Introduction to Artificial Neural Networks

By: Mojtaba Ziyadi

Table of Contents

ILLINOIS CENTER FOR TRANSPORTATION

Clarifying Some Definitions

Artificial Intelligence
" The study and design of intelligent agent (a system that perceives its environment and takes action that maximize its chance of success"
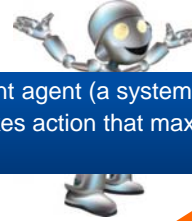
Machine Learning
Takes input empirical data → patter........ons

Supervised Learning

Data Mining
Focuses on discovery of unknown properties......

Unsupervised Learning

ILLINOIS CENTER FOR TRANSPORTATION

Machine Learning

- ❑ **Artificial Neural Networks (ANN)**
- ❑ **Support Vector Machines (SVM)**
- ❑ **Genetic Algorithms (GA)**
- ❑ **Bayesian Networks**
- ❑ **etc.**

ILLINOIS CENTER FOR TRANSPORTATION

Human Brain vs. Computers

❑ 10^9 ---- Each neuron connected to 10^5 neurons

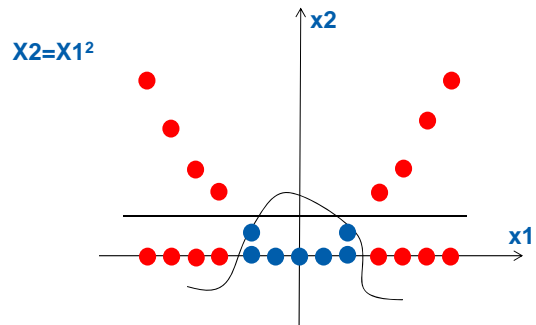|	Brain	Computer
processing elements	10^9 Neurons	10^8 Transistors
style of computation	Parallel - Distributed	Serials - Centralized
processing speed	100 Hz	10^9 Hz
energy use	30 W	30 W (CPU)
Fault Tolerant	yes	no

ILLINOIS CENTER FOR TRANSPORTATION
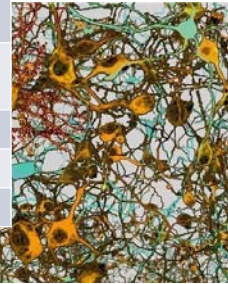
What Happens in the Brain?

❑ **A Neuron has;**

Dendrites (inputs)

Cell Body

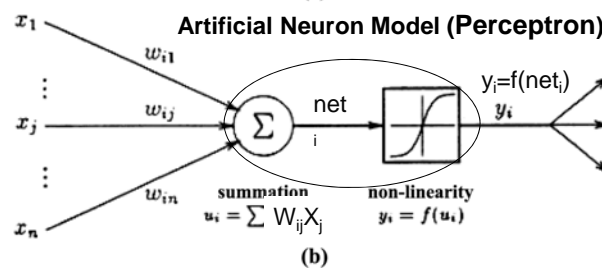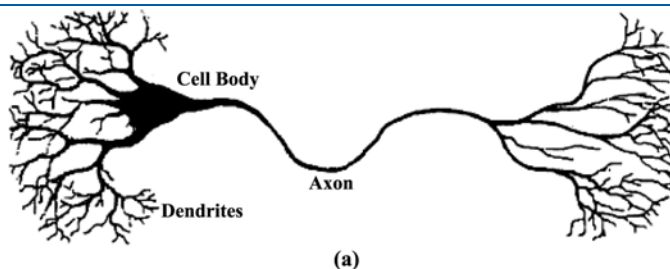**Axon (output)**

ILLINOIS CENTER FOR TRANSPORTATION
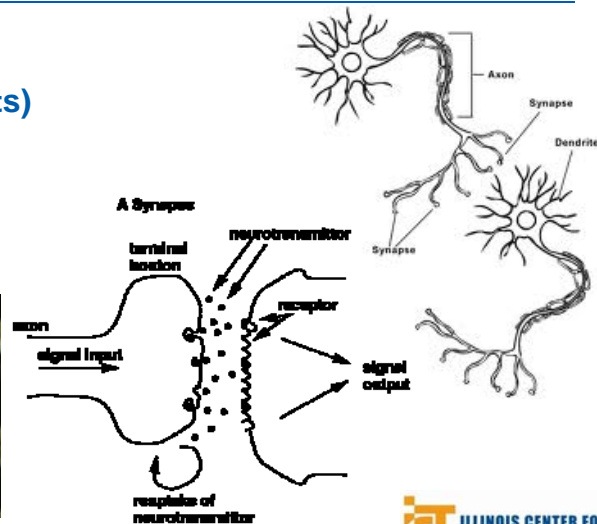
What is Artificial Neural Network?

