Artificial Intelligence

Artificial Neural Networks

Invited lecture by

Sami Ghawi

MSc. Informatics, Technical University of Munich
sami.ghawi@tum.de

Watch this lecture and download the slides from

http://jarrar-courses.blogspot.com/2011/11/artificial-intelligence-fall-2011.html

Table of Contents

Keywords: Learning, Machine learning, neural networks, supervised learning, unsupervised learning, text recognition

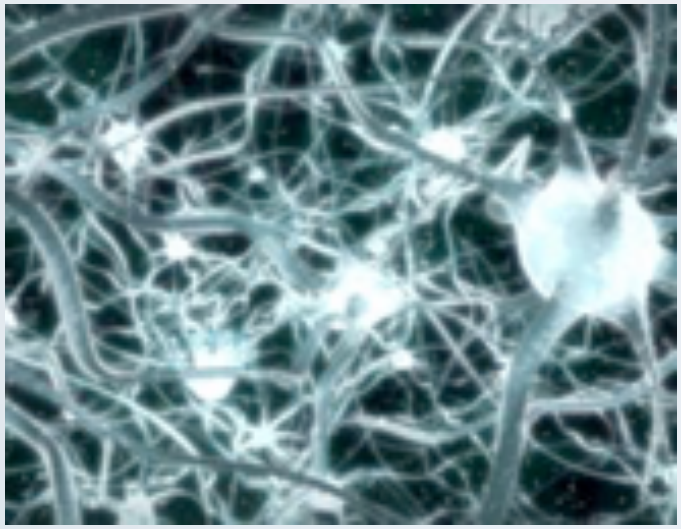
التعلم الآلي، تعليم الآلة، التعليم الموجه ، التعليم غير الموجه، الشبكات العصبونية، التعرف على الكتابات

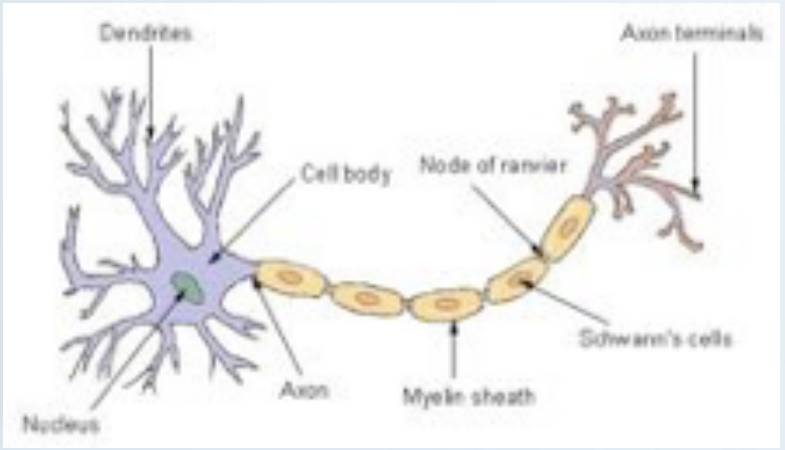
Introduction

Biological Neural Networks

A series of interconnected neurons whose activation defines a recognizable linear pathway.

A neuron cell (ةيبصع ةيلخ) receives the incoming signals through dendrites (تانصغت). It processes the summed inputs and forwards the output signal to other cells through the Axon (رواحم). The axon is connected through synapses (يبصعلا كابتشالا طاقن) to other cells dendrites. [1]

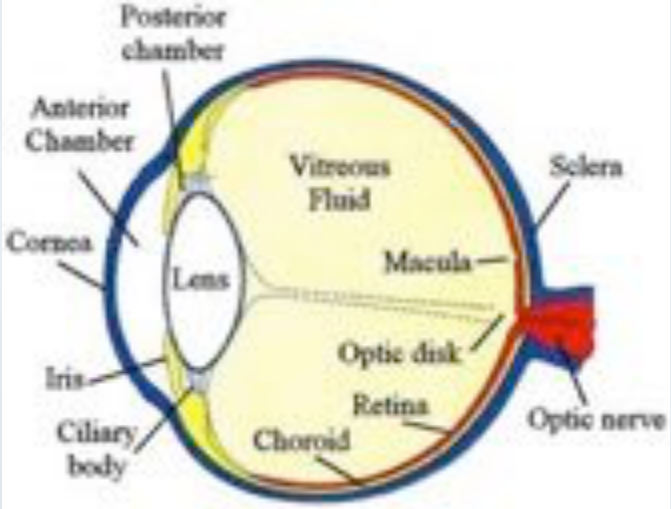




Introduction

Human Eye [2]

- Retina (شبكية) measures 5x5 cm and contains 10^8 sampling elements of rods (نبوتية) and cones (مخروطية).
- The eye gives a data rate of 2.75 GBytes/s.



