Master of Technology

Unit 2/6: Computational Intelligence I

Introduction to Computational Intelligence and Neural Networks

Dr. Zhu Fangming
Institute of Systems Science
National University of Singapore

Objectives

- To provide a basic introduction to computational intelligence and neural networks
- To provide biological and historical motivation for neural network computing
- To discuss general architecture of neural networks and application examples

Outline

- Introduction to computational intelligence
- Introduction to neural networks
- Biological neuron and artificial neuron
- Pioneering research work of neural networks
- General architecture of neural networks
- NN learning categories
- Application examples

Computational Intelligence (CI)

- A broad definition of CI is the study of
  - » adaptive mechanisms to enable or facilitate "*intelligent*" behavior of machine in complex, uncertain, and changing environments
    - ♦ to learn or adapt to new situations, to generalize, abstract, discover and associate
- *CI* is understood as solving various problems of AI with the use of computer to perform numerical calculations
- Such computations are connected with application of the techniques such as:
  - » Neural networks, Fuzzy logic, Evolutionary algorithms, Rough sets, etc.

CI & Soft Computing (SC)

- SC includes principal members:
  - » fuzzy logic (FL), neural networks (NN), evolutionary computation (EC)
    - ♦ Has neuro-fuzzy techniques as the back-bone
    - ♦ Has neural networks to help automatic learning and knowledge construction of knowledge based systems (KBS)
    - ♦ Has EC to enhance the performance of KBS

- CI and SC are sometimes used interchangeably, when indicating the major techniques of the field

CI Research Activities

- IEEE Computational Intelligence Society
  - » http://cis.ieee.org/
  - » Organizes numerous conferences in the field
  - » Publishes journals & magazines
    - ♦ IEEE Transactions on *Neural Networks*
    - ♦ IEEE Transactions on *Fuzzy Systems*
    - ♦ IEEE Transactions on *Evolutionary Computation*
    - ♦ *Computational Intelligence* magazine

Introduction to Neural Networks

- **In many real-world applications, computers are expected to perform complex pattern recognition tasks.**
- **Complex pattern recognition involves thinking & learning**
- **Learning involves both memorising and generalising**
- **Recognition of complex patterns needs parallel processing**
  - » **human can easily handle**
  - » **conventional computing paradigms are not suitable to solve this type of problems**
- **We therefore borrow features from physiology of brain as the basis for our new processing models — *Artificial Neural Networks* (ANN) or *Neural Networks* (NN)**

Introduction to Neural Networks

- **Computers can do things being exactly told by programming**

- **Computers can make "inferences" for decision making with complete rule set. However, it is necessary to "tell"computers the relationship between a given situation and a conclusion or an action to be suggested**

- **As conventional expert systems, computers will not generate new conclusion, and will not "learn" new knowledge automatically**

- **Can computer really do human-like "thinking"and "learning" by itself ?**

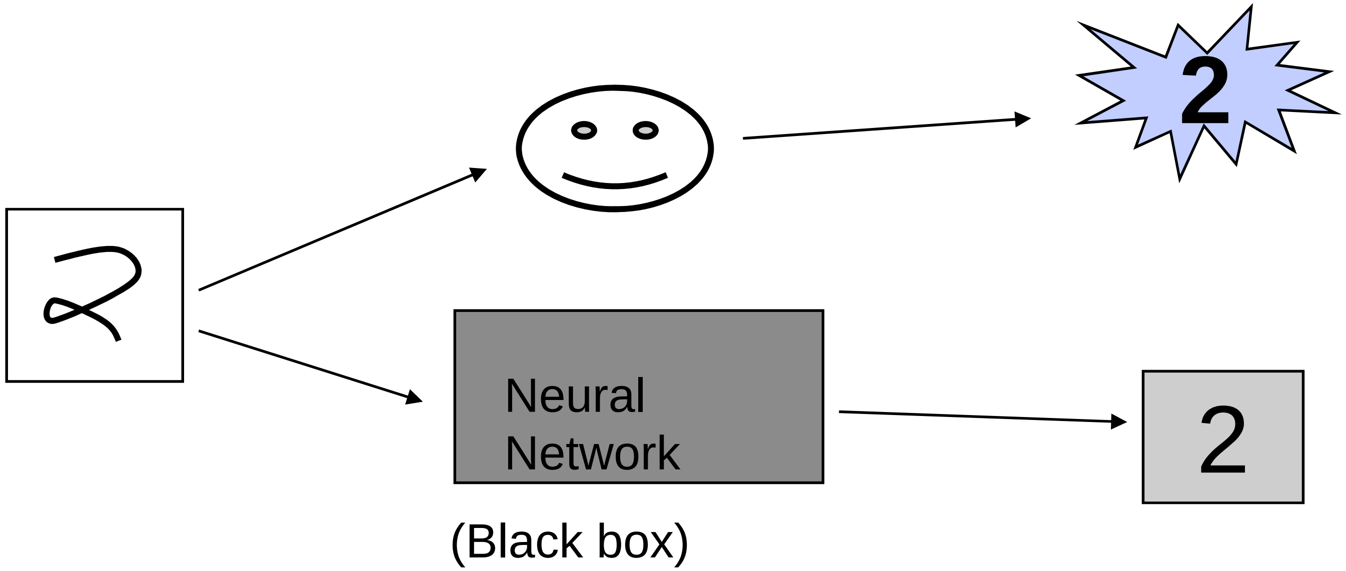
  — Neural Networks provide an affirmative answer