- ❏ **Mathematical model of biological neural networks**
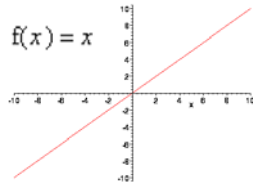- ❏ **Non linear statistical data modeling tool**

ILLINOIS CENTER FOR TRANSPORTATION

Architecture of Neural Networks

Cell Body

Axon

Dendrites

(a)

Artificial Neuron Model (Perceptron)

$y_i = f(net_i)$

summation $u_i = \sum W_{ij}X_j$ non-linearity $y_i = f(u_i)$

(b)

ILLINOIS CENTER FOR TRANSPORTATION

Architecture of Neural Networks

Some activation functions;

Identity

$$f(x) = x$$

$$f(x) = 0 \quad \text{if } x \leq 0 \quad \text{and} \quad f(x) = 1 \quad \text{if } 0 < x$$
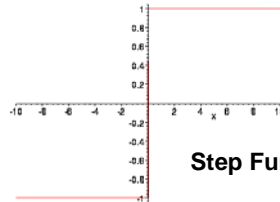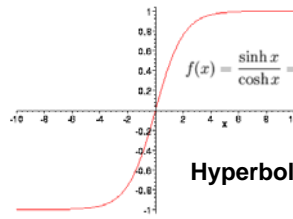
Step Function

$$f(x) = \frac{1}{1 + e^{(-a\,x)}}$$
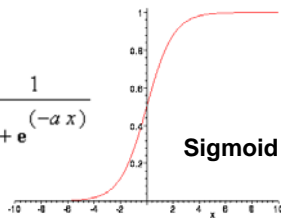
Sigmoid

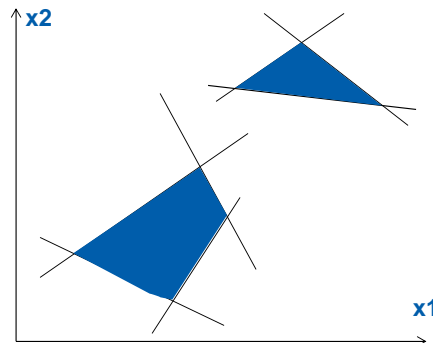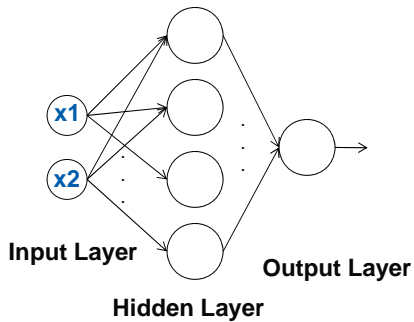$$f(x) = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1}$$

Hyperbolic Tangent

ILLINOIS CENTER FOR TRANSPORTATION

Architecture of Neural Networks

Multilayer Perceptron

x1

x2

Input Layer

Output Layer

Hidden Layer

x2

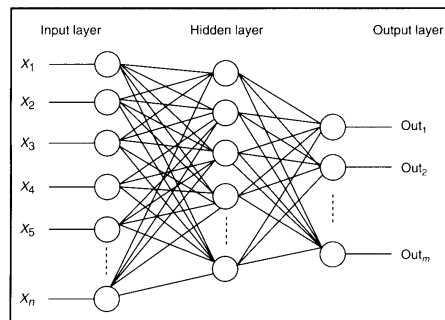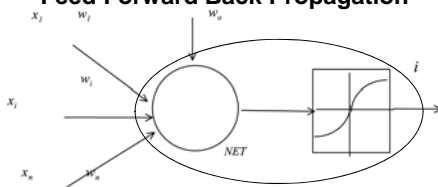x1

ILLINOIS CENTER FOR TRANSPORTATION

The learning Process

- ❑ **We have a set of data samples**
- ❑ **Each sample is an input (variables) – output(responses) pair**
- ❑ **In training the goal is finding the best Weights (W_{ij})**

x   y

Input Layer

Output Layer

Hidden Layer

ILLINOIS CENTER FOR TRANSPORTATION
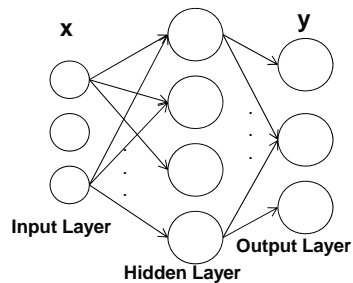
Training Algorithm

Feed Forward Back Propagation

Input layer   Hidden layer   Output layer

x_1
x_2
x_3 Out$_1$
x_4 Out$_2$
x_5
x_n Out$_m$

Forward pass:

Hidden layer:

$$net_j = w_0 + \sum_i w_{ij} x_i$$

$$H_j = f(net_j)$$

Output layer:

$$net_j = v_0 + \sum_j v_{jk} H_j$$

$$Y_k = f(net_j)$$

Error Function:

$$E = \sum_p E_p = \sum_p \frac{1}{2}(T_p - Y_p)^2$$

Back propagation:

