PANA ACADEMY

Neural Networks

Unit 9.6

Introduction

PANA ACADEMY

A neuron is a cell in brain whose principle function is the collection, Processing, and dissemination of electrical signals. Brains Information processing capacity comes from networks of such neurons. Due to this reason some earliest AI work aimed to create such artificial networks.

A neural network can be defined as a model of reasoning based on the human brain.

The brain consists of a densely interconnected set of nerve cells, or basic information processing units, called neurons

The human brain incorporates nearly 10 billion neurons and 60 trillion connections, synapses, between them. By using multiple neurons simultaneously, the brain can perform its functions much faster than the fastest computers in existence today.

Each neuron has a very simple structure, but an army of such elements constitutes a tremendous processing power

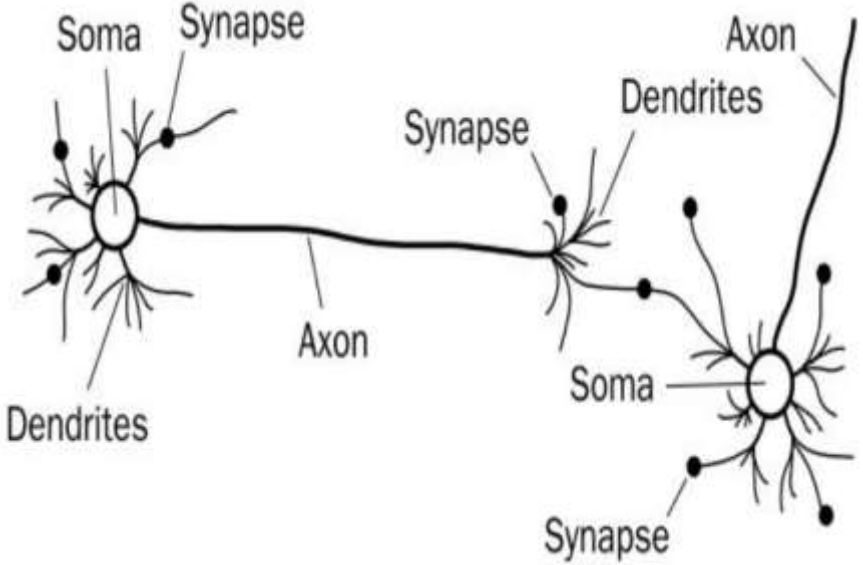
Neuron: fundamental functional unit of all nervous system tissue

Soma: cell body, contain nucleus

Dendrites: a number of fibres, input

Axon: single long fibre with many branches, output

Synapse: junction of dendrites and axon, each neuron form synapse with 10 to 100000 other neurons



PANA ACADEMY

Neural Network

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information.

The key element of this paradigm is the novel structure of the information processing system.

It is composed of a large number of highly interconnected processing elements (neurones) working in unison to solve specific problems. ANNs, like people, learn by example.

An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process.

PANA ACADEMY

Why use Neural Network?

Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques.

A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze. Other advantages include:

1. Adaptive learning: An ability to learn how to do tasks based on the data given for training or initial experience.

2. Self-Organisation: An ANN can create its own organisation or representation of the information it receives during learning time.

3. Real Time Operation: ANN computations may be carried out in parallel, and special hardware devices are being designed and manufactured which take advantage of this capability.

4. Fault Tolerance via Redundant Information Coding: Partial destruction of a network leads to the corresponding degradation of performance. However, some network capabilities may be retained even with major network damage

Neural networks take a different approach to problem solving than that of conventional computers.

Conventional computers use an algorithmic approach i.e. the computer follows a set of instructions in order to solve a problem.

Unless the specific steps that the computer needs to follow are known the computer cannot solve the problem.

That restricts the problem solving capability of conventional computers to problems that we already understand and know how to solve.

But computers would be so much more useful if they could do things that we don't exactly know how to do.

Neural networks process information in a similar way the human brain does.

The network is composed of a large number of highly interconnected processing elements(neurones) working in parallel to solve a specific problem.

Neural networks learn by example.

They cannot be programmed to perform a specific task.