Introduction

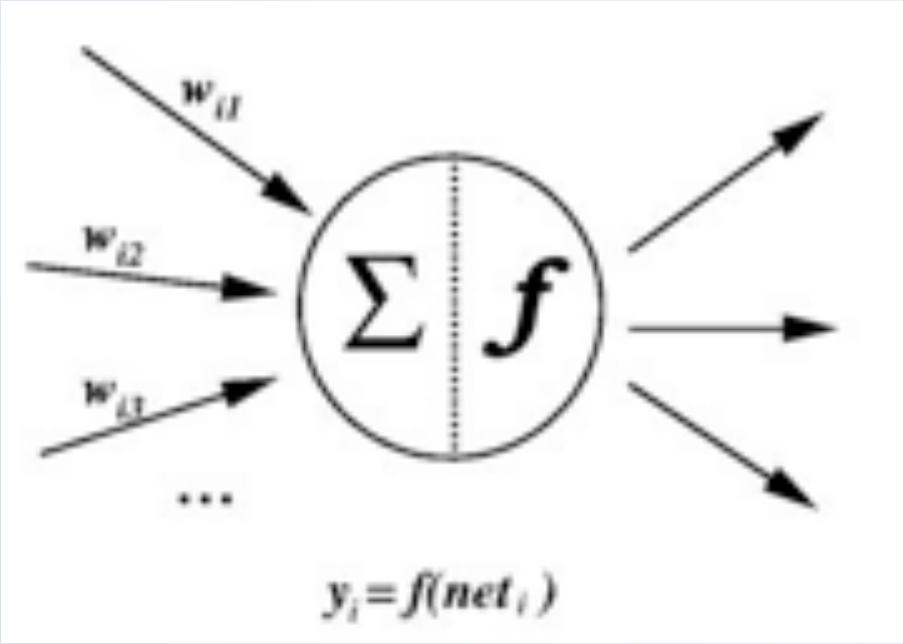
IT and Biology [2]

- We shouldn't try to copy biology.
- The human nervous system contains 85 billion neurons that are heavily interconnected.
- Computers cannot perform the computations similar to the human brain.
- We do not understand how the brain functions.
- However, we can use the principles in biology instead of the exact implementation.