- **Learning from examples is the hallmark of any neural network**

Example: Character Recognition

- **Example:**

Character Recognition by Human Brain and NN



Some Definitions of Artificial Neural Networks

- **... a neural network is a system composed of many simple processing elements operating in parallel whose function is determined by network structure, connection strengths, and the processing performed at computing elements or nodes.** ***--------DARPA Neural Network Study*** **(1988)**

- **A neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:**

  » **Knowledge is acquired by the network through a learning process.**

  » **Inter neuron connection strengths known as synaptic weights are used to store the knowledge.**

----Haykin (1994)

Some Definitions of Artificial Neural Networks…

- **A neural network is a circuit composed of a very large number of simple processing elements that are neural based. Each element operates only on local information. Furthermore each element operates asynchronously; thus there is no overall system clock.**

-----Nigrin (1993)

- **Artificial neural systems, or neural networks, are physical cellular systems which can acquire, store, and utilize experiential knowledge.**

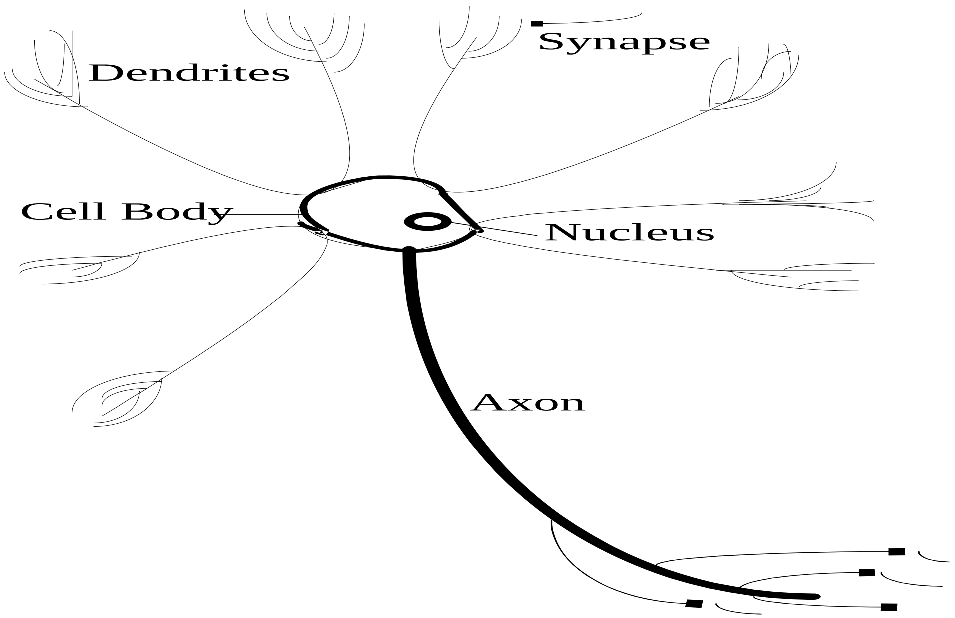
-----Zurada (1992)