The examples must be selected carefully otherwise useful time is wasted or even worse the network might be functioning incorrectly.

The disadvantage is that because the network finds out how to solve the problem by itself, its operation can be unpredictable.

On the other hand, conventional computers use a cognitive approach to problem solving; the way the problem is to solved must be known and stated in small unambiguous instructions.

These instructions are then converted to a high level language program and then into machine code that the computer can understand.
These machines are totally predictable; if anything goes wrong is due to a software or hardware fault.

Mcculloch and Pitts model

The McCulloch-Pitts neuron is a foundational model in the field of artificial neural networks, proposed by Warren McCulloch and Walter Pitts in 1943.

This model represents a simplified abstraction of biological neurons and serves as a basis for understanding more complex neural network architectures.

Units of Neural Network

Nodes(units): Nodes represent a cell of neural network.

Links: another node.

Activation: Activations are inputs to or outputs from a unit.

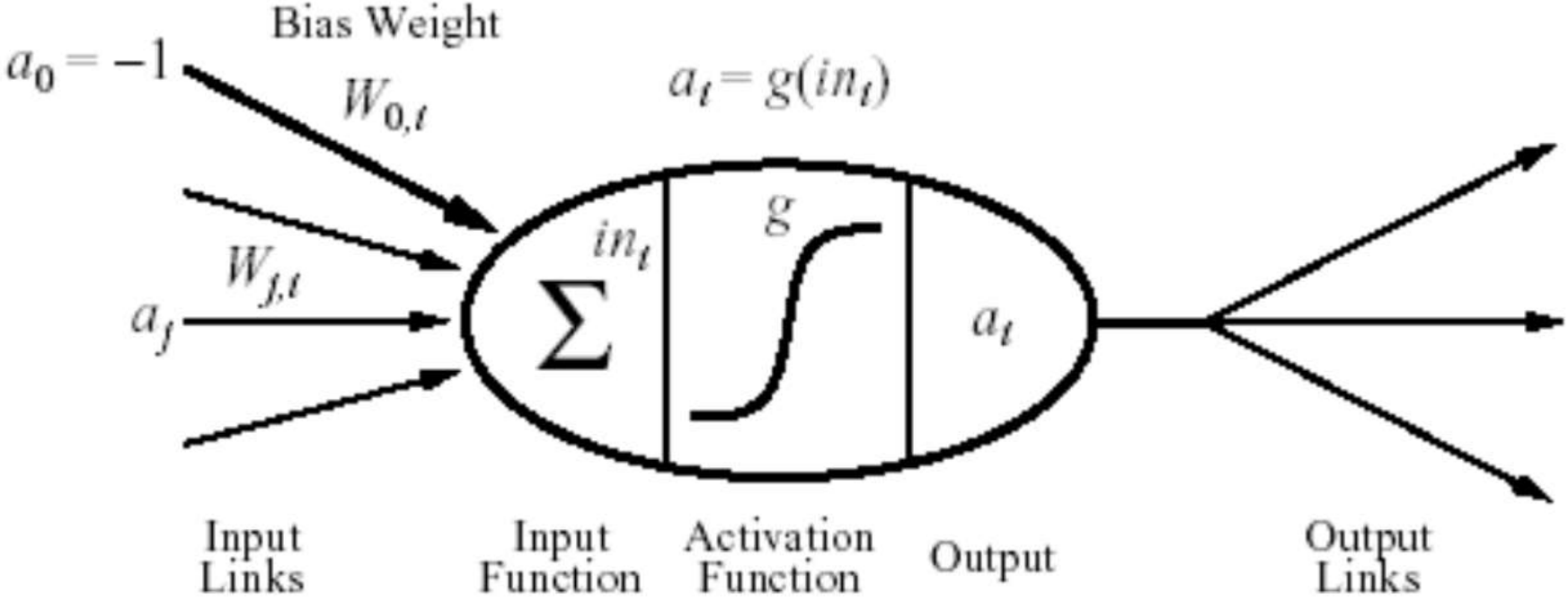
Weight: Each link has weight associated with it which determines strength and sign of the connection.

Activation function: A function which is used to derive output activation from the input activations to a given node is called activation function.

Bias Weight: Bias weight is used to set the threshold for a unit. Unit is activated when the weighted sum of real inputs exceeds the bias weight.

PANA ACADEMY

Simple Neural Network model



It fires when a linear combination of its inputs exceeds some threshold.

PANA ACADEMY

A neural network is composed of nodes (units) connected by directed links A link from unit j to i serve to propagate the activation a_j from j to i. Each link has some numeric weight $W_{j,i}$ associated with it, which determines strength and sign of connection.

Each unit first computes a weighted sum of it's inputs:

$$in_i = \sum_{J=0}^{n} W_{j,i}\ a_j$$

Then it applies activation function g to this sum to derive the output:

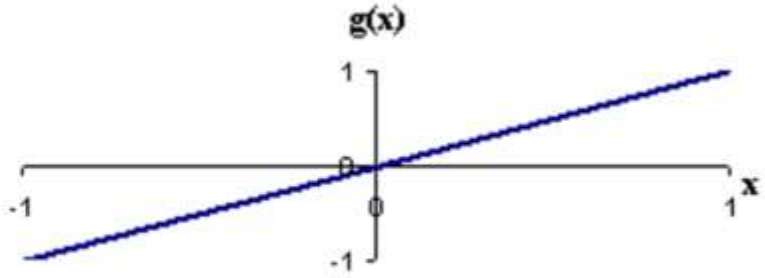
$$a_i = g(in_i) = g\left(\sum_{J=0}^{n} W_{j,i}\ a_j\right)$$

Here, a_j output activation from unit j and $W_{j,i}$ is the weight on the link j to this node.
Activation function typically falls into one of three categories:

- Linear
- Threshold (*Heaviside function)*
- Sigmoid
- Sign

For **linear activation functions**, the output activity is proportional to the total weighted output.

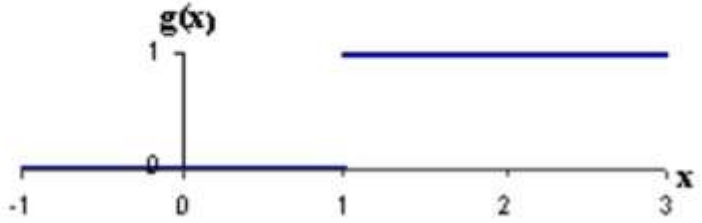
$g(x) = k\,x + c$, where k and x are constant



For **threshold activation functions**, the output are set at one of two levels, depending on whether the total input is greater than or less than some threshold value.

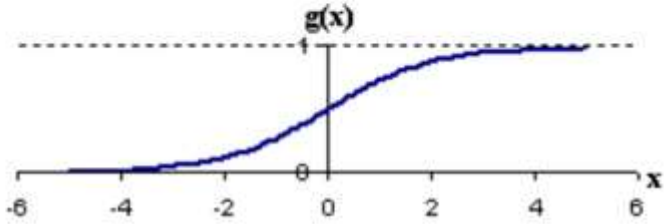
$g(x) = 1$ if $x >= k$

$= 0$ if $x < k$

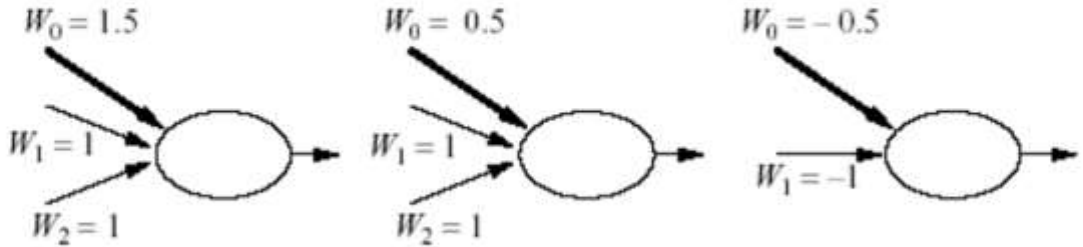


For **sigmoid activation functions**, the output varies continuously but not linearly as the input changes. Sigmoid units bear a greater resemblance to real neurons than do linear or threshold units. It has the advantage of differentiable.

$g(x) = 1/(1 + e^{-x})$

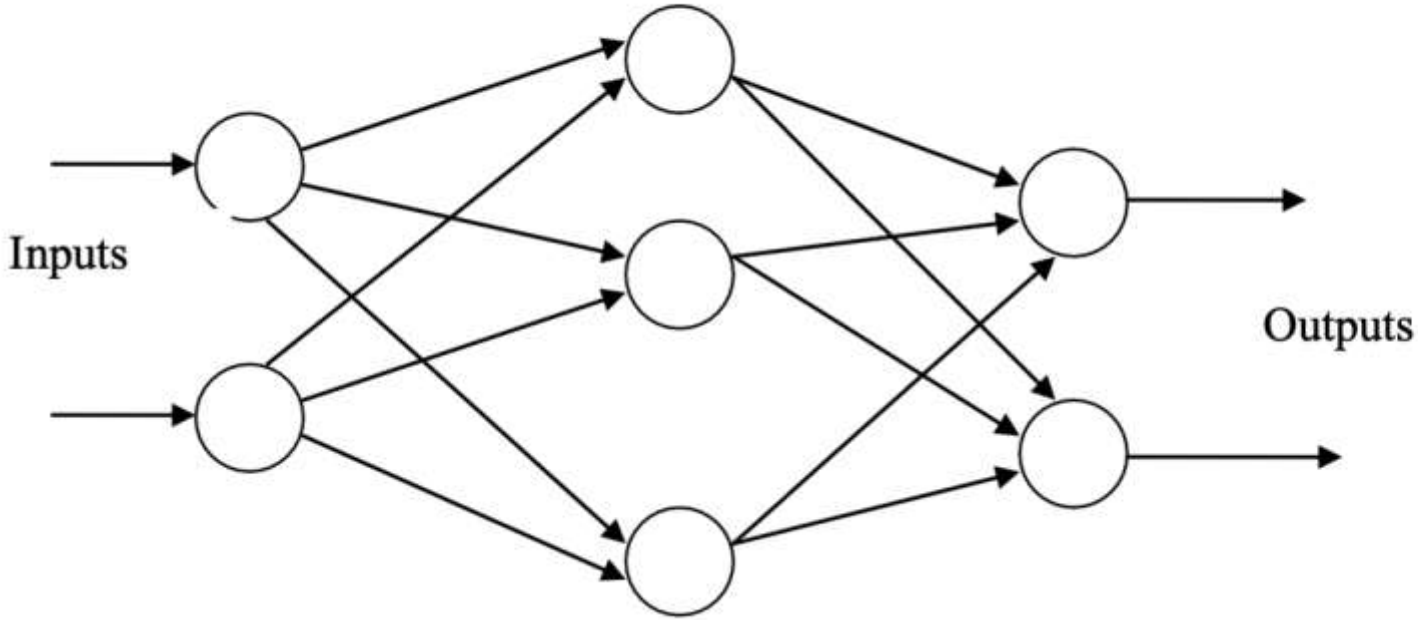


Realizing logic gates by using Neurons:

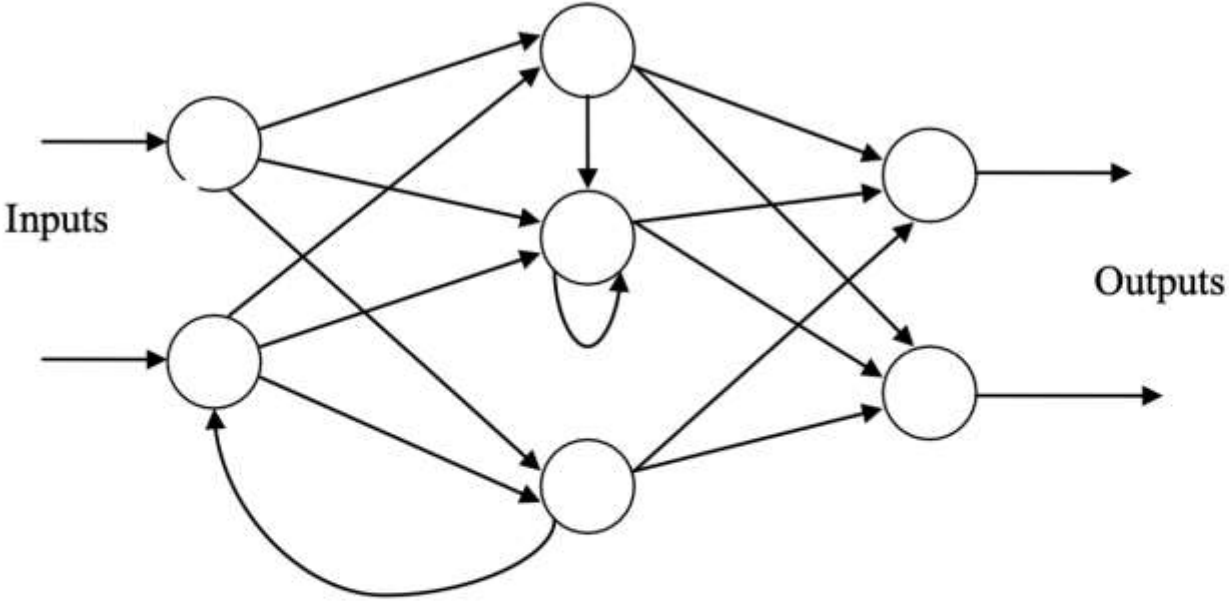


Feed-forward networks:

Feed-forward ANNs allow signals to travel one way only; from input to output. There is no feedback (loops) i.e. the output of any layer does not affect that same layer. Feed-forward ANNs tend to be straight forward networks that associate inputs with outputs. They are extensively used in pattern recognition. This type of organization is also referred to as bottom-up or top-down.



Feedback networks (Recurrent networks:)



Feedback networks (figure 1) can have signals traveling in both directions by introducing loops in the network. Feedback networks are very powerful and can get extremely complicated. Feedback networks are dynamic; their 'state' is changing continuously until they reach an equilibrium point. They remain at the equilibrium point until the input changes and a new equilibrium needs to be found. Feedback architectures are also referred to as interactive or recurrent.

- The neuron computes the weighted sum of the input signals and compares the result with a **threshold value**, θ. If the net input is less than the threshold, the neuron output is −1. But if the net input is greater than or equal to the threshold, the neuron becomes activated and its output attains a value +1.
- The neuron uses the following transfer or **activation function**
- $X = \sum_{i=1}^{n} x_i w_i \qquad Y = \begin{cases} +1 \; if \; X \geq \theta \\ -1 \; if \; X \leq \theta \end{cases}$
- This type of activation function is called a **sign function**

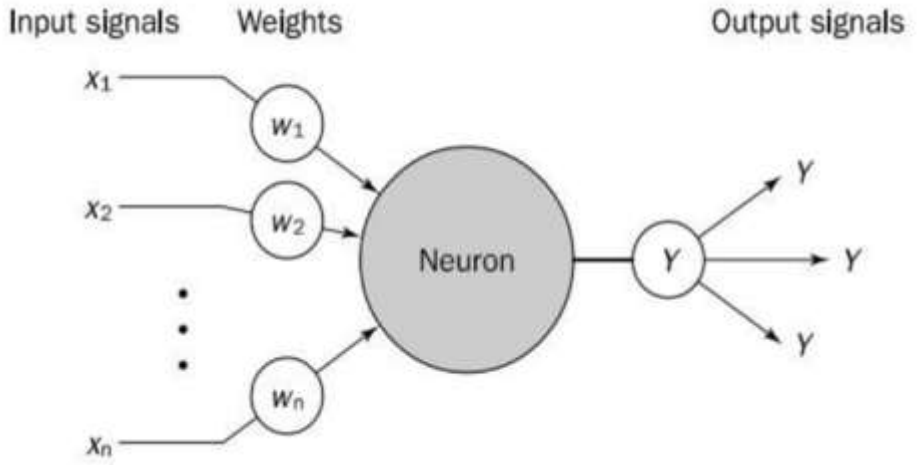