Introduction

Mapping Biology to Machine Learning

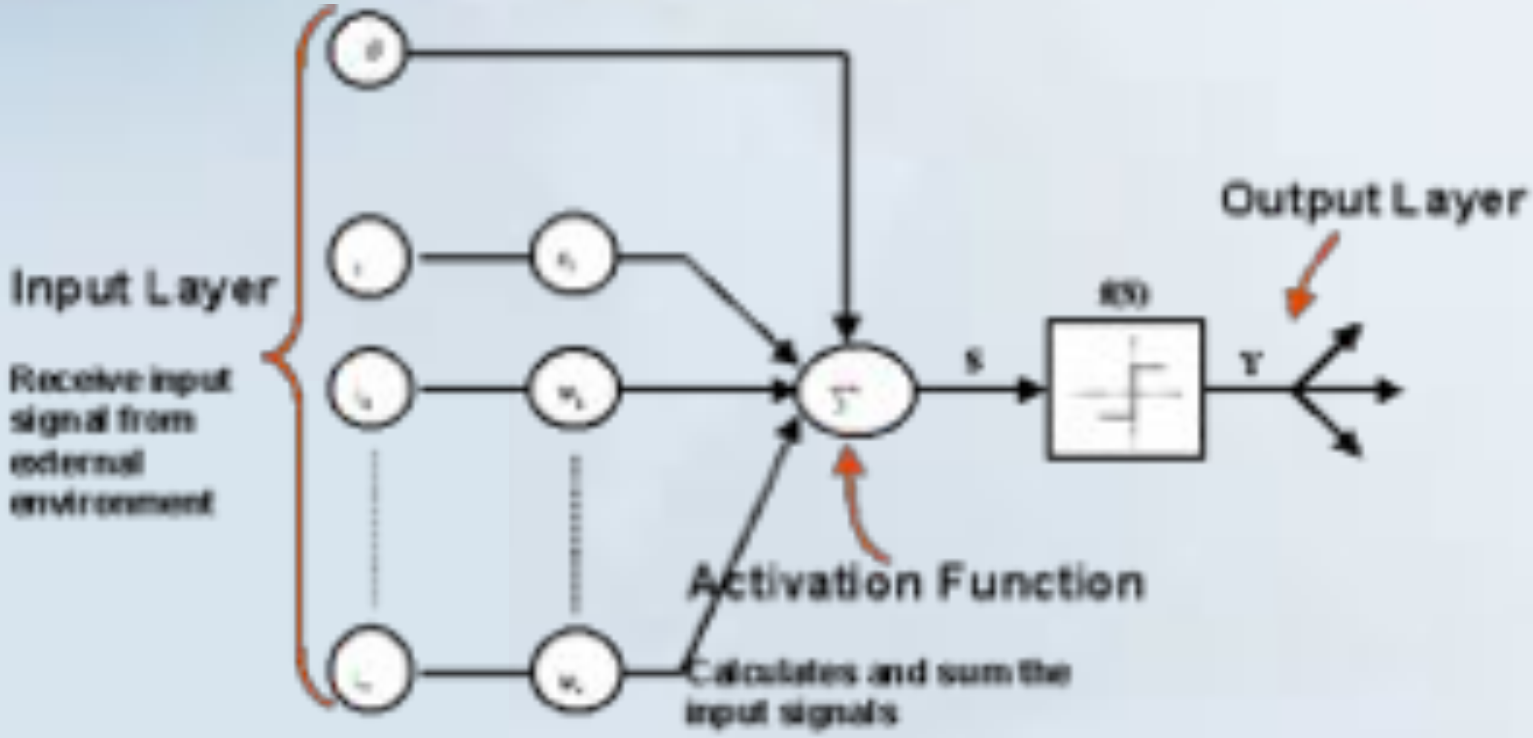
- Every neuron sums the input and calculates the output using a transfer function.
- The synapses are represented by real numbers called weights.
- The dendrites are represented by connections between neurons.
- The neural network is represented by layers to simplify the structure.



History

McCulloch-Pitts Neuron (1943)

– Neurophysiologist Warren McCulloch and mathematician Walter Pitts wrote a paper on how neurons might work. (A logical calculus of the ideas immanent in nervous system)



History

Hebb Learning Rule (1949)

- Psychologist Donald Hebb wrote a book about how humans can learn. [3]
- His work stated that neural pathways are strengthened each time they are used.
  - If two nerves fire at the same time, the connection between them is enhanced.
- The change of weight between a neuron-i and neuron-j is calculated as:

$$\Delta w_{ij} = \alpha a_i a_j$$

where alpha is a learning constant.

History

Rosenblatt's Perceptron (1957) [4,5]

- A Supervised linear classifier invented by Frank Rosenblatt.
- Mathematical definition:

$$f(x) = \begin{cases} 1 & \text{if } w \cdot x + b > 0 \\ 0 & \text{otherwise} \end{cases}$$

- Algorithm:

- Initialize the weights to either 0 or small random value.
- Loop over training set (x, d).
  - Calculate the output f(x):

$$y_j(t) = f[\mathbf{w}(t) \cdot \mathbf{x}_j] = f[w_0(t) + w_1(t)x_{j,1} + w_2(t)x_{j,2} + \cdots + w_n(t)x_{j,n}]$$

  - Calculate the weights:

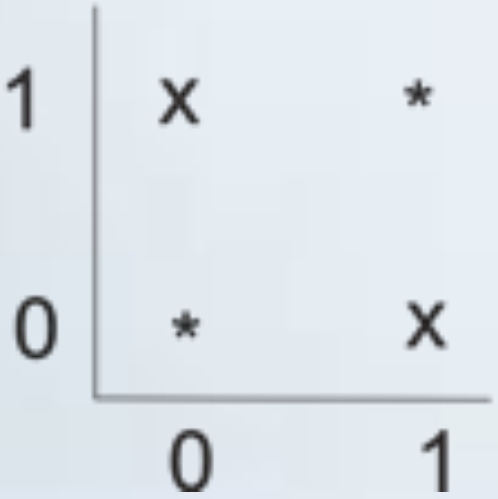
$$w_i(t+1) = w_i(t) + \alpha(d_j - y_j(t))x_{j,i}$$

