Intelligent Systems in Business

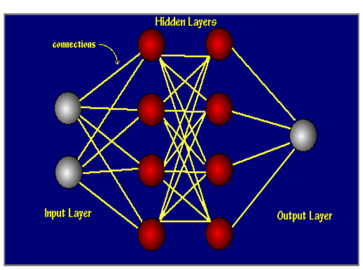
Neural Networks (Part I)

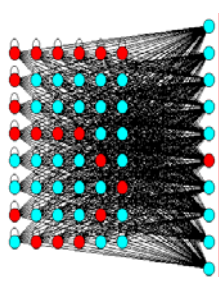
Iulian Nastac

Introduction

- **What is an Artificial Neural Network?**

  A powerful data modeling tool that is able to capture and represent complex input/output relationships





Definition

- ANN is a network of many very simple processors ("units" or "neurons"), each possibly having a (small amount of) local memory.
- The units are connected by unidirectional communication channels ("connections"), which carry numeric data.
- The units operate only on their local data and they receive the inputs via the connections.

A formal definition of an ANN according to Haykin is:

"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:

1. Knowledge is acquired by the network through a learning process.

2. Interneuron connection strengths known as synaptic weights are used to store knowledge."

What are Neural Networks used for?

- Clustering / classification
- Modeling
- Forecasting

Business applications:

- Credit rating and risk assessment
- Insurance risk evaluation
- Fraud detection
- Marketing analysis
- Inventory control
- Prediction of share and commodity prices
- Prediction of economic indicators
- Insider dealing detection
- Future sales
- Production Requirements
- Market Performance
- Economic Indicators

Who needs Neural Networks?

- People who have to work with or analyze data in any form.
- People who in business, finance, or industry and whose problems are either complex , laborious, 'fuzzy' or simply un-resolvable with present methods.
- People who simply want to improve on their current techniques and gain competitive advantage.

7

Why are they the best method for data analysis?

- Neural networks outperform current methods of analysis because they can successfully
- Deal with the non-linearities of the world
- Be developed from data without an initial system model
- Handle noisy or irregular data from the real world
- Quickly provide answers to complex issues
- Be easily and quickly updated
- Interpret information from tens or even hundreds of variables or parameters
- Readily provide generalized solutions

History of the Development of Neural Networks

- Artificial Neural Networks (ANN) have been hailed as the greatest technological advance since the transistor

- The ANN design is based on the neural structure of the brain

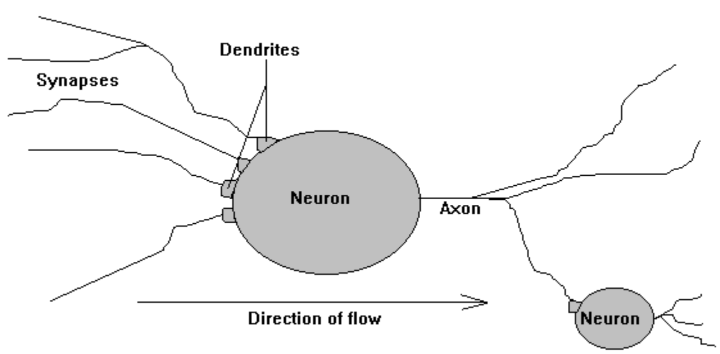
- ANN are computer (software & hardware) simulations of the biological neural networks

History (cont.)

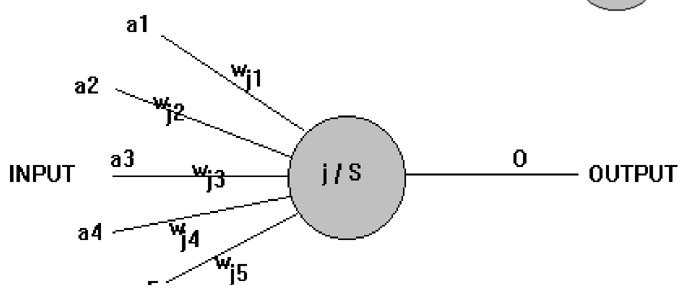
- 1940 - McCulloch and Pitts introduced the first neural network computing model
- 1949 - Donald Hebb: *Organization of Behavior*
  - The concept of neurons and how they work
- 1950 Rosenblatt's developed a two-layer network, the perceptron
  - The perceptron was capable of learning certain classifications by adjusting connection weights
- 1969 Minsky and Papert published *Perceptrons*
  - Although the perceptron was successful in classifying certain patterns, it had a number of limitations.
  - The perceptron is not able to solve the classic XOR (exclusive or) problem.
  - Such limitations led to the decline of the field of neural networks.
- In the early 1980's, researchers showed renewed interest in neural networks
  - Boltzmann machines, Hopfield nets, competitive learning models, multilayer networks, Kohonen (SOM) nets, adaptive resonance theory model, etc.

