Data mining, machine learning, and uncertainty reasoning

林偉川

Memory

- Memory refers to the enduring neural alternations induced by the interaction of an organism with its environment
- It must be accessible to the nervous system to influence future behavior
- An activity pattern must be stored in memory through a learning process
- Memory and learning are intricately connected

2

Memory

- When a activity pattern is learned, it stored in the brain where it can be recalled later when required
- Memory is divided into “short-term” and “long-term” memory, depending on the retention time

3

Memory

- Short-term memory refers to a compilation of knowledge representing the “current” state of the environment.
- Any discrepancies between knowledge stored in short-term memory and a “new” state are used to update the short-term memory
- Long-term memory refers to knowledge stored for a long time or permanently

4

Memory characteristics

- Memory is distributed
- Both the stimulus pattern and the response (stored) pattern of an associated memory consist of data vectors
- Information is stored in memory by setting up a spatial pattern of neural activities across a large number of neurons
- Information contained in a stimulus not only determines its storage location in memory but also an address for its retrieval

5

Memory characteristics

- Memory exhibits a high degree of resistance to noise and damage
- There may be interactions between individual patterns stored in memory (otherwise the memory would have to be large for it to accommodate the storage of a large number of patterns in perfect isolation from each other)

6

Memory characteristics

- There is the distinct possibility for the memory to make errors during the recall process

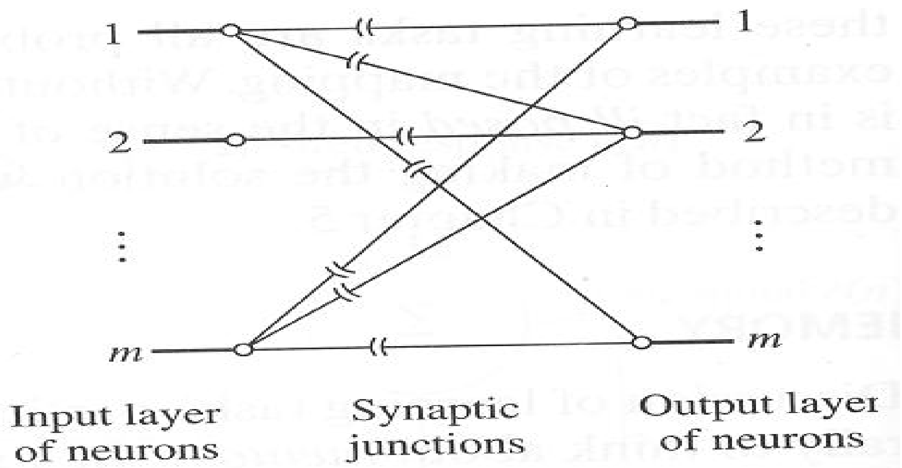
7

Distributed Memory

- The neural activities form a spatial pattern inside the memory that contains information about the stimuli
- The memory is said to perform a distributed mapping that transforms an activity pattern in the input space into another activity pattern in the output space (synapases ➔兩個神經原的相接處)