  - Break loop if the number of iterations have been reached or the error is less than: $\frac{1}{s}\sum_{j}^{s}[d_j - y_j(t)]$
- End Loop

History

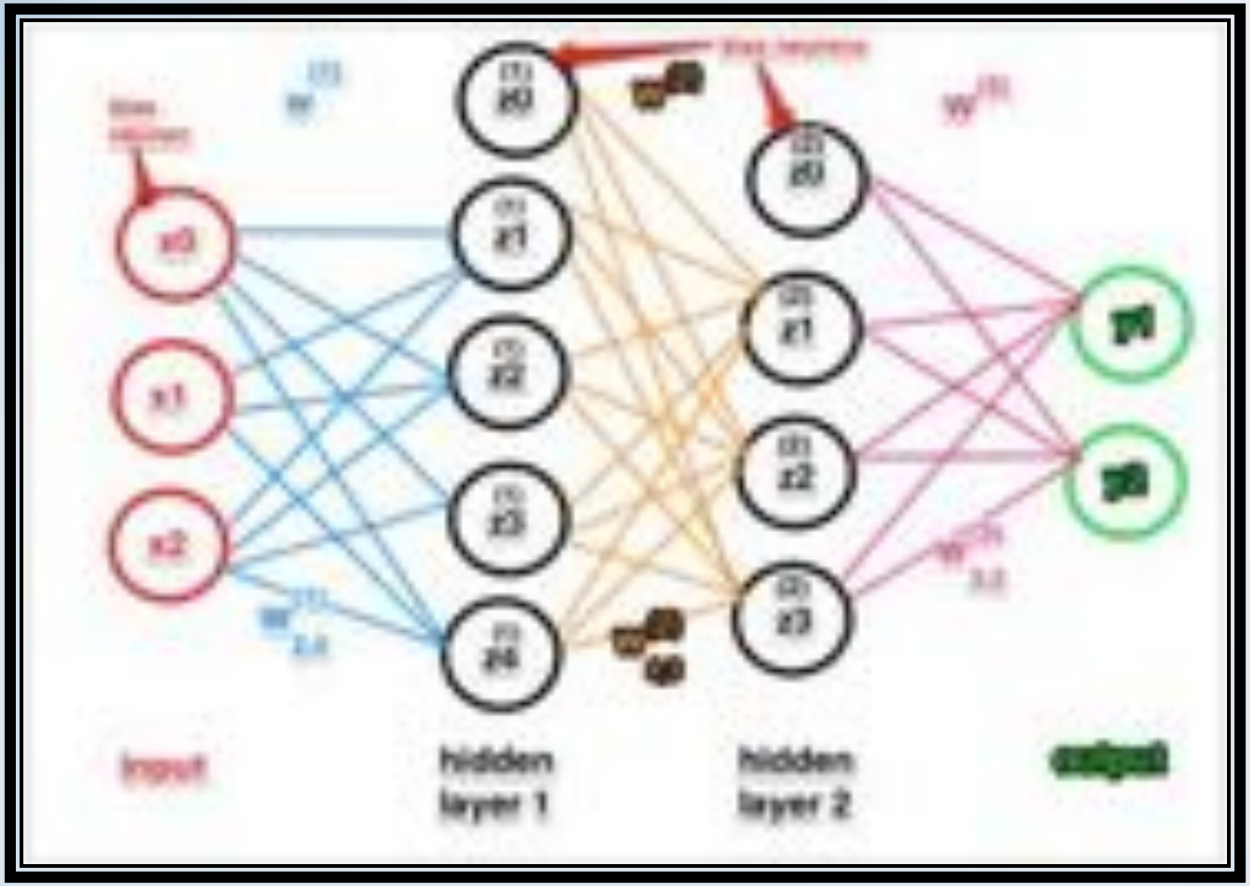
Minsky & Papert (1969) [6]

- They published a paper enumerating many of the flaws in neural networks.
- They used the functionality of XOR as an example of how the neural networks fail.
- If a model cannot solve a very simple problem, then we cannot use it in the real world.
- They suggested to use multi-layer neural networks.