From Biological Neuron to Artificial Neuron
— Biological Inspiration for Artificial Neural Networks

- **The basic biological computing element — *the neuron***

  This extremely small computer is a multiple signal processor based on

  electrochemical processing principles



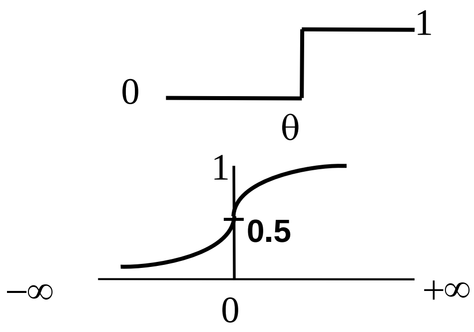
From Biological Neuron to Artificial Neuron ...

- **A biological neuron is the basic biological computing element**
- **A neuron is a small cell that receives electrochemical stimuli from multiple sources and responds by generating electrical impulses that are transmitted to other neurons or effector cells**
- **There are something like 10^{10} to 10^{12} neurons in the human nervous system and each is capable of storing several bits of “information”**
- **About 10% of the neurons are input and output, the remaining 90% are interconnected with other neurons which store information or perform various transformations on the signals being propagated through the network**

From Biological Neuron to Artificial Neuron
— An Artificial Neuron

An artificial neuron — **A Single Neural Computing Element**

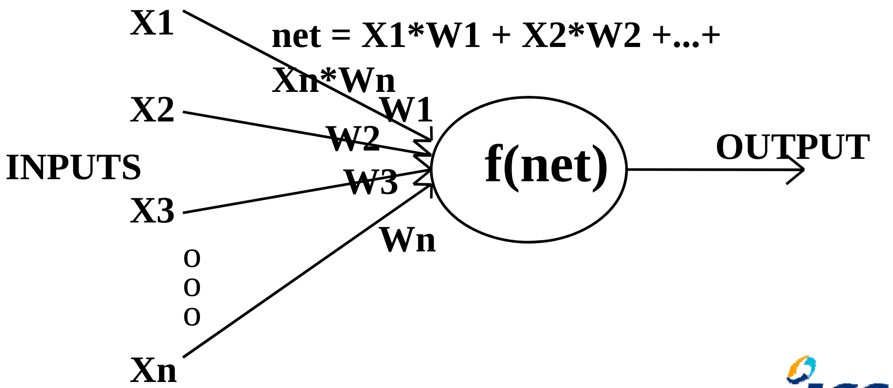
- **N input signals each weighted for its importance**
- **Signals are added to produce a cumulative (net) input**
- **The neuron transforms the net input to produce an output**
  - » **net = $\Sigma_i x_i w_i$**
- **Activation / transfer function**
  - » **hard-limiting function**



$$f(\text{net}) = \begin{cases} 0 & \text{if net} < \text{threshold} \\ 1 & \text{otherwise} \end{cases}$$

  - » **sigmoid function**

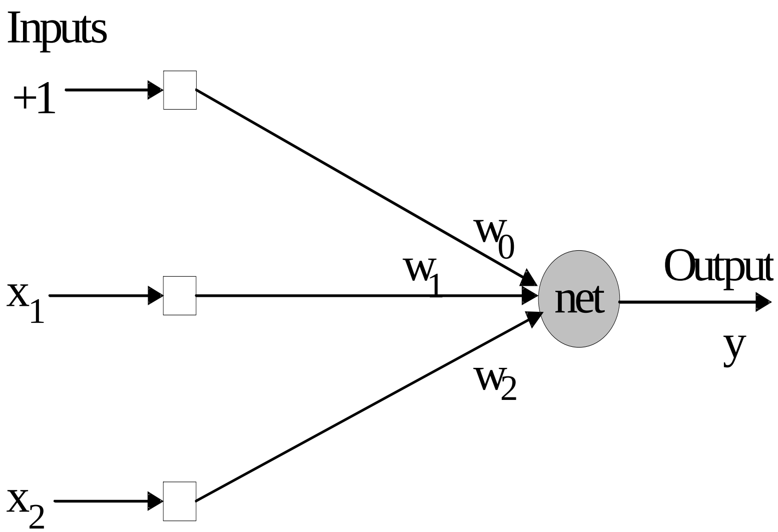
$$f(\text{net}) = \frac{1}{1 + e^{-net}}$$