Artificial neuron vs. biological neuron

Biological Neuron



Artificial Neuron



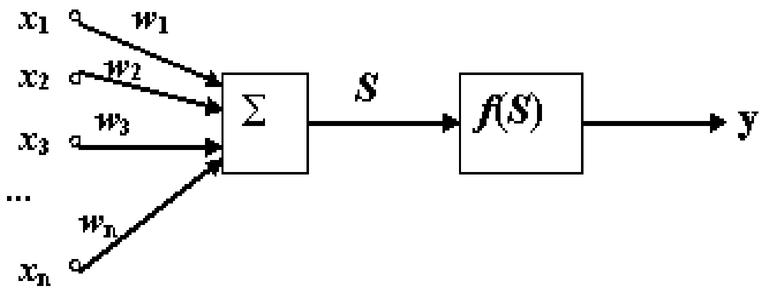
Biological Neural Networks	Artificial Neural Networks
• biological brain is an incredibly complex system of more than a 100 billion neurons of different types	• groups/layers of usually no larger than hundreds neurons
• neurons are highly interconnected (not all) with each other via more than a 150 billion synapses	• the number of interconnections is usually between n and n^2, where n is the number of neurons
• action potentials are fired from each neuron to others (depending on the task the brain is performing) which are electric pulses whose intensity level varies	• most ANNs use analogue values usually represented by floating point numbers in software simulations • in hardware implementations, ANNs are the fastest possible tools

12

Some of the more prominent ANN paradigms

1) PERCEPTRON, (1957-Rosenblatt)
2) ADALINE and MADALINE, (1959- 62-Widrow and Hoff)
3) AVALANCHE, (1967-Grossberg)
4) CEREBELLATION, (1969-Marr, Albus & Pellionez)
5) BACKPROPAGATION (BPN), (1974-85-Werbos, Parker, Rumelhart), more commonly referred to as MULTI-LAYER PERCEPTRON (MLP)
6) BRAIN STATE IN A BOX, (1977-Anderson)
7) NEOCOGNITRON, (1978-84-Fukushima)
8) ADAPTIVE RESONANCE THEORY (ART), (1976-86-Carpenter, Grossberg)
9) SELF-ORGANISING MAP, (1982-Kohonen)
10) HOPFIELD, (1982-Hopfield)
11) BI-DIRECTIONAL ASSOCIATIVE MEMORY, (1985-Kosko)
12) BOLTZMANN/CAUCHY MACHINE, (1985-86-Hinton, Sejnowsky, Szu)
13) COUNTERPROPAGATION, (1986-Hecht-Nielsen)
14) RADIAL BASIS FUNCTION, (1988-Broomhead, Lowe)
15) PROBABILISTIC (PNN), (1988-Specht)
16) GENERAL REGRESSION NEURAL NETWORK (GRNN), (1991-Specht)

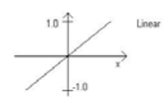
The Basic Model of the Neuron



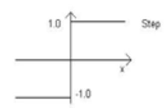
$$S = \sum_{i=1}^{n} x_i w_i \qquad\qquad y = f(S) = f\left(\sum_{i=1}^{n} w_i x_i\right)$$

Activation Functions

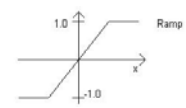
1) Linear $\quad f(x) = ax$



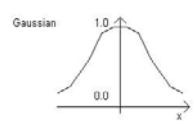
2) Step (Hard Limiter) $\quad f(x) = \begin{cases} a & \text{if } x \le t \\ b & \text{if } x > t \end{cases}$