Feed-forward Neural Network

- This neural network consists of multiple layers of nodes in a directed graph [7,8,9].
- Each layer is fully connected to the next layer (but not with bias).



Feed-forward Neural Network

- Only one input layer per network. The number of nodes depends on the data.
- Only one output layer per network.
  - Regression mode: one node only per output (i.e. price gets one node).
  - Classification mode:
    - One node per output. (i.e. one node for spam filter. 0 for spam and 1 for non-spam)
    - One node per label (i.e. two nodes for spam filter).
- Number of hidden layers is a variable. (refer to [9])
  - Number of units per hidden layer is also a variable. (refer to [9])
- The neural networks have two phases:
  - Training Phase
  - Execution Phase

Execution Phase

- The execution phase returns the output of a given input.
- The output of every neuron is computed as follows (one hidden layer):
  - Calculate the activations of the first layer

$$a_j = \sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}$$

  - Transform using an activation function

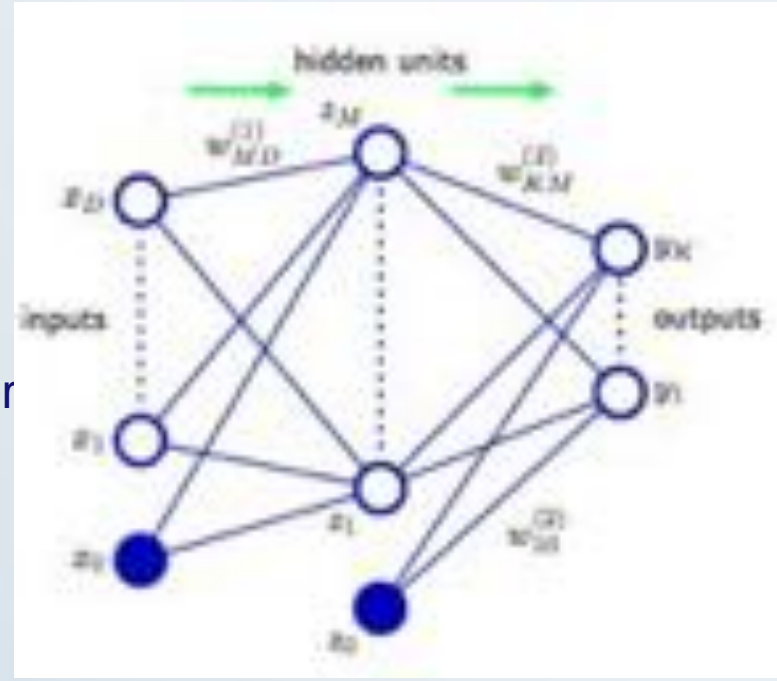
$$z_j = h(a_j)$$

  - Find the activations of the second layer

$$a_k = \sum_{j=1}^{M} w_{kj}^{(2)} z_j + w_{k0}^{(2)}$$

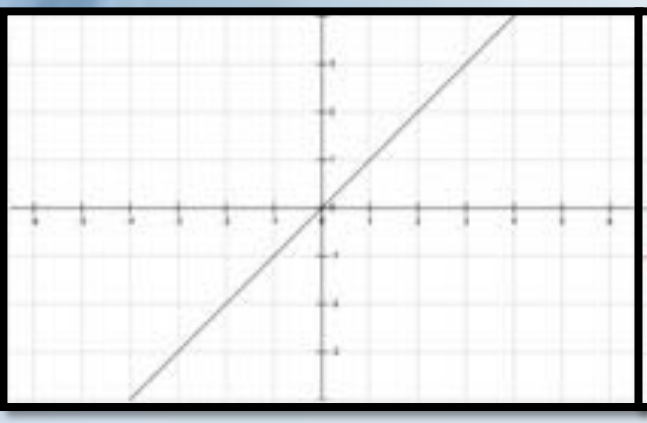
  - Transform using an activation function