An Example of Simple NN
— two inputs & bias —

- **The bias input is a fixed positive signal. The weight w_0 on the bias input is also adjustable**
- $y = f(\text{net}) = f(w_0 + x_1w_1 + x_2w_2)$
- **Suppose the activation function is a hard-limiting function:**

$$f(x) = \begin{cases} 0 & \text{if } x \le 0 \\ 1 & \text{otherwise} \end{cases}$$



- **Instances:**
  - » $x_1 = 1, x_2 = 0, w_0 = -3, w_1 = 2, w_2 = 6$, net = -1, y = 0
  - » $x_1 = 1, x_2 = 0, w_0 = -2, w_1 = 3, w_2 = 5$, net = 1, y = 1

Why & How NN Can Learn

- **From the simple NN example, we can see that given the same inputs, the output will be different when different weights are used.**

- **Why NN can learn**
  - » **by changing weights, the behaviour of neural network can be changed**

- **How NN can learn**
  - » **changing weights purposely to reduce the difference between the actual output and the expected output**

What NN Can Do

- **NNs can learn to model complex systems from examples**
- **NNs are good at difficult perception tasks like vision or speech understanding**
- **NNs can often outperform other methods such as statistical pattern recognition and regression techniques**
- **NNs can mimic many intelligent traits found in humans:**

  learning, generalisation, tolerance to missing or noisy data, associative recall
- **Parallel processing is more powerful and faster than sequential processing**
- **NNs are tolerant to faults. They are more reliable and robust than conventional systems**
- **VLSI technology makes hardware models of NNs cheap and practical as intelligent computing systems**

Applications of Neural Networks

- Some of the successful applications of neural networks
  - » Content addressable memories
  - » Process control
  - » Data compression
  - » Fault diagnosis
  - » Forecasting time series
  - » General mapping
  - » Functional approximation
  - » Multi-sensor data fusion
  - » Optimization
  - » Pattern recognition
  - » Image Processing
  - » Natural language processing
  - » Data mining
  - » Data visualization

- Areas
  - » All branches of Science and Engineering
  - » Economics, Finance, Management, Social Sciences
  - » Medicine, Agriculture

Some of Pioneering Research Work of Neural Networks

- **W. McCulloch & W. Pitts (1943): Neurons as computers**
  - » **described a logical calculus of neural networks that united the studies of neurophysiology and mathematical logic**
  - » **showed that a network with a sufficient number of simple units, and synaptic connections would, in principle, compute any function**
  - » **Influenced by this idea, von Neumann used idealized switch-delay elements derived from the McCulloch-Pitts neuron for the development of first generation of computer**
- **D. Hebb (1949): Neural learning theory**
  - » **postulate of learning: the effectiveness of a variable synapse between two neurons is increased by the repeated activation on one neuron by the other across that synapse**
  - » **provided a source of inspiration for the development of computational models of learning and adaptive systems**
- **M. Minsky (1954, 1961):**
  - » **Theory of Neural-Analog reinforcement system**
  - » **neural network**

Some of Pioneering Research Work of Neural Networks ...

- W.K. Taylor (1956): Associative memory
  - » contributions also from Steinbuch, Willshaw and others
  - » Anderson (1972), Kohonen (1972), Nakano (1972): correlation matrix memory
- F. Rosenblatt (1958), M. Minsky & S. Papert (1969): Perceptrons
  - » a novel method of supervised learning
  - » perceptron convergence theorem (Rosenblatt 1960)
- B. Widrow & Hoff (1960): Least mean-square (LSM) algorithm
  - » Adaline (ADAptive LINear Element)
  - » Madaline (Multiple ADALINE) (Widrow 1962): one of the earliest trainable layered neural networks
- S. Amari (1967): Stochastic gradient method for adaptive pattern classification
- S. Grossberg (1970s, 1980): Adaptive Resonance Theory (ART)
  - » established a new principle of self-organization

Some of Pioneering Research Work of Neural Networks ...