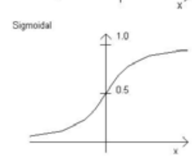
3) Ramp $\quad f(x) = \begin{cases} a & \text{if } x \ge c \\ x & \text{if } |x| < c \\ -a & \text{if } x \le -c \end{cases}$



4) Gaussian $\quad f(x) = exp(-x^2/v)$

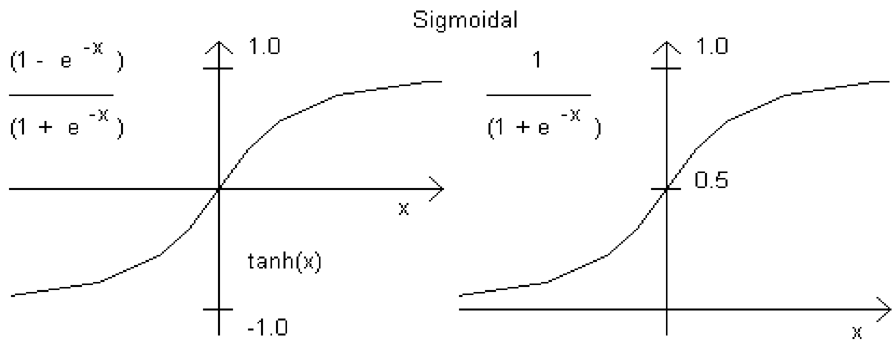


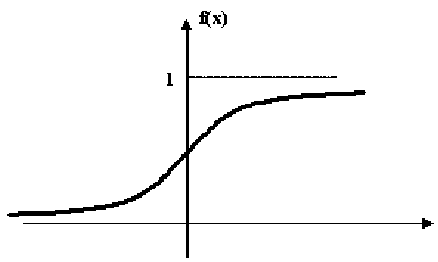
5) Sigmoidal $\quad f(x) = 1/(1 + exp(-x))$



15

Sigmoidal Activation Functions

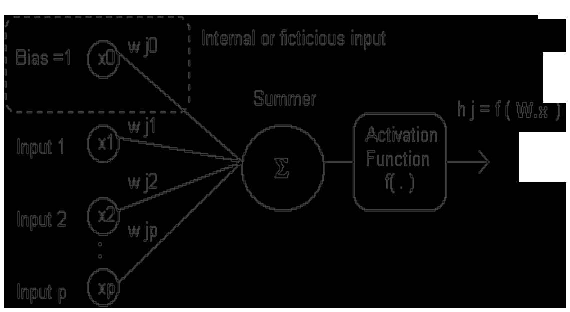




$$f(x) = \frac{1}{1 + e^{-k \cdot x}}$$

16

Basic Learning Model of neuron or Processing Element (PE)



$$h_j = f(\mathbf{W}(k).\chi(k)) \qquad (1)$$

$$\mathbf{W}(k) = [w_{j0}, w_{j1}, \ldots, w_{ji}, \ldots, w_{jp}]$$

$$\chi(\mathbf{k}) = [x_0, x_1, \ldots, x_i, \ldots, x_p]^T$$

$e_j = d_j - h_j$ output error

d_j desired PE output

$$w_{ji}(k+1) = w_{ji}(k) + \Delta w_{ji}(k) = w_{ji}(k) + \mu E(h_j(k), d_j(k)) \qquad (2)$$

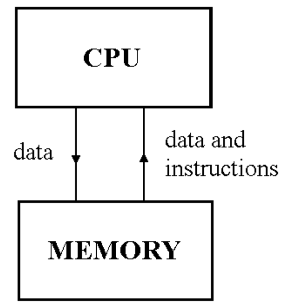
μ the learning rate

$E(h_j(k), d_j(k))$ error function

17

The von Neumann machine and the symbolic paradigm

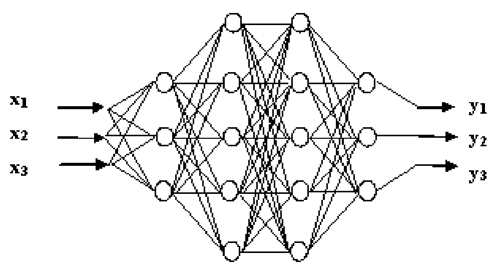
The computer repeatedly performs the following cycle of events:



1) fetch an instruction from memory.
2) fetch any data required by the instruction from memory.
3) execute the instruction (process the data).
4) store results in memory.
5) go back to step 1).