$$y_k = \sigma(a_k)$$

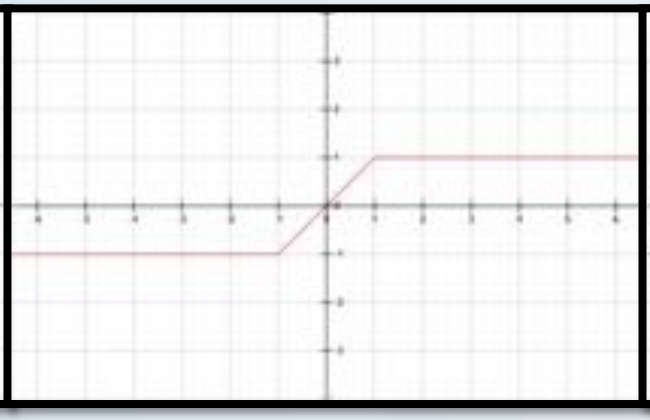


Execution Phase

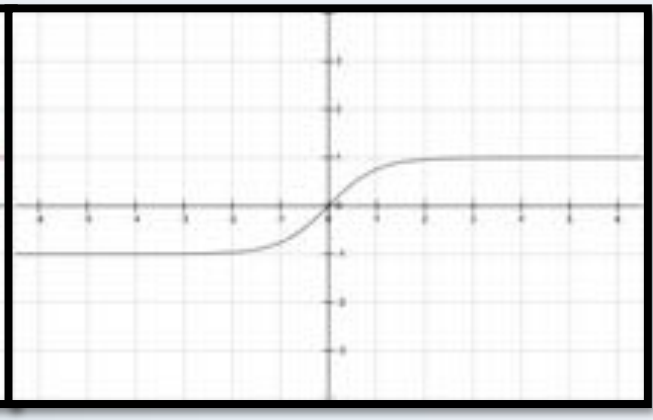
- Some types of activation (transfer) functions



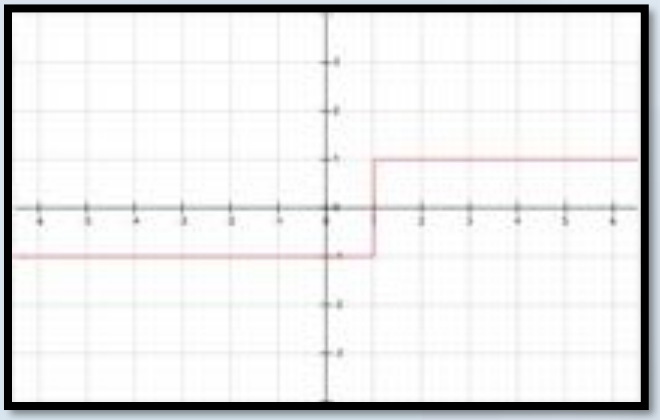
linear



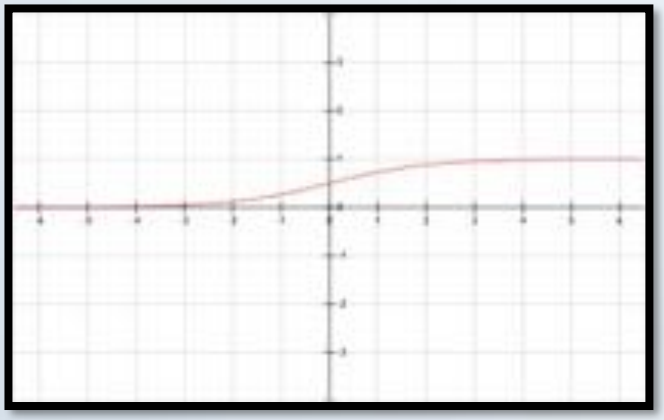
Piecewise linear



tanh



Step function



Sigmoid

Training Phase

- This phase aims to calculate the weights that minimize the error [8]:

$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N} \|\mathbf{y}(\mathbf{x}_n, \mathbf{w}) - \mathbf{t}_n\|^2.$$

- It is a minimization problem with updates as: $\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} - \eta \nabla E(\mathbf{w}^{(\tau)})$

- Initialize the weights randomly
- Do
  - Loop over training data (x,t)
    - Calculate y=execution_phase(x)
    - Calculate error between t and y
    - Calculate delta for all weights between hidden and output
    - Calculate delta for all weights between input and hidden.
    - Update the weights
  - End loop
- Until error is less than a threshold or number of epochs are succeeded