- **L. Cooper et al. (1982, 1987): Reduced Coulomb Energy networks**
  - » **an architecture with incremental learning capability**
- **J. Hopfield (1982): Recurrent Hopfield Nets**
  - » **the principle of storing information in dynamically stable network**
- **T. Kohonen (1982): Self-organizing Maps**
  - » **using a one- or two-dimensional lattice structure**
- **Barto, Sutton, & Anderson (1983): Reinforcement learning**
- **D. Rumelhart & et al. (1986): back-propagation algorithm (BP)**
- **R. Hecht-Nielsen (1987): counter propagation neural network**
- **Jang, Pal and Group, Keller and others (1992)**
  - » **Fuzzy Neural Nets, Neuro-Fuzzy Systems etc**

Some of Pioneering Research Work of Neural Networks ...

- **Broomhead & Lowe (1988), Moody & Darken (1989): Radial Basis Function (RBF) network**
  - » **provided an alternative to multilayer perceptrons**
- **P. Baldi and K. Hornik (1989): Auto associative neural networks**
- **D.F. Specht (1991): General Regression Neural Networks (GRNN)**
- **A. M. Kramer (1991): Nonlinear principal component neural networks**
- **Nikola Kasabov (1998): Evolving connectionist systems (ECOS) for real time data mining**
- **Geoffrey Hinton, Yann LeCun, et al. Since 2000, Deep Neural Network, Convolutional Neural Networks, Deep Learning…**

...

National University of Singapore

General Architecture of Neural Networks

- **A neural network has a parallel-distributed architecture**
  - » **with a large number of nodes and connections**
  - » **each connection points from one node to another and is associated with a weight**
- **Construction of a neural network involves the following tasks:**
  - » **determine the network properties:**
    - ♦ **the network topology (framework as well as interconnection scheme)**
    - ♦ **the type of connections**
    - ♦ **the range of weight**
  - » **determine the node properties:**
    - ♦ **the activation range**
    - ♦ **the activation (transfer) function**
  - » **determine the system dynamics:**
    - ♦ **the weight initialisation scheme**
    - ♦ **the activation-calculating formula**
    - ♦ **the learning rule (algorithm)**

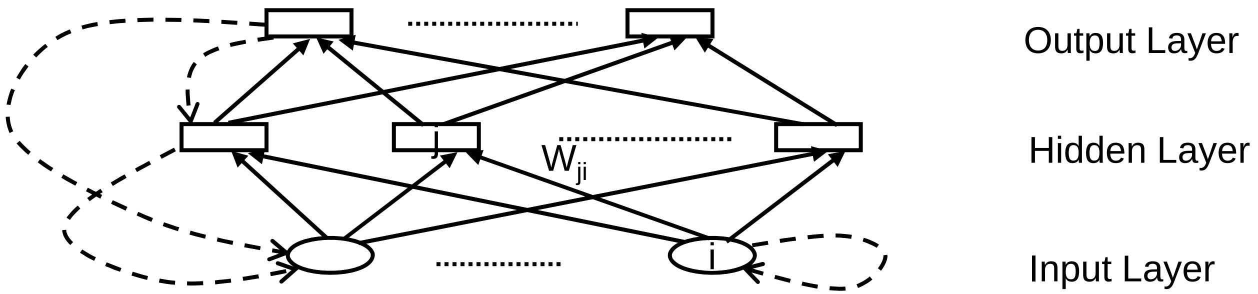
ISS National University of Singapore

General Architecture of Neural Networks (cont.)
— Network Properties —

- **Framework (in general, but not for all NNs)**
  - » **the input layer — input units (nodes)**
    - ♦ **encode the instance presented to the network for processing**
      - **example: attribute values of an object**
  - » **the hidden layer — hidden units (nodes)**
    - ♦ **are not directly observable**
    - ♦ **provide non-linearity for the network**
    - ♦ **an NN may have more than one hidden layers**
  - » **the output layer — output units (nodes)**
    - ♦ **encode possible concepts (or values) to be assigned to the instance under consideration**
      - **example: a class of objects**

General Architecture of Neural Networks (cont.)
— Network Properties —

- **Interconnection scheme**
  - » **feedforward networks**
    - ♦ **all connections point in one direction (input→ output)**
  - » **recurrent networks**
    - ♦ **there are feedback connections or loops**
  - » **symmetrical connections**
    - ♦ **node-i ⟷ node-j,** $W_{ij} = W_{ji}$
  - » **asymmetrical = not symmetrical**