The advantages of neural networks machine over von Neumann machine

- **computing speed**
- **parallelism**
- **reliability**
- **pre-processing data**
- **programming**



Neural network design involves at least five main tasks:

1) Data collection

2) Raw data preprocessing

3) Feature extraction from preprocessed data

4) Selection of an ANN type and topology (architecture)

5) ANN training, testing and validation

Basic Concepts

The basic attributes of an ANN can be separated into:

- **architectural properties**
- **functional properties**

Architectural issues

1) Topology
2) Number of layers in the network
3) Number of neurons per layer
4) Type and parameters of the neuron which is usually the same throughout
5) Processing and pattern recall speed
6) Number of calculations per iteration during learning and recall
7) Implementation technology

Functional properties refer to the network's operation

1) Learning method and efficiency
2) Learning recall
3) Information association
4) Continuous comparison of new information with existing knowledge
5) Classification of new information
6) Development of new classifications if applicable
7) Noise immunity
8) Reliability
9) Error and failure tolerance
10) Performance

What are the main types of Neural Network learning?

- **supervised**
- **unsupervised**

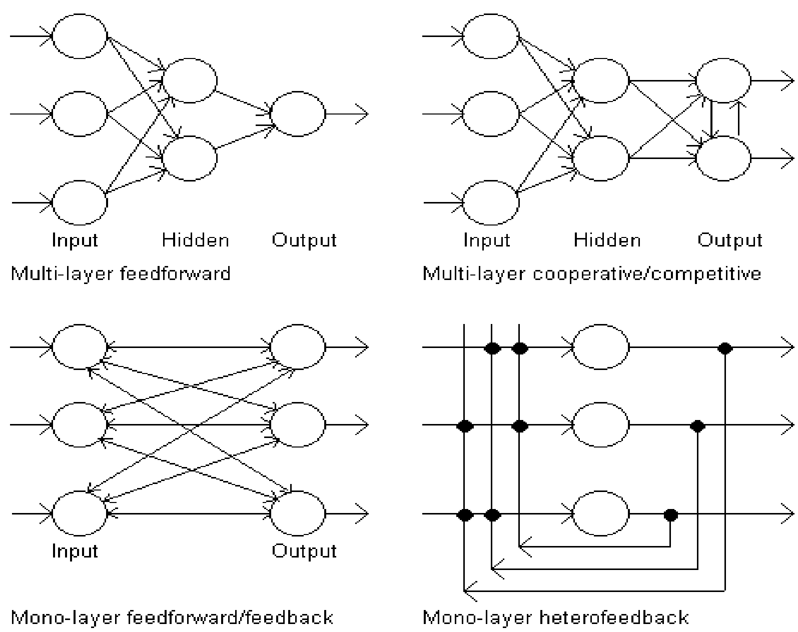
Supervised Learning

- Process of training by giving examples of the task (i.e., learning with a teacher)

- Training set is composed by pairs of vectors (patterns):
  - the first pattern of each pair is an example of an input pattern that the network might have to process
  - the second pattern is the output pattern that the network should produce for that input which is known as a target output pattern for whatever input pattern.

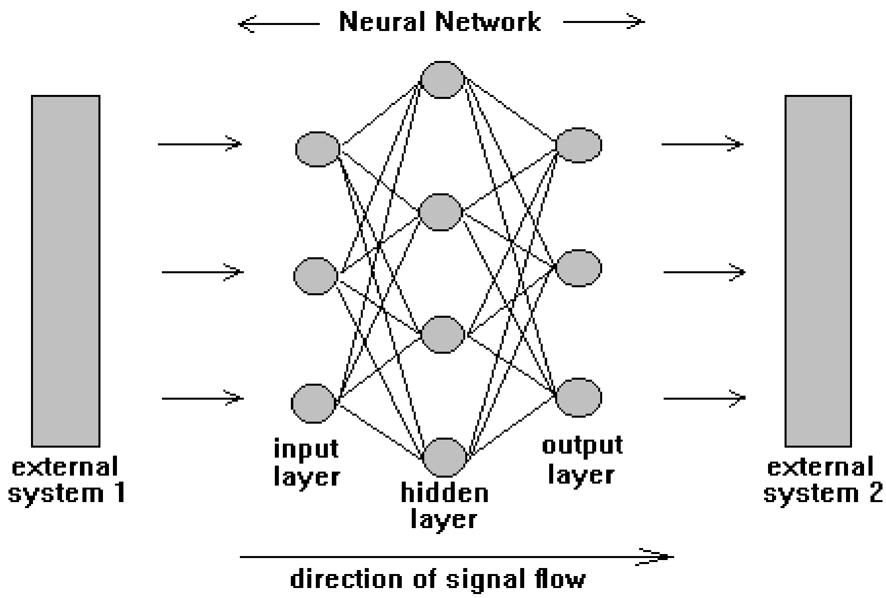
This technique is mostly applied to feed forward type of neural networks.