Training Phase

- Delta is defined as follows [8]:

$$\delta_j \equiv \frac{\partial E_n}{\partial a_j}$$

- Delta for weights between hidden and output:

$$\delta_k = y_k - t_k$$

- Delta for weights between input and hidden:

$$\delta_j = h'(a_j) \sum_k w_{kj} \delta_k$$

For more information about the derivations of delta, refer to [8]

Other Neural Networks

- Radial Basis Function Network (Wikipedia)
- Self-Organizing Networks (Wikipedia)
- Learning Vector Quantization (Wikipedia)
- Recurrent Neural Networks (Wikipedia)
- Modeling Neural Networks (Wikipedia)

Text Recognition

- Text Recognition is the conversion of images into written words.
- Character Recognition can be classified into groups:
  - By Acquisition
    - Offline Character Recognition
    - Online Character Recognition
  - By Script
    - Handwritten Text
    - Printed Text

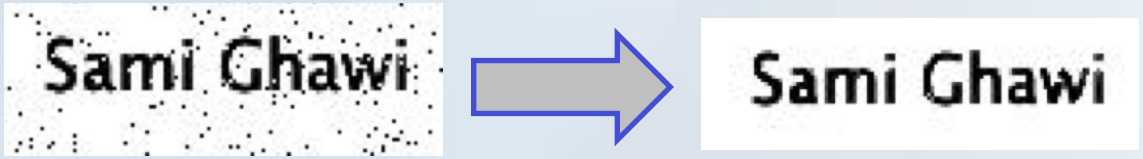
Text Recognition

- Text Recognition is divided to five different stages:
  - Preprocessing
  - Segmentation
  - Feature extraction
  - Recognition
  - Segmentation

Text Recognition

Preprocessing

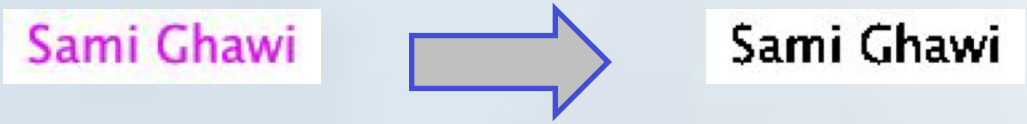
- Noise Reduction
  - Remove the noise that is available in the input image
  - Removing Gaussian noise is done by Gaussian filter
  - Removing Salt and Pepper noise is done by



Text Recognition

Preprocessing

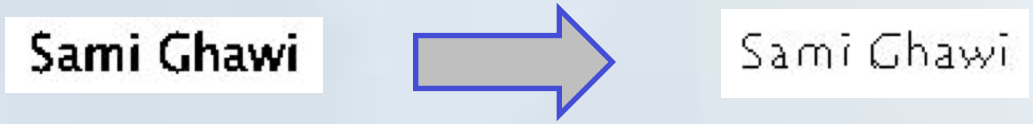
- Convert to binary
  - Convert the image to gray scale from colored image
  - Use a threshold to convert it to binary image (0,1)



Text Recognition

Preprocessing

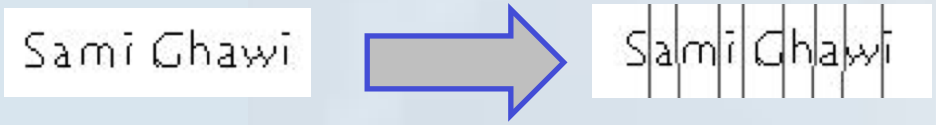
- Thin the image
  - We want to deal with thinned letters in order to find features
  - Can be done by thinning algorithms for skeletonizations



Text Recognition

Segmentation

- Line Segmentation
  - Break pages into lines
- Word Segmentation
  - Break Lines into words
- Character Segmentation
  - Break words into characters or sub-characters