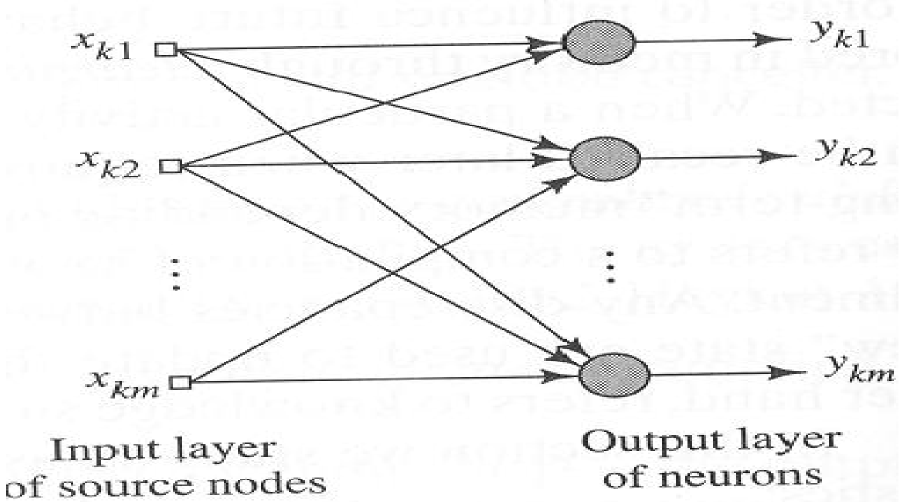
8

Associative memory model component of a nervous system



(a) Associative memory model component of a nervous system

Associative memory model using artificial neuron



(b) Associative memory model using artificial neurons

Another types of learning methods

- Another categorizes the learning methods into another group, off-line or on-line.
- When the system uses input data to change its weights to learn the domain knowledge, the system could be in training mode or learning mode. (on-line)
- When the system is being used as a decision aid to make recommendations, it is in the operation mode, this is also sometimes called recall. (off-line)

11

On-line learning methods

- In on-line or real time learning, when the system is in operating mode (recall), it continues to learn while being used as a decision tool. This type of learning has a more complex design structure.

12

Off-line learning methods

- In the off-line learning methods, once the systems enters into the operation mode, its weights are fixed and do not change any more. Most of the networks are of the off-line learning type. ➔ agent mode or avatar

EX. The movie of 13^{th} floor

The on-line or off-line game

13

Artificial discovery

- Supervised learning where the learner seeks to develop a concept description from examples that have been pre-classified by the teacher
- Unsupervised learning whose task is to generate conceptual taxonomies from non-classified objects

14

Learning types

- The learning ability of a neural network is determined by its architecture and by the algorithmic method chosen for training.
- The training method usually consists of one of three schemes: Unsupervised learning, Reinforcement learning, Back Propagation (BP)

15

Learning types

- The brain basically learns from experience. Neural networks are sometimes called machine learning algorithms
- Changing of its connection weights (training) causes the network to learn the solution to a problem.
- The strength of connection between the neurons is stored as a weight-value for the specific connection. The system learns new knowledge by adjusting these connection weights.

16

Unsupervised learning

- The hidden neurons must find a way to organize themselves without help from the outside.
- In this approach, no sample outputs are provided to the network against which it can measure its predictive performance for a given vector of inputs. This is learning by doing.

17

Reinforcement learning

- The neural network method works on reinforcement from the outside. The connections among the neurons in the hidden layer are randomly arranged, reshuffled as the network is told how close it is to solving the problem.

18

Reinforcement learning

- Reinforcement learning is also called supervised learning, because it requires a teacher. The teacher may be a training set of data or an observer who grades the performance of the network results.
- Both unsupervised and reinforcement suffer from relative slowness and inefficiency relying on a random shuffling to find the proper connection weights.

19

BP Learning

- This method is proven highly successful in training of multilayered neural nets. The network is not just given reinforcement for how it is doing on a task.
- Information about errors is also filtered back through the system and is used to adjust the connections between the layers, thus improving performance. Another kinds of supervised learning.

20

BP Learning

- Since the real uniqueness or 'intelligence' of the network exists in the values of the weights between neurons, we need a method of adjusting the weights to solve a particular problem.

BP Learning

- A BP network learns by example, that is, we must provide a learning set that consists of some input examples and the known-correct output for each case.
- We use these input-output examples to show the network what type of behavior is expected, and the BP algorithm allows the network to adapt.

BP Learning

- The BP learning process works in small iterative steps: one of the example cases is applied to the network, and the network produces some output based on the current state of it's synaptic weights (initially, the output will be random).
- This output is compared to the known-good output, and a mean-squared error signal is calculated. The error value is then propagated backwards through the network, and small changes are made to the weights in each layer.