Feedforward connections

Recurrent connections

General Architecture of Neural Networks (cont.)
— Node Properties —

- **The activation level of node can be**
  - » **discrete (e.g., 0 and 1, when the activation function is hard-limiting)**
  - » **continuous across a range (e.g., [0, 1] when a sigmoid function is used for activation function)**
- **Data representation in NNs (for input layer)**
  - » **Discrete feature value**
    - ♦ **can be encoded by a single input unit. In this case, an activation 1 corresponds to “yes” and 0 corresponds “no”.**
  - » **Continuous feature value**
    - ♦ **by continuous representation: each feature is encoded by an input unit, the feature value is mapped into the unit activation (normalised to the interval [0, 1])**
    - ♦ **by discrete representation: the range of continuous values is divided into multiple intervals, each interval encoded by an input unit**
- **Description of training data**
  - » **input vector (feature vector)** **e.g. [0, 1, 0]**
  - » **output vector**

General Architecture of Neural Networks (cont.)
— System Dynamics —

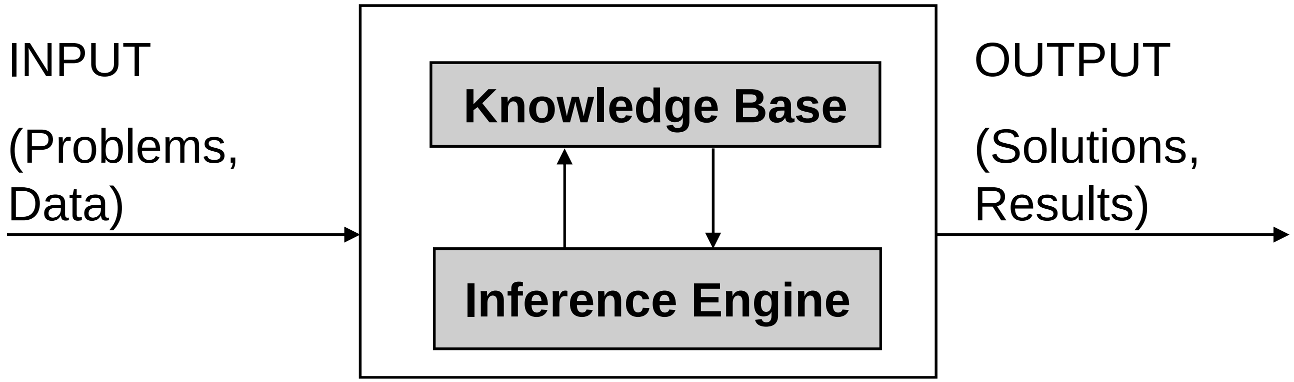
- **The weight initialisation scheme**
  - » **it is specific to the particular neural network model chosen**
  - » **in many cases, initial weights are randomised to small real numbers**
- **The learning rule**
  - » **one of the most important features to specify for a neural network**
  - » **determine how to adapt connection weights in order to optimise the network performance**
  - » **indicate how to calculate the weight adjustment during each training cycle**
  - » **is suspended after training is completed**
- **The activation calculation**
  - » **the important inference behaviour of a NN — how to compute the activation levels across the network**
  - » **training of a NN also involves the same calculation — weight adjustment based on the errors (differences between the actual activation and the desired activation)**

Building a KE System Based on NN Approach

- **General steps**
  - » **Select a suitable neural network model based on the nature of the problem domain**
  - » **Construct a neural network according to the characteristics of the application (with the consideration of available data)**
  - » **Train the neural network with the learning procedure of the selected model**
  - » **Use the trained network for solving problems. If the performance is not satisfactory, then go to one of the previous steps**
- **What will help**
  - » **To select NN model, determine NN architecture, design learning and inference algorithm**
    - ♦ **Familiarity with existing applications and experience**
    - ♦ **Understanding of NN models**
  - » **For better performance**
    - ♦ **Good training data**