Common ANN Topologies



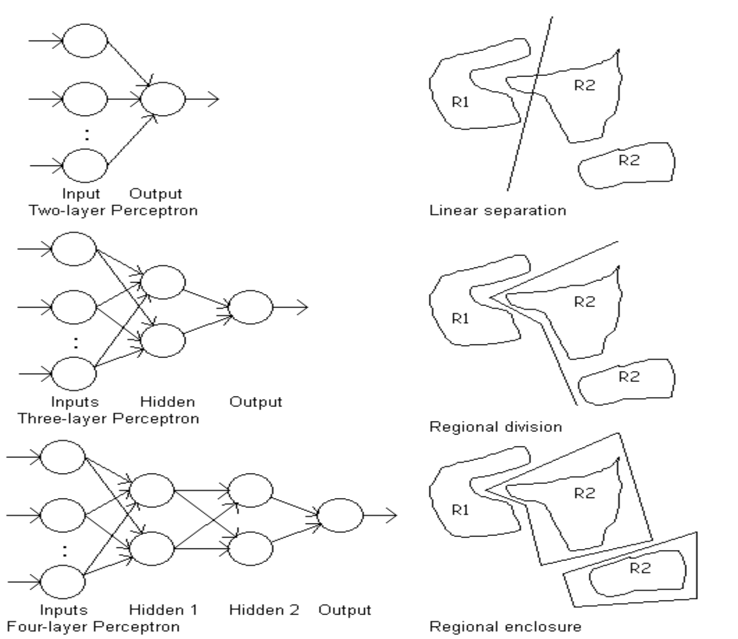
26

Feedforward Neural Networks



27

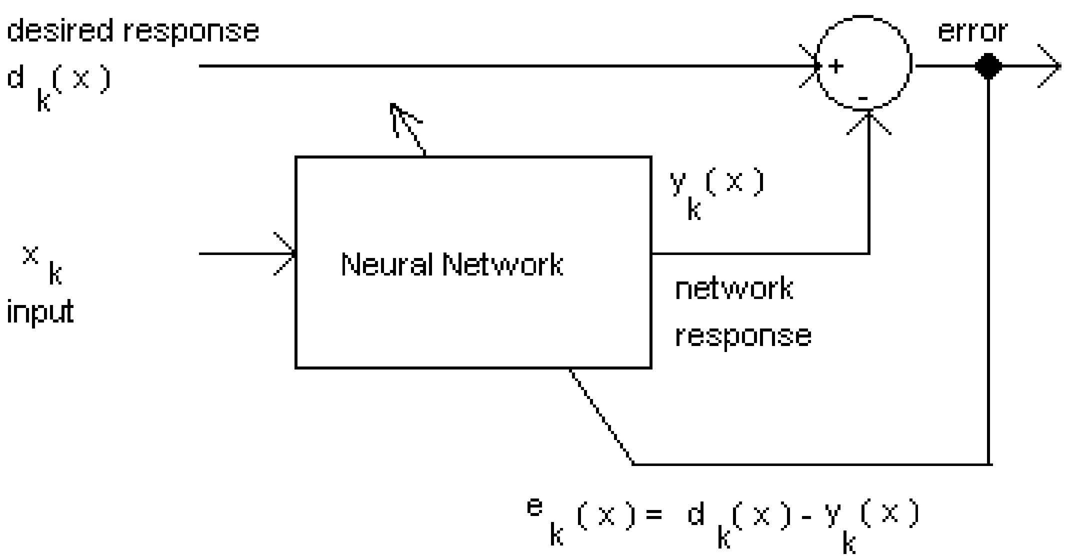
General Multi-layer Perceptron Neural Network