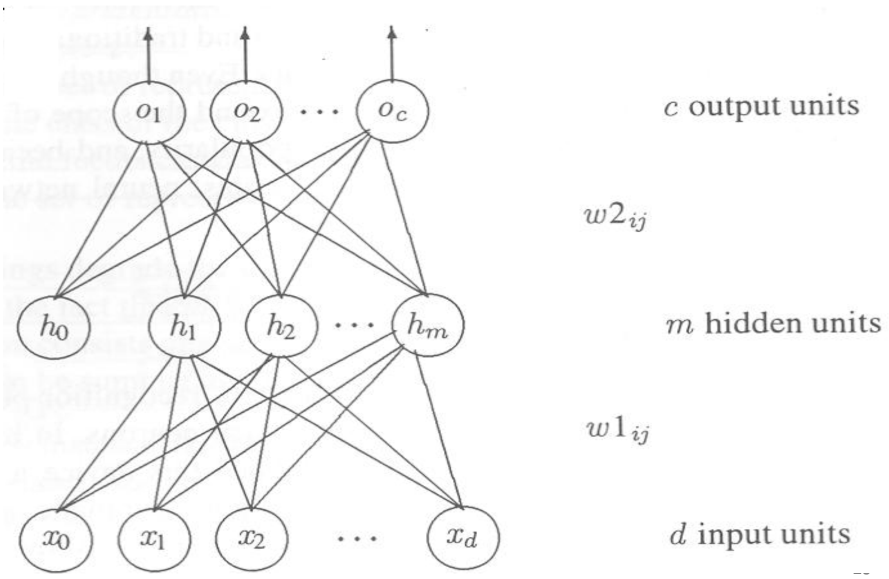
23

BP Learning

- The weight changes are calculated to reduce the error signal for the case in question. The whole process is repeated for each of the example cases, then back to the first case again, and so on. The cycle is repeated until the overall error value drops below some pre-determined threshold.
- At this point we say that the network has learned the problem "well enough" - the network will never exactly learn the ideal function, but rather it will asymptotically approach the ideal function.

24

Multilayer perceptron



BP Learning

- In many complicated tasks, researchers made good experience when they used two or more hidden layers
- BP learning is a kind of automatic adjustment of the weight in the neural network
- $W1_{12}$ 表示由x_1到h_2間的Weight
- $W2_{12}$ 表示由h_1到o_2間的Weight

26

BP Learning algorithm

1. Define the configuration of the neural net in terms of the number of units in each layer
2. Set the initial weight $w1_{ij}$ and $w2_{ij}$ to small random values, for the interval [-0.1,0.1]
3. Select an example and denote its attribute values by $x_1, \ldots, x_k$. Attach the example to the input layer

27

BP Learning algorithm

4. Propagate the input values from the input layer to the hidden layer. The output value of the jth unit in the hidden layer is calculated by the function:

$$h_j = \frac{1}{1 + e^{-\sum_i \boxed{w1_{ij} \bullet x_i}}}$$

Sum up from input
Multiply weight

Propagate the values obtained to the output layer. The output value of the jth unit in this layer is calculated by the function:

$$o_j = \frac{1}{1 + e^{-\sum_i \boxed{w2_{ij} \bullet h_i}}}$$

Sum up from hidden layer
Multiply weight

28

BP Learning algorithm

5. Compare the outputs o_j with the teacher's classifications y_j; calculate the correction error as $\delta 2_j = o_j(1- o_j)(y_j - o_j)$

   if the error is under some predefined threshold then stop

   adjust the weights by the following formula: $w2_{ij}(t+1)=w2_{ij}(t)+ \delta 2_j * h_i * \eta$, where $w2_{ij}(t)$ are the respective weight values at time t and η is a constant such that $\eta \in (0,1)$