Text Recognition

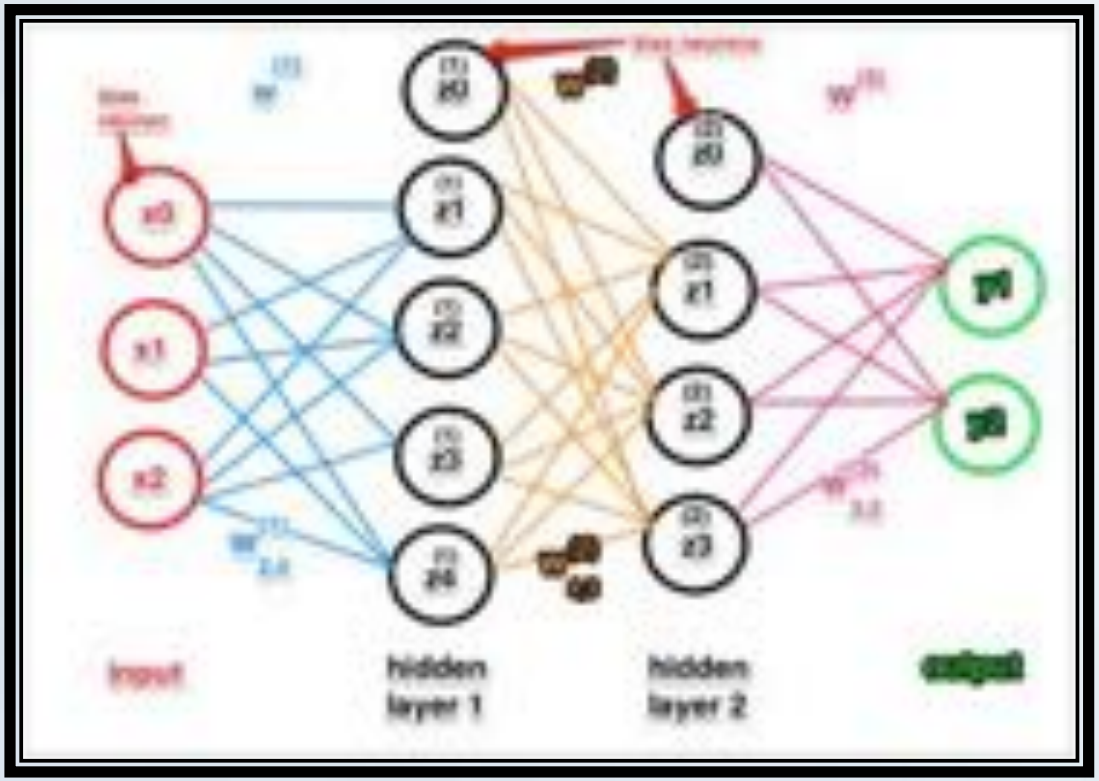
Feature extraction

- Features can be divided to two categories:
  - Local Features (lines, concave part, convex part, edges, branches)
  - Global features (dimensions, number of holes)
- We can either learn the features using machine learning or have a static set.
- In this example, we can use statically defined local features such as lines (horizontal, vertical, diagonal), curves and global features as holes.
- The following letters have the corresponding features (that we try to learn):
  - E: 3 horizontal lines + 1 vertical line
  - A: 1 horizontal lines + 2 diagonal lines
  - O: 1 hole

Text Recognition

Recognition

- The features extracted in the last step are used as inputs
- The number of input nodes equals the number of features
- The number of output nodes equals the number of characters or sub-characters.



Text Recognition

Classification

- The following stage decides the resulting text from the characters we got in the recognition stage.
- It can either be statically defined if the combinations are clear or using machine learning technique if it is complicated.

References

1. Biological Neural Networks – Wikipedia : http://en.wikipedia.org/wiki/Biological_neural_network
2. Introduction to Vision Sensing – University of Text Arlington: http://ranger.uta.edu/~gianluca/teaching/CSE4392-5369_F12/3_CSE4392-5369_IntroVision_Mariottini.pdf
3. D. Hebb, The Organization of Behavior. Wiley, New York, 1949.
4. Rosenblatt's Perceptron - Wikipedia: http://en.wikipedia.org/wiki/Perceptron
5. F. Rosenblatt, The Perceptron - a perceiving and recognizing automaton. Report 85-460-1, Cornell Aeronautical Laboratory, 1957.
6. M. Alstyne, Remaking the PDP Network: A three node solution to the XOR problem. May, 1985.
7. J. Heaton, Introduction to Neural Networks for Java, 2nd edition. Heaton Research, Inc. 2005.
8. C. Bishop. Pattern Recognition and Machine Learning. Springer, 2006.
9. Answers provided by Machine learning books authors: ftp://ftp.sas.com/pub/neural/FAQ3.html