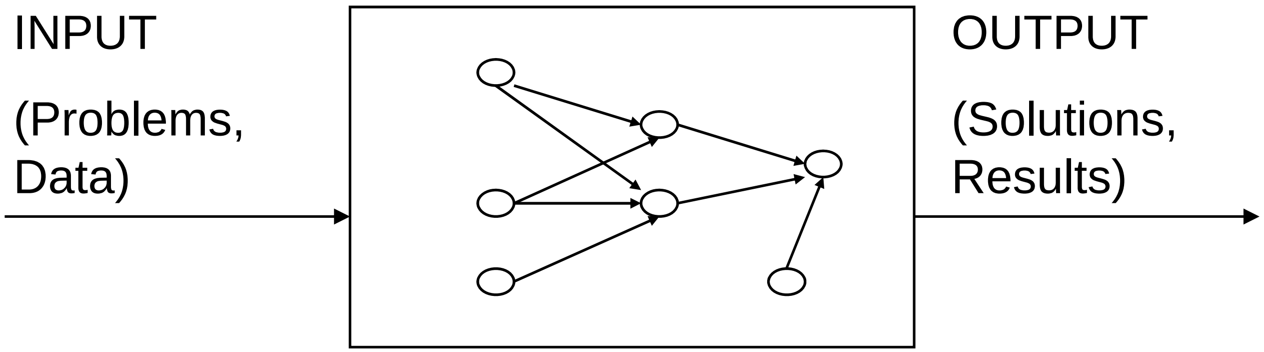
Knowledge-based Information Processing

- Key components of knowledge-based information processing
  - » knowledge representation
  - » knowledge-based inference
- A suitable knowledge representation of a specific problem depends on how the knowledge has been obtained (through learning, acquisition, etc)
- A suitable inference mechanism of a specific problem depends on how the knowledge is represented



Neural Network Systems

- **A neural network architecture after "learning" provides a representation of the knowledge learned, and it also determines the inference mechanism as well**



- ***Although the neural network approach rejects the notion of separating knowledge from the inference mechanism, it does not reject the importance of knowledge in many tasks that require intelligence. It just uses a different way to store and manipulate knowledge.* (L. Fu)**

NN Learning Categories - A Brief Review of Learning

- **Definition of learning**
  - » ***Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more efficiently and more effectively the next time.***
- **Learning can refer to**
  - » **acquiring new knowledge**
    - ♦ **acquisition of concepts**
    - ♦ **understanding of their meanings and relationships to each other and to the domain concerned (symbolic information)**
  - » **skill enhancement**
    - ♦ **distinguish from knowledge acquisition: repetition and practice**
    - ♦ **reinforcing a pattern of neural connections for performing the desired function**
- **AI Learning**
  - » **Rote learning** **(learning by memorization)**
  - » **Learning by instruction** **(learning by being told)**
  - » **Learning by Analogy** **(deductive + inductive)**
  - » **Learning by induction**
    - ♦ **from examples** **(= supervised learning)**
    - ♦ **by observation** **(= unsupervised learning)**
    - ♦ **by discovery** **(= unsupervised learning)**

Neural Network Learning Methods

Supervised

- **assumes the availability of a teacher or supervisor who classifies the training examples into classes**
- **uses pattern-class information to detect pattern misclassifications as a feedback to themselves**
- **error information contributes to the learning process by rewarding accurate classification and/or punishing misclassifications**

Unsupervised

- **uses unlabeled instances, so must identify the pattern-class information as a part of the learning process**
- **less computational complexity and less accuracy than supervised learning**
- **can be used for scientific discovery**
- **clusters instances into groups or decision classes by selecting "winning" neurons and modifying associated weights**

Category of Neural Network Learning
— by Learning Methods —

Learning Methods

Supervised

- Perceptron
- GRNN, PNN
- MLFF with BP
- RBF (partly)
- RCE
- LVQ
- CNN

Reinforcement

- (not covered)

Unsupervised

- ART
- SOM
- RBF (partly)
- VQ

RCE —	Reduced Coulomb Energy
GRNN —	General (Generalized) Regression Neural Network
PNN —	Probabilistic Neural Network
RBF —	Radial Basis Function
SOM —	Self-Organizing Map
ART —	Adaptive Resonance Theory
VQ —	Vector Quantization
LVQ —	Learning Vector Quantization
CNN —	Convolutional Neural Network

