clustering generated a partition of the data which helps decision making, the specific decision making activity of interest to us. Clustering is used in an unsupervised learning.

Examples

- A pattern is a physical object or an abstract notion. While talking about the classes of animals, a description of an animal would be a pattern. While talking about various types of balls, then a description of a ball is a pattern. In the case balls considered as pattern, the classes could be football, cricket ball, table tennis ball etc. Given a new pattern, the class of the pattern is to be determined. The choice of attributes and representation of patterns is a very important step in pattern classification. A good representation is one which makes use of discriminating attributes and also reduces the computational burden in pattern classification.
- An obvious representation of a pattern will be a **vector**. Each element of the vector can represent one attribute of the pattern. The first element of the vector will contain the value of the first attribute for the pattern being considered.

Example

- While representing spherical objects, (25, 1) may be represented as an spherical object with 25 units of weight and 1 unit diameter. The class label can form a part of the vector. If spherical objects belong to class 1, the vector would be (25, 1, 1), where the first element represents the weight of the object, the second element, the diameter of the object and the third element represents the class of the object.

- **Advantages:**
- Pattern recognition solves classification problems
- Pattern recognition solves the problem of fake bio metric detection.
- It is useful for cloth pattern recognition for visually impaired blind people.
- We can recognize particular object from different angle.
- **Disadvantages:**
- Syntactic Pattern recognition approach is complex to implement and it is very slow process.
- Sometime to get better accuracy, larger dataset is required.
- It cannot explain why a particular object is recognized.
  Example: my face vs my friend’s face.

- **Applications:**
- **Image processing, segmentation and analysis**
  Pattern recognition is used to give human recognition intelligence to machine which is required in image processing.
- **Computer vision**
  Pattern recognition is used to extract meaningful features from given image/video samples and is used in computer vision for various applications like biological and biomedical imaging.
- **Seismic analysis**
  Pattern recognition approach is used for the discovery, imaging, and interpretation of temporal patterns in seismic array recordings. Statistical pattern recognition is implemented and used in different types of seismic analysis models.

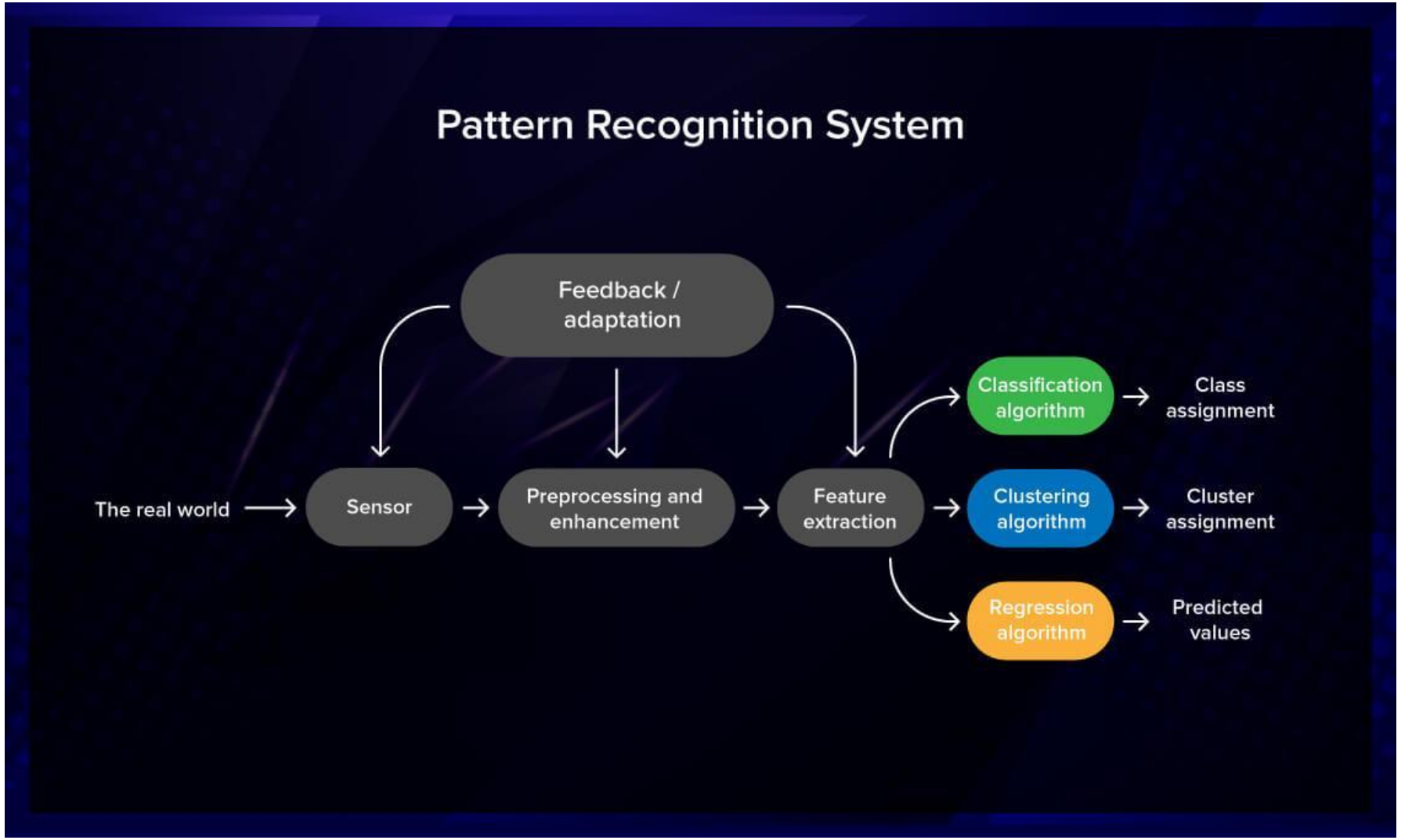
- **Radar signal classification/analysis**
  Pattern recognition and Signal processing methods are used in various applications of radar signal classifications like AP mine detection and identification.
- **Speech recognition**
  The greatest success in speech recognition has been obtained using pattern recognition paradigms. It is used in various algorithms of speech recognition which tries to avoid the problems of using a phoneme level of description and treats larger units such as words as pattern
- **Finger print identification**
  The fingerprint recognition technique is a dominant technology in the biometric market. A number of recognition methods have been used to perform fingerprint matching out of which pattern recognition approaches is widely used.