Gradient Descen Algorithm

$$\Delta V_j = -\eta \frac{\partial E}{\partial V_j} + \alpha V_{old}$$

$$V_{new} = V_{old} + \Delta V_j$$

where η is the learning rate, $\eta \in (0,1]$, and
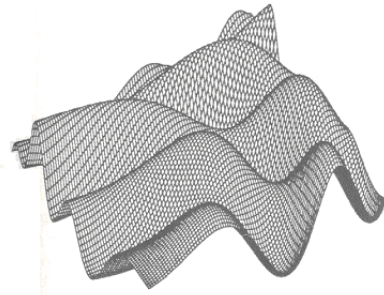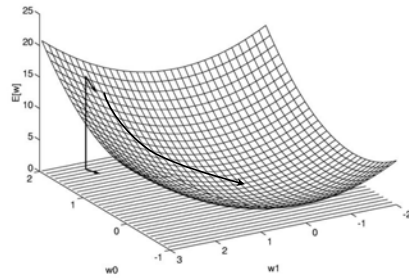α is the momentum coefficient, $\alpha \in [0,1]$.

$$\Delta W_i = -\eta \frac{\partial E}{\partial W_i} + \alpha W_{old}$$

$$W_{new} = W_{old} + \Delta W_i$$
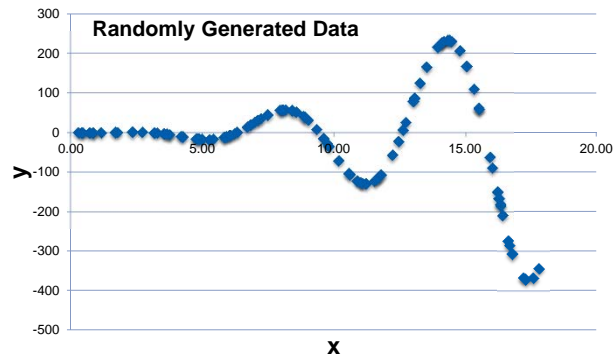
Training Algorithm

Other Training Algorithms

- ❑ **Levenberg-Marquardt**

- ❑ **Conjugated Gradient**
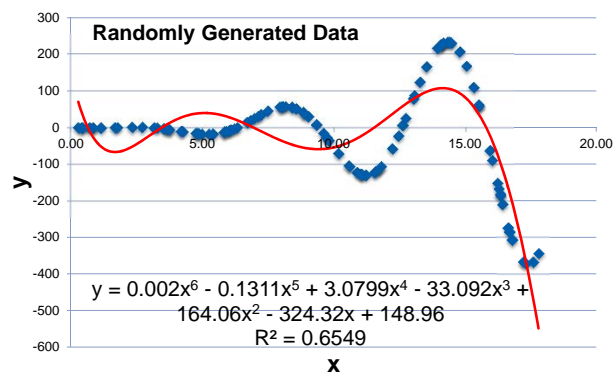
- ❑ **Gauss-Newton**

- ❑ **Steepest Gradient**

- ❑ **…**

An Example

$y = x^2 \sin(x) \log(x)$

Randomly Generated Data

ILLINOIS CENTER FOR
TRANSPORTATION

An Example

$y = x^2 \sin(x) \log(x)$

Randomly Generated Data

$y = 0.002x^6 - 0.1311x^5 + 3.0799x^4 - 33.092x^3 + 164.06x^2 - 324.32x + 148.96$
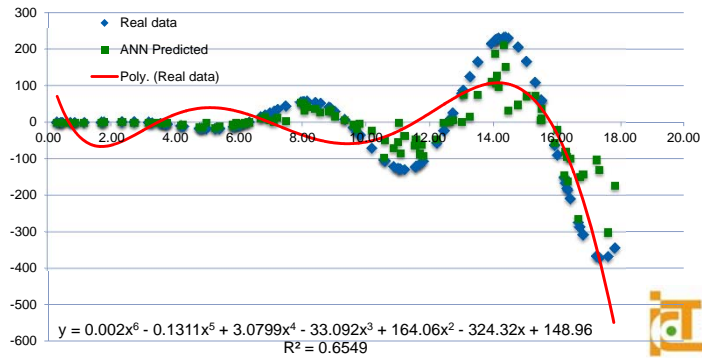$R^2 = 0.6549$

ILLINOIS CENTER FOR
TRANSPORTATION

An Example … Training
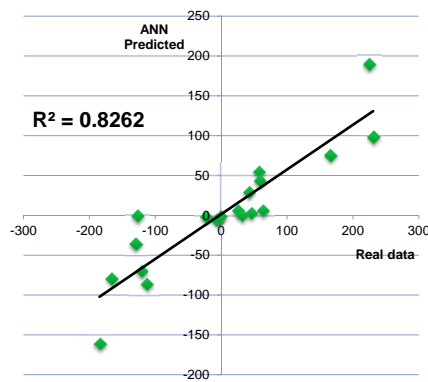
$y = x^2 \sin(x) \log(x)$

80% Training
20% Testing

Neural Network Structure;
- MLP
- 1 hidden layer with 5 neurons
- Tanh activation function
- Training terminated at 600 iteration
- GDM training algorithm

$y = 0.002x^6 - 0.1311x^5 + 3.0799x^4 - 33.092x^3 + 164.06x^2 - 324.32x + 148.96$
$R^2 = 0.6549$

An Example … Test Results

Real Data vs. ANN Predicted

$R^2 = 0.8262$

Real Data vs. Polynomial Reg. Predicted

$R^2 = 0.1151$