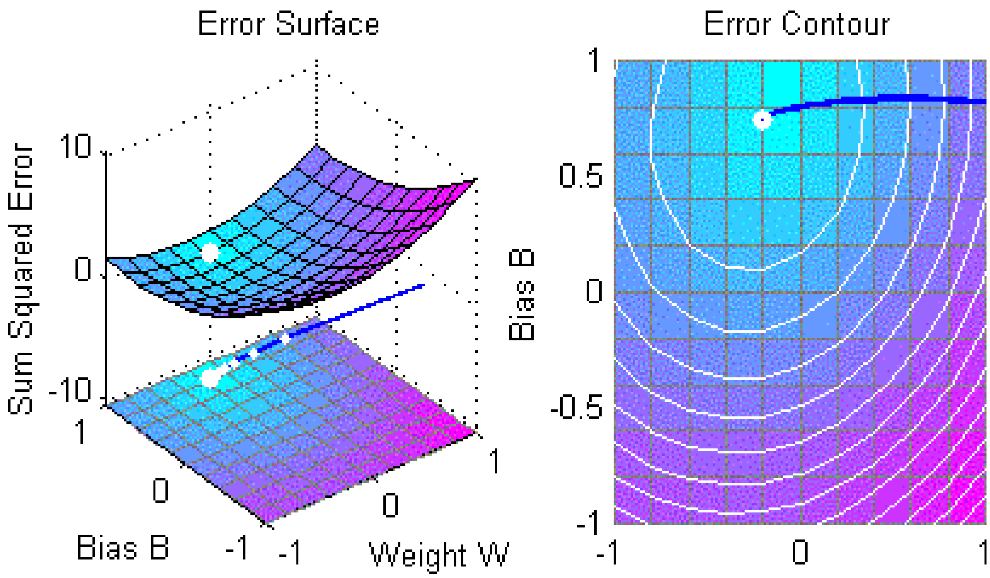
28

Artificial Neural Network as a Black Box

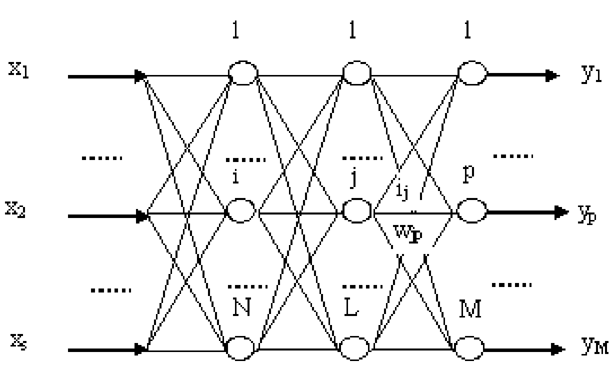
A Supervised Learning Scheme



Minimum of cost function



A Basic Supervised Learning Algorithm (Backpropagation - BKP)



$$e_p = y_p^* - y_p$$

$$E = \frac{1}{2}\sum_{p=1}^{M} e_p^2 = \frac{1}{2}\sum_{p=1}^{M}(y_p^* - y_p)^2$$