- Pattern recognition is the process of recognizing regularities in data by a machine that uses machine learning algorithms. In the heart of the process lies the classification of events based on statistical information, historical data, or the machine’s memory.
- For the machine to search for patterns in data, it should be preprocessed and converted into a form that a computer can understand. Then, the researcher can use classification, regression, or clustering algorithms depending on the information available about the problem to get valuable results:

- **Classification.** In classification, the algorithm assigns labels to data based on the predefined features. This is an example of supervised learning.
- **Clustering.** An algorithm splits data into a number of clusters based on the similarity of features. This is an example of unsupervised learning.
- **Regression.** Regression algorithms try to find a relationship between variables and predict unknown dependent variables based on known data. It is based on supervised learning.

- **What should a pattern recognition system be able to do?**

- If you want to assess how good or bad a pattern recognition system is, you need to pay attention to what it can do:
- Identify a familiar pattern quickly and accurately.
- Classify unfamiliar objects.
- Recognize shapes and objects from different angles.
- Uncover patterns and objects, even when partly hidden.
- Automatically recognize patterns.


Pattern Recognition System
Feedback / adaptation
The real world
Sensor
Preprocessing and enhancement
Feature extraction
Classification algorithm
Class assignment
Clustering algorithm
Cluster assignment
Regression algorithm
Predicted values

- A pattern recognition system needs some input from the real world that it perceives with sensors. Such a system can work with any type of data: images, videos, numbers, or texts.
- Having received some information as the input, the algorithm performs **preprocessing**. That is segmenting something interesting from the background. For example, when you are given a group photo and a familiar face attracts your attention, this is preprocessing.
- Preprocessing is tightly connected with **enhancement**. By this term, researchers understand an increase in the ability of a human or a system to recognize patterns even when they are vague. Imagine you are still looking at the same group photo but it is 20 years old. To make sure that the familiar face in the photo is really the person you know, you start comparing their hair, eyes, and mouth. This is when enhancement steps into the game.

- The next component is **feature extraction**. The algorithm uncovers some characteristic traits that are similar to more than one data sample.
- The result of a pattern recognition system will be either a class assignment (if we used classification), or cluster assignment (in case of clustering), or predicted values (if you apply regression).

use of pattern recognition

- Pattern recognition techniques are useful for solving classification problems, uncovering fraud, predicting volcanic eruptions, or diagnosing dangerous diseases with higher accuracy than humans. What is an example of pattern recognition?
- **Image processing, segmentation, and analysis**
- Pattern recognition is used for image processing. For example, a machine learning algorithm can recognize hundreds of bird species better than humans, even if the image is dark or noisy.

- **Computer vision**
- **Speech recognition**
- Virtual assistants such as Alexa or Siri use speech recognition systems to process whole chunks of speech instead of working with separate words of phonemes.
- **Fingerprint identification**
- A number of recognition methods have been used to perform fingerprint matching. The pattern recognition approach is widely used both for criminalistics and in your own smartphone. If you have a fingerprint lock on your phone, that is pattern recognition that steps into the game every time you unlock it.

- **Stock market analysis**
- The stock market is hard to predict. However, even there, there are patterns that can be recognized and used. Modern apps for investors use AI to provide their users with consulting services. Some examples are Blumberg, Tinkoff, Kosho, and SofiWealth.
- **Medical diagnosis**
- Algorithms for pattern recognition trained on real data can be used for cancer diagnosis. histopathology images are used from biopsy for feature extraction and applied an artificial neural network to produce the results.

- **Features** may be represented as continuous, discrete or discrete binary variables. A feature is a function of one or more measurements, computed so that it quantifies some significant characteristics of the object.
- **Example:** consider our face then eyes, ears, nose etc are features of the face.
  A set of features that are taken together, forms the **features vector**.
- **Example:** In the above example of face, if all the features (eyes, ears, nose etc) taken together then the sequence is feature vector([eyes, ears, nose]). Feature vector is the sequence of a features represented as a d-dimensional column vector.

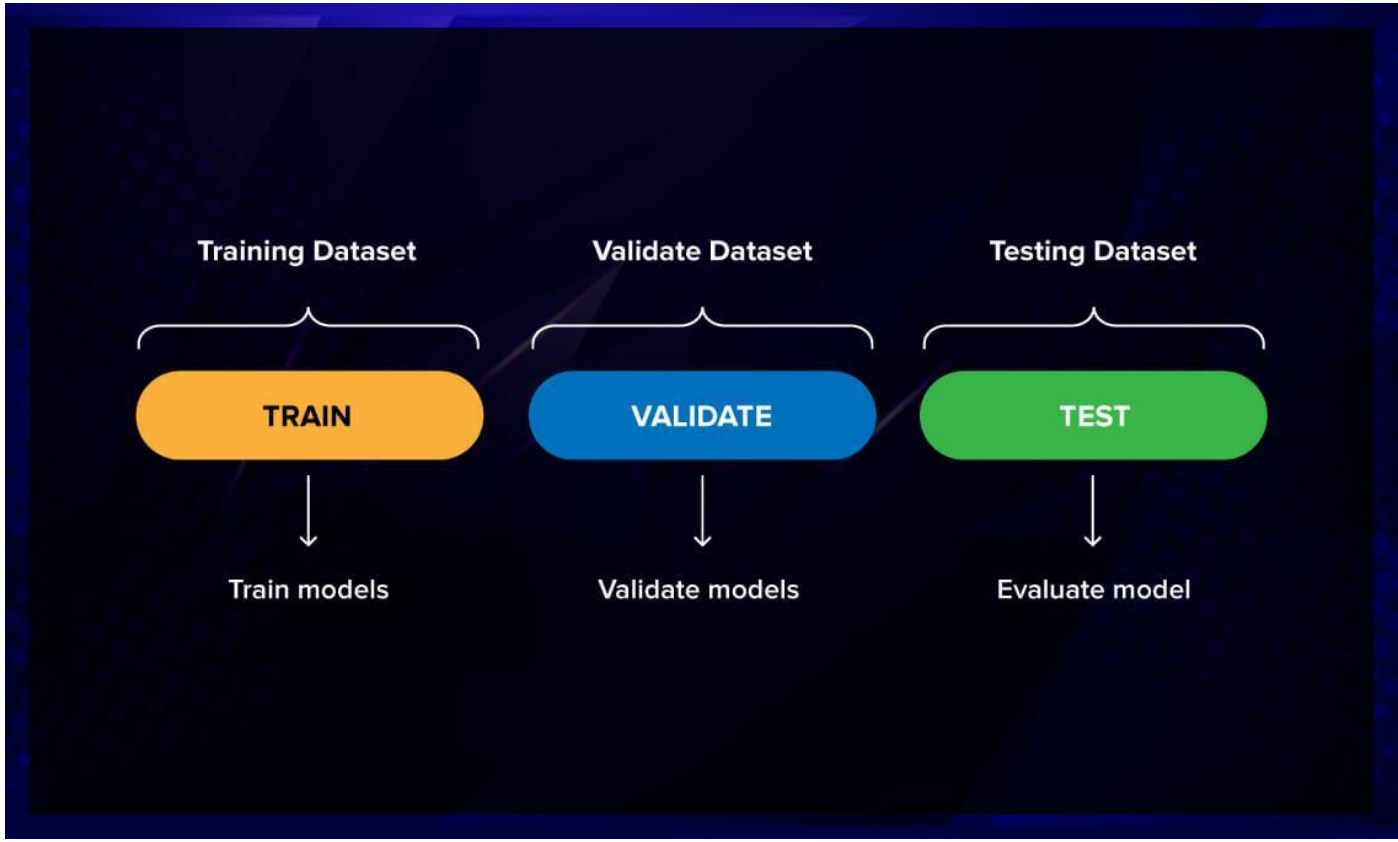
Pattern recognition process

- Pattern recognition system should recognize familiar pattern quickly and accurately
- Recognize and classify unfamiliar objects
- Accurately recognize shapes and objects from different angles
- Identify patterns and objects even when partly hidden
- Recognize patterns quickly with ease, and with automaticity.
- **Training and Learning in Pattern Recognition**
  Learning is a phenomena through which a system gets trained and becomes adaptable to give result in an accurate manner. Learning is the most important phase as how well the system performs on the data provided to the system depends on which algorithms used on the data. Entire dataset is divided into two categories, one which is used in training the model i.e. Training set and the other that is used in testing the model after training, i.e. Testing set.

- **Training set:**
  Training set is used to build a model. It consists of the set of images that are used to train the system. Training rules and algorithms used give relevant information on how to associate input data with output decision. The system is trained by applying these algorithms on the dataset, all the relevant information is extracted from the data and results are obtained. Generally, 80% of the data of the dataset is taken for training data.
- **Testing set:**
  Testing data is used to test the system. It is the set of data which is used to verify whether the system is producing the correct output after being trained or not. Generally, 20% of the data of the dataset is used for testing. Testing data is used to measure the accuracy of the system. Example: a system which identifies which category a particular flower belongs to, is able to identify seven category of flowers correctly out of ten and rest others wrong, then the accuracy is 70 %

Training a pattern recognition system

- To build a pattern recognition system, you need to choose a model and prepare the data. For pattern recognition, neural networks, classification algorithms (Naive Bayes, Decision Tree, Support Vector Machines), or clustering algorithms (k-means, Mean Shift, DBSCAN) are often used.
- Next, you will work with data. Divide it into three sets:



- **Training set.** We use the training set to train the model. You need to select representative samples and make the program process it using training rules. For example, if you are building a security system based on face recognition, you will need a variety of photos of your employees. All the relevant information will be extracted from these data. Generally, 80% of all the data makes a training dataset.
- **Validation set.** This set is used to fine-tune the model. You use it to verify that any increase in accuracy over the training data set will also show an increase in accuracy over a data set that has not been shown to the network before. Because if the accuracy over the training data set increases, but the accuracy over the validation data set stays the same or decreases, then you're over fitting your model and you should stop training it.
- **Testing set.** Testing data is used to test whether the outputs given by the system are accurate. About 20% of the data is used for testing.

Pattern Recognition Tasks Performed by Different Functional Units of ANN

Basic Functional units

- Feedforward ANN
- Feedback ANN
- Feedforward and Feedback (Competitive Learning) ANN

Pattern Recognition Problem

- We have a set of input patterns and the corresponding output patterns.
- Depending on the nature of the output patterns and the nature of the task environment, the problem could be identified as one of association or classification or mapping.
- The given set of input-output pattern pairs form only a few samples of an unknown system. From these samples the pattern recognition model should capture the characteristics of the system.

Pattern Recognition Tasks by Feedforward ANN

- Pattern association
- Pattern classification
- Pattern mapping

Pattern Association

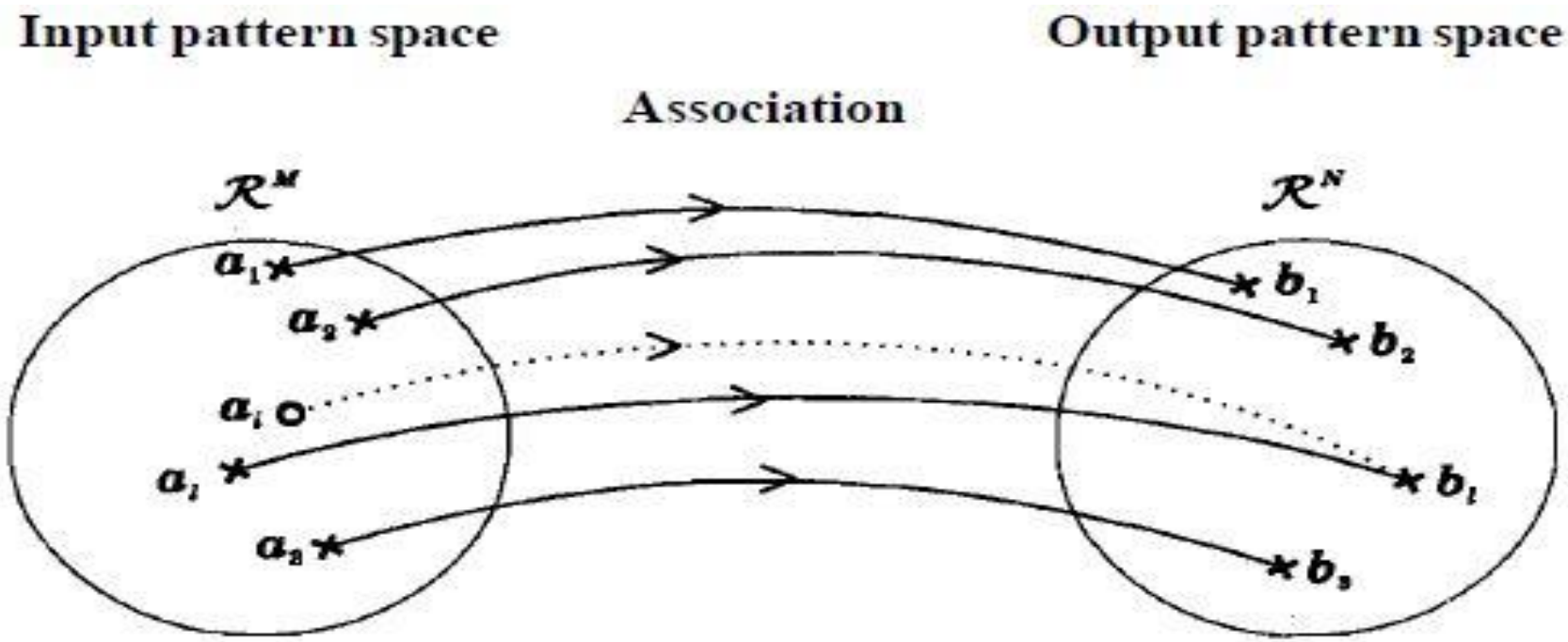
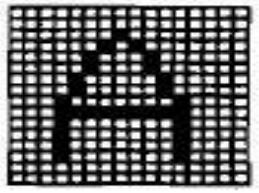
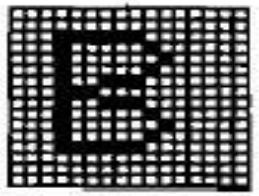


Figure 1:Illustration of pattern association task.

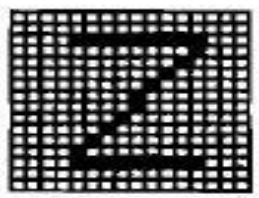
Pattern Association Example



$\rightarrow [0\ 0\ 0\ 0\ 0]^T$



$\rightarrow [0\ 0\ 0\ 0\ 1]^T$



$\rightarrow [1\ 1\ 0\ 0\ 1]^T$

Pattern Classification

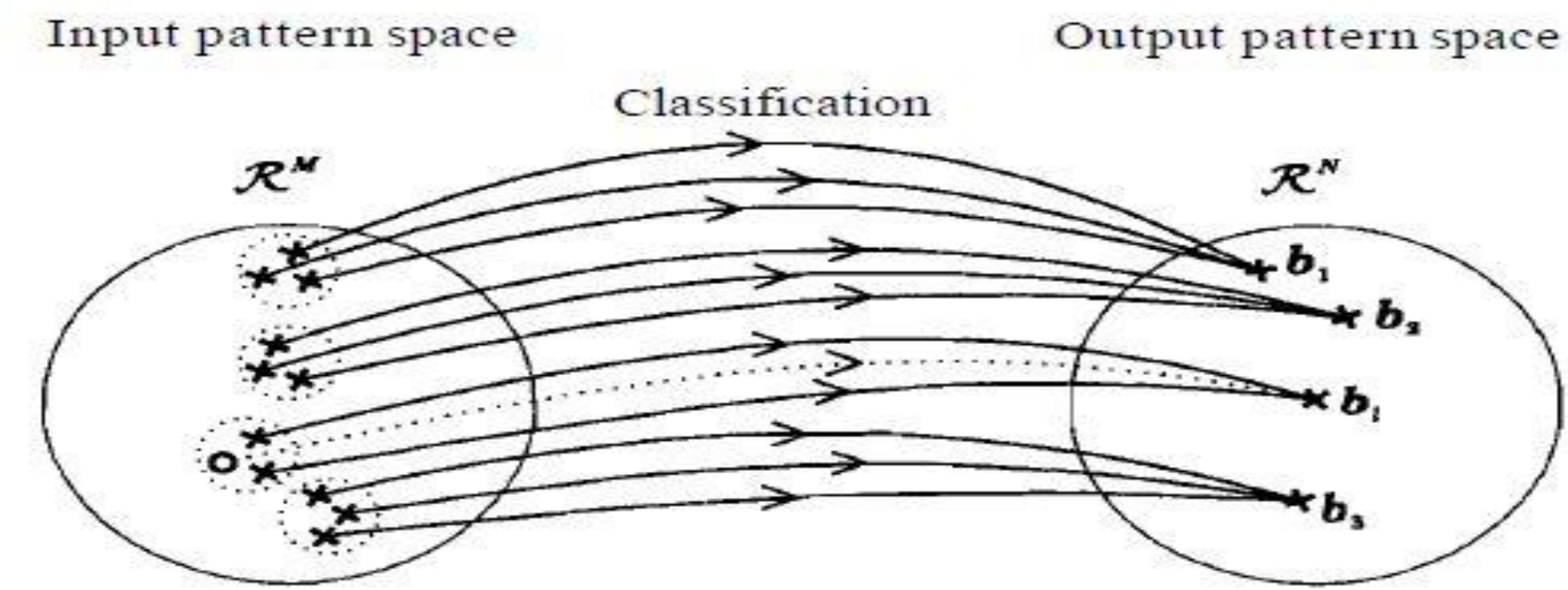
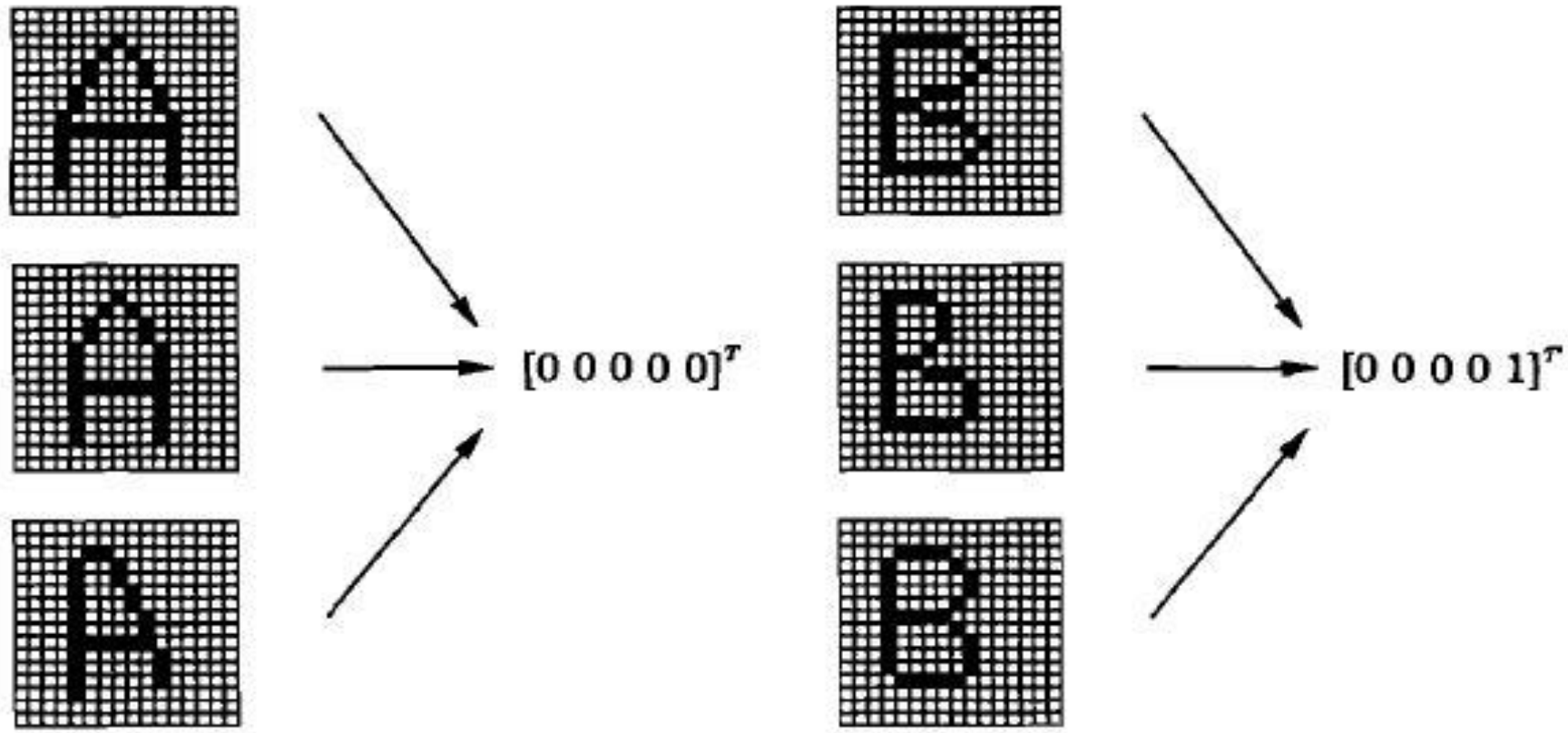


Figure 2:Illustration of pattern classification task

Pattern Classification Example



Pattern Mapping

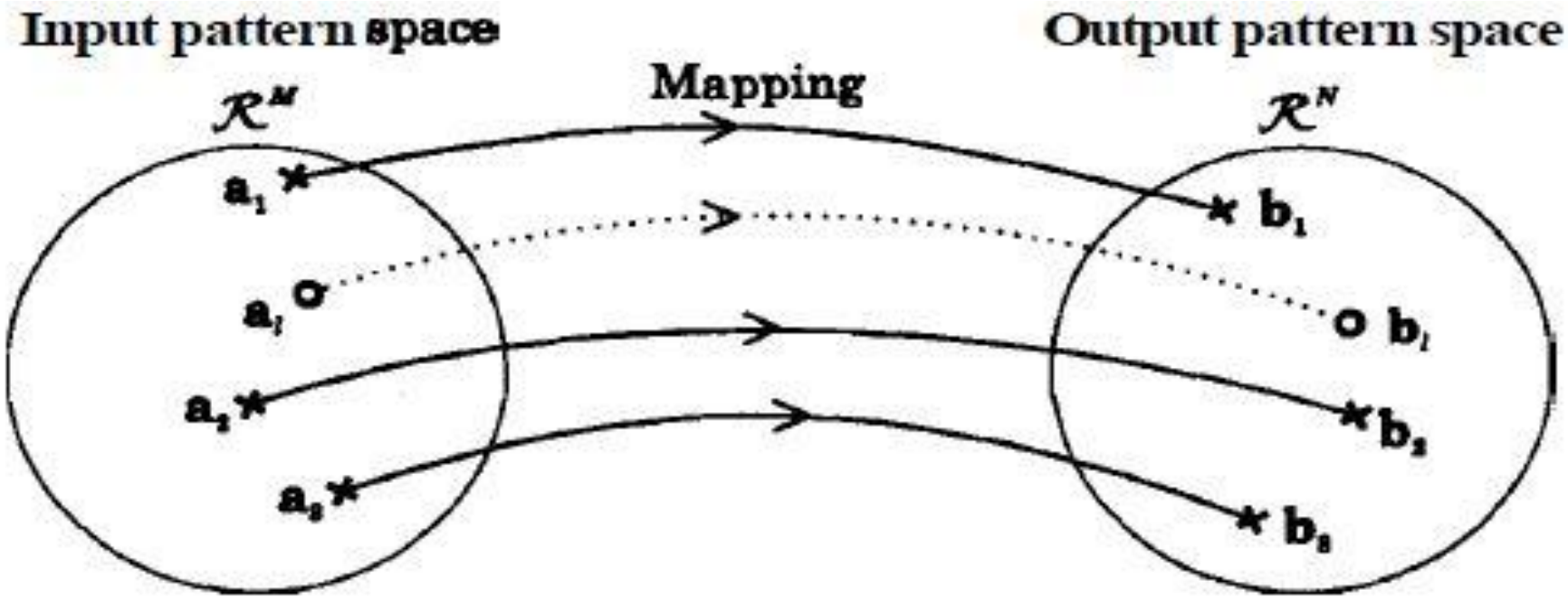


Figure 3: Illustration of pattern mapping task.

Pattern Recognition Tasks by Feedback ANN

- Auto association
- Pattern storage
- Pattern environment storage

Auto association

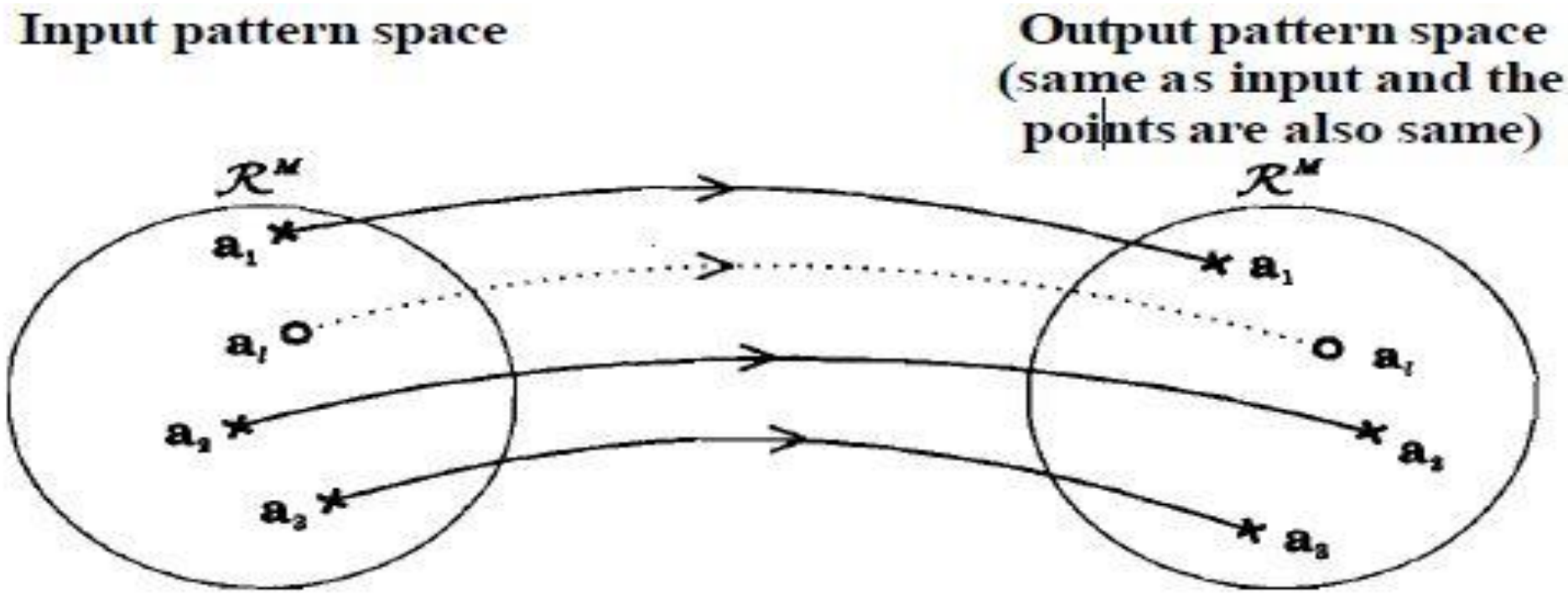


Figure 4: Illustration of pattern autoassociation task.

Pattern storage

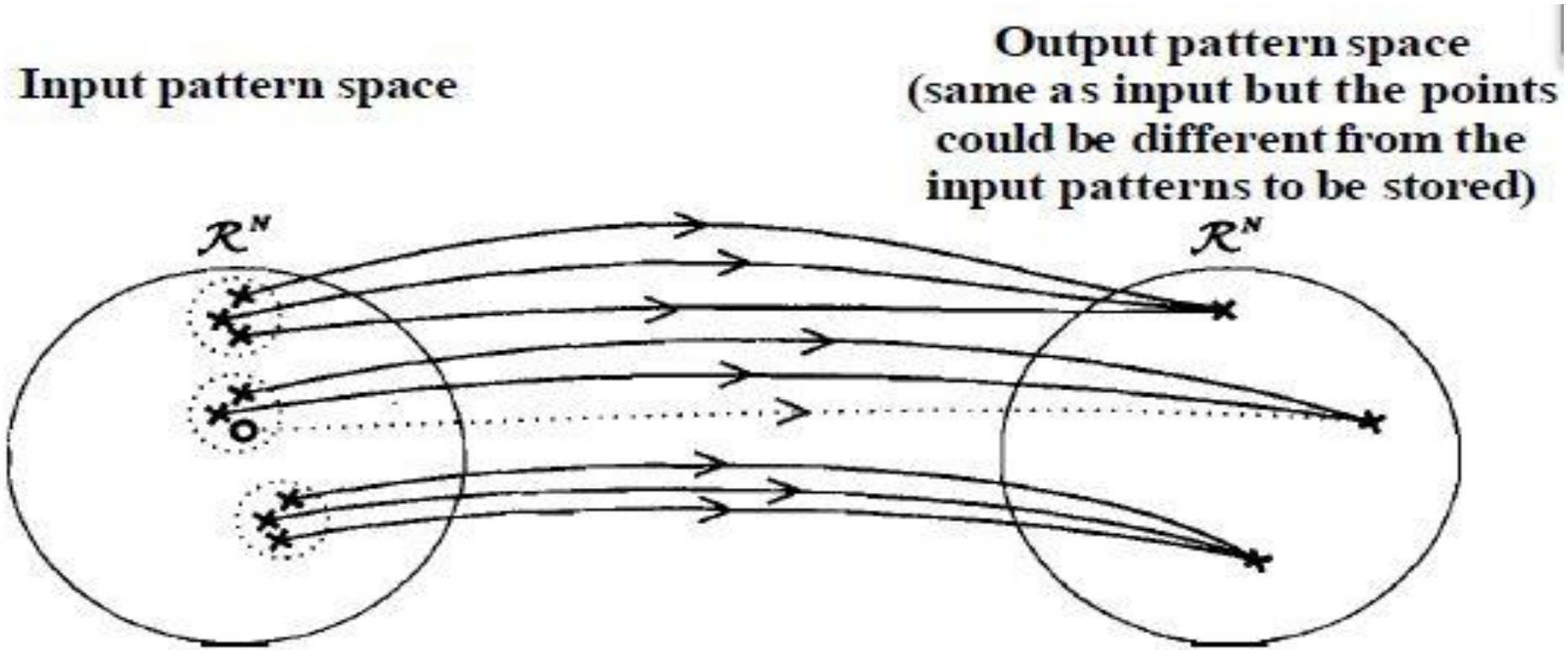


Figure 5: Illustration of pattern storage task

Pattern environment storage

- If a set of patterns together with their probabilities of occurrence are specified, then the resulting specification is called pattern environment.
- The design of a network , to store a given pattern environment aims at recall of the stored patterns with the lowest probability of error. This is called a pattern environment storage problem.

Pattern Recognition Tasks by Competitive Learning ANN

- Temporary pattern storage
- Pattern clustering
- Feature mapping

Temporary Pattern Storage

- If a given input pattern is stored in a network, even in a transformed form, in such a way that the pattern remains only until a new pattern input is given, then the problem becomes that of a short term memory or temporary storage problem.

Pattern Clustering

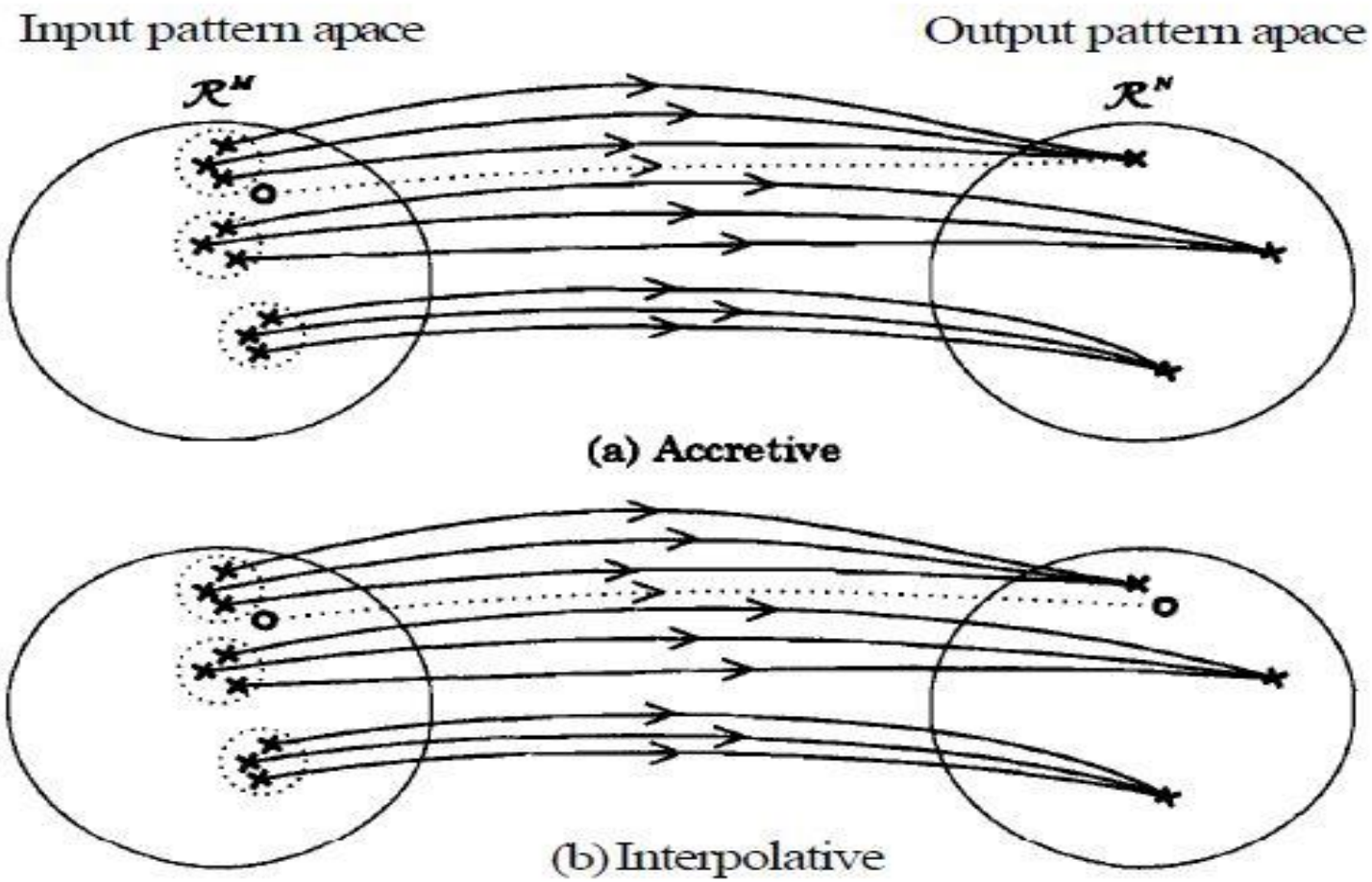


Figure 6: Illustration of two types of pattern clustering tasks.

Feature Mapping

- If similarities of the features of the input patterns have to be retained in the output, the problem becomes one of feature mapping.
- In this, a given set of input patterns are mapped onto output patterns in such a way that the proximity of the output patterns reflect the similarity of the features of the corresponding input patterns.
- When a test input pattern is given, it will generate an output which is in the neighborhood of the outputs for similar patterns.
- Typically the number of output patterns are fixed, but they are much larger than in the pattern clustering case, and they are organized physically in the network in such a way that the neighborhood pattern labels reflect closeness of features.