Category of Neural Network Learning (cont.)
— by Other Perspective —

- **Learning rules**
  - » **Error correction**
    - ♦ **MLFF with BP**
    - ♦ **GRNN**
    - ♦ **PNN**
    - ♦ **RBF**
    - ♦ **CNN**
  - » **Competitive**
    - ♦ **ART**
    - ♦ **SOM**
  - » **Hebbian (correlative)** (not covered)
    - ♦ **correlative weight adjustment**
    - ♦ **basic form of learning rule**

$$\Delta W_{ji} = O_j O_i$$

Category of Neural Network Learning (cont.)
— by Other Perspective —

- **Procedure of learning**
  - » **Incremental learning**
    - ♦ **a learning system updates its hypotheses (or knowledge) when a new instance arrives without reviewing old instances**
    - ♦ **suitable for a system which continually receives input**
    - ♦ **is efficient since it need not store and reprocess old instances**

Category of Neural Network Learning (cont.)
— by Applications —

- **Categorised in terms of applications:**
  - » **Classification**
    - ♦ **assignment of the input data to one of the finite number of categories**
      - ***Multilayer Perceptron, GRNN, PNN***
  - » **Association**
    - ♦ **auto-association — retrieval of an object (memory) based on part of the object (memory) itself**
    - ♦ **hetero-association — retrieval of an object (memory) in one set using another object (memory) in a different set**
      - **Hopfield Nets, Associative memory**
  - » **Optimization**
    - ♦ **finding the best solution (e.g., minimising a certain cost function)**
      - **Hopfield Nets, Boltzmann Machine**
  - » **Self-organization**
    - ♦ **organizing or clustering received information using adaptive learning capabilities**
      - ***SOM networks, ART and variants***

Examples of Application

- *Classifying cells for cancer diagnosis*

  » **The structure and other characteristics of cells observed from urine samples of patients can provide an accurate indication of bladder cancer.**

  » **A simple two category classification scheme of "well" or "not well" is sufficient for the cell diagnosis task when well chosen, discriminant features are used.**

  » **Several approaches to the classification problem have been proposed**

  - **the tree classifier system accuracy was on the order of 23%**
  - **only the MLFF neural networks approach has achieved levels of classification accuracy acceptable for clinical use (on the order of 96%)**

Examples of Application …

- *Fault Identification in Telephone Switching System*

  - » **When a fault does occur, troubleshooting is carried out in an exhaustive unit elimination process until the faulty unit is eventually isolated.**

  - » **Intermittent faults may occur at intermittent times and for unpredictable duration. The fault may disappear before the troubleshooting process is complete.**

  - » **To assist engineers in identifying such faults, ANNs were trained to help locate and identify a variety of faults using created data.**

Examples of Application ...

- *Forecasting Stock Returns*
  - » **LBS Capital Management, a mutual fund management company based in Florida, manages over $600 million by tracking and predicting the returns on some 3,000 U.S. and foreign-based companies.**
  - » **Some 3,000 neural networks using BP were trained every weekend. Inputs are both fundamental factors (per share income, price to earnings ratio, annual growth rate) and technical factors (closing price, volume, momentum).**
  - » **When the projected return on a security shows up an outlier (i.e. with its percent of return being in a high or low percentile range relative to other securities), the stock is selected for possible acquisition (long) or sale (short). These are the securities that potentially have greater likelihood for either appreciation or depreciation and hence offer more opportunity for higher returns.**

Examples of Application ...

- *Predicting Credit Worthiness for Loan Applications*
  - » **Chase Manhatten Bank developed *Creditview*.**
    - ♦ **Assesses the credit worthiness of public corporations seeking business loans, and performs three-year forecasts that provide a detailed listing of items**
  - » **One NN module generates candidate variables that may indicate the future financial condition of a company, with inputs:**
    - ♦ **historical input data from financial database**
    - ♦ **good and bad obligor data**
    - ♦ **input on industry norms, financial data for the normalisation of specific industry categories**
  - » **Another NN module is used to predict the financial health of the company three years into the future**

Examples of Application ...

- *Handwritten Character recognition*
  - » There have been many successful applications in this area.
  - » In most ANNs used for this purpose, one single pixel point will be treated as one input for the corresponding ANN.
  - » A well trained neural network system may recognise hand written character with an accuracy rate (character level) up to 98%.

- *Automatic Identification of Individuals*
  - » Automatic methods of validating the identity of individuals have broad applications in the area of security and crime prevention.
  - » Neural networks have been applied to identify individuals based on data
    - ♦ Speech
    - ♦ Facial image
    - ♦ Finger-print