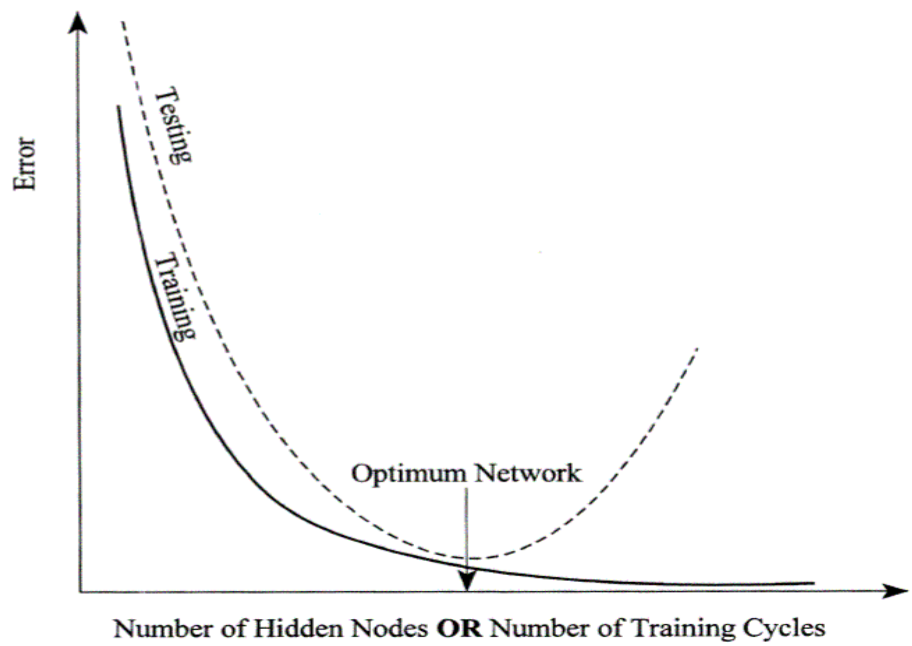
$$w^{k+1} = w^k + \Delta w^k$$

$$\Delta w = -a\frac{\partial E}{\partial w}$$

$$\frac{\partial E}{\partial w_{jp}} = -(y_p^* - y_p)\frac{\partial f_p}{\partial S_p}\frac{\partial S_p}{\partial w_{jp}}$$

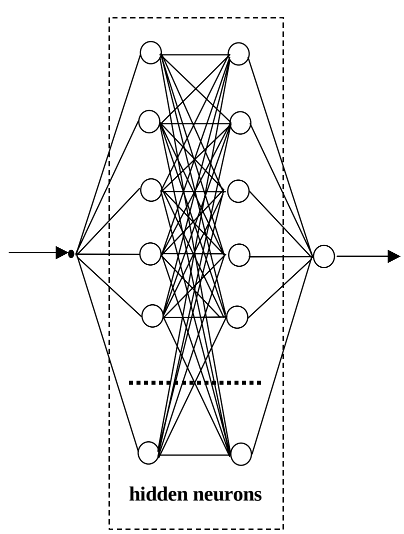
$$w_{jp}^{k+1} = w_{jp} + a(y_p^* - y_p)y_p(1 - y_p)i_j$$

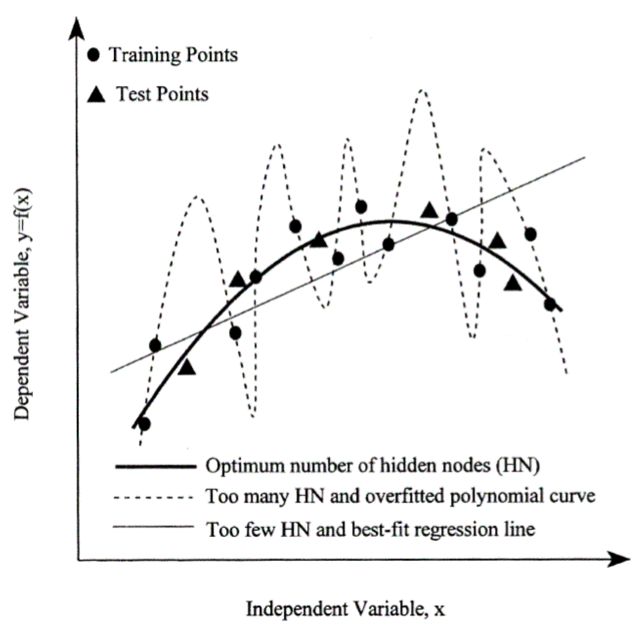
Learning error versus test error



33

Effect of hidden layer size on network generalization





References

- Basheer, I.A., Hajmeer, M., Artificial neural networks: fundamentals, computing, design, and application, Journal of Microbiological Methods, Elsevier Science, Vol. 43, 2000, pp. 3-31.
- Berthold, M. and Hand, D.J. Intelligent Data Analysis, Springer, 1999
- Dhar, V. and Stein, R. Intelligent Decision Support Systems, Prentice Hall, 1997
- Nastac, I., Koskivaara, E., A Neural Network Model for Prediction: Architecture and Trainig Analysis, TUCS Technical Report 521, Turku Centre for Computer Science, 2003