29

BP Learning algorithm

6. Calculate the correction error for the hidden layer by means of the formula $\delta 1_j = h_j(1- h_j)\Sigma_i \delta 2_i * w2_{ij}$ and adjust the weight $w1_{ij}$ by the formula $w1_{ij}(t+1)=w1_{ij}(t)+ \delta 1_j * x_j * \eta$
7. Go to step 3

30

Neural Network's Learning laws

- There are a variety of learning laws which are in common use. These laws are mathematical algorithms used to update the connection weights.
- **Hebb's Rule in 1949**
  The first and the best known learning rule. This basic rule is: If a neuron receives an input from another neuron, and if both are highly active (mathematically have the same sign), the weight between the neurons should be strengthened.

31

Learning laws -- **Hopfield Law**

- **Hopfield Law**
  This law is similar to Hebb's Rule with the exception that it specifies the magnitude of the strengthening or weakening.

32

Learning laws -- **Hopfield Law**

- It states, "if the desired output and the input are both active or both inactive, increment the connection weight by the learning rate, otherwise decrement the weight by the learning rate." (Most learning functions have some provision for a learning rate, or a learning constant. Usually this term is positive and between 0 and 1.)

33

Learning laws -- Delta Rule

- The Delta Rule is a further variation of Hebb's Rule. This rule is based on the idea of continuously modifying the strengths of the input connections to reduce the difference (the delta) between the desired output value and the actual output of a neuron.

34

Learning laws -- Delta Rule

- This rule changes the connection weights in the way that minimizes the mean squared error 【誤差均方】 of the network. The error is back propagated into previous layers one layer at a time.

35

Learning laws -- Delta Rule

- The process of back-propagating the network errors continues until the first layer is reached. The network type called Feed forward.
- Back-propagation derives its name from this method of computing the error term. ➔Windrow-Hoff Learning Rule and the Least Mean Square Learning Rule.

36

Learning laws -- **Kohonen's Rule**

- **Kohonen's Learning Law**
  In this procedure, the neurons compete for the opportunity to learn, or to update their weights.

37

Learning laws -- **Kohonen's Rule**

- The processing neuron with the largest output is declared the winner and has the capability of inhibiting its competitors as well as exciting its neighbors.
- Only the winner is permitted output, and only the winner plus its neighbors are allowed to update their connection weights.

38

Where are Neural Networks being used

- Basically, most applications of neural networks fall into the following five categories: **Prediction, Classification, Data association, Data Conceptualization, Data Filtering, Data mining**

39

Where are Neural Networks being used

- **Prediction**
  Uses input values to predict some output. e.g. pick the best stocks in the market, predict weather, identify people with cancer risk.
- **Classification**
  Use input values to determine the classification. e.g. Is the input the letter A, is the blob of the video data a plane and what kind of plane is it.

40

Where are Neural Networks being used

- **Data association**
  Like classification but it also recognizes data that contains errors. E.g. not only identify the characters that were scanned but identify when the scanner is not working properly.
- **Data Conceptualization**
  Analyze the inputs so that grouping relationships can be inferred. e.g. extract from a database the names of those most likely to by a particular product.

41

Where are Neural Networks being used

- **Data Filtering**
  Smooth an input signal. e.g. take the noise out of a telephone signal.

42

Using Artificial neural network for data mining

- Artificial neural networks are useful for analytical problems such as:
  - A large amount of example data is available and it is difficult to specify a parametric model for the data
  - High input dimension and relationships exist within the data that are not fully understood, with many potential models that could be specified

43

Using Artificial neural network for data mining

- There are potentially stable patterns in the data that are subtle or deeply hidden
- The data exhibit significant un-characterizable non-linearity
- Iterative use of the data is required to detect patterns
- Problems are solved by generating predictions of complicated phenomena rather than by generating explanations

44

Conclusion of BP Learning algorithm

- The algorithm captures only the fundamental principle of learning in multilayer perceptrons, and its practical use in many realistic application suffers from various shortcomings and pitfalls that the user must be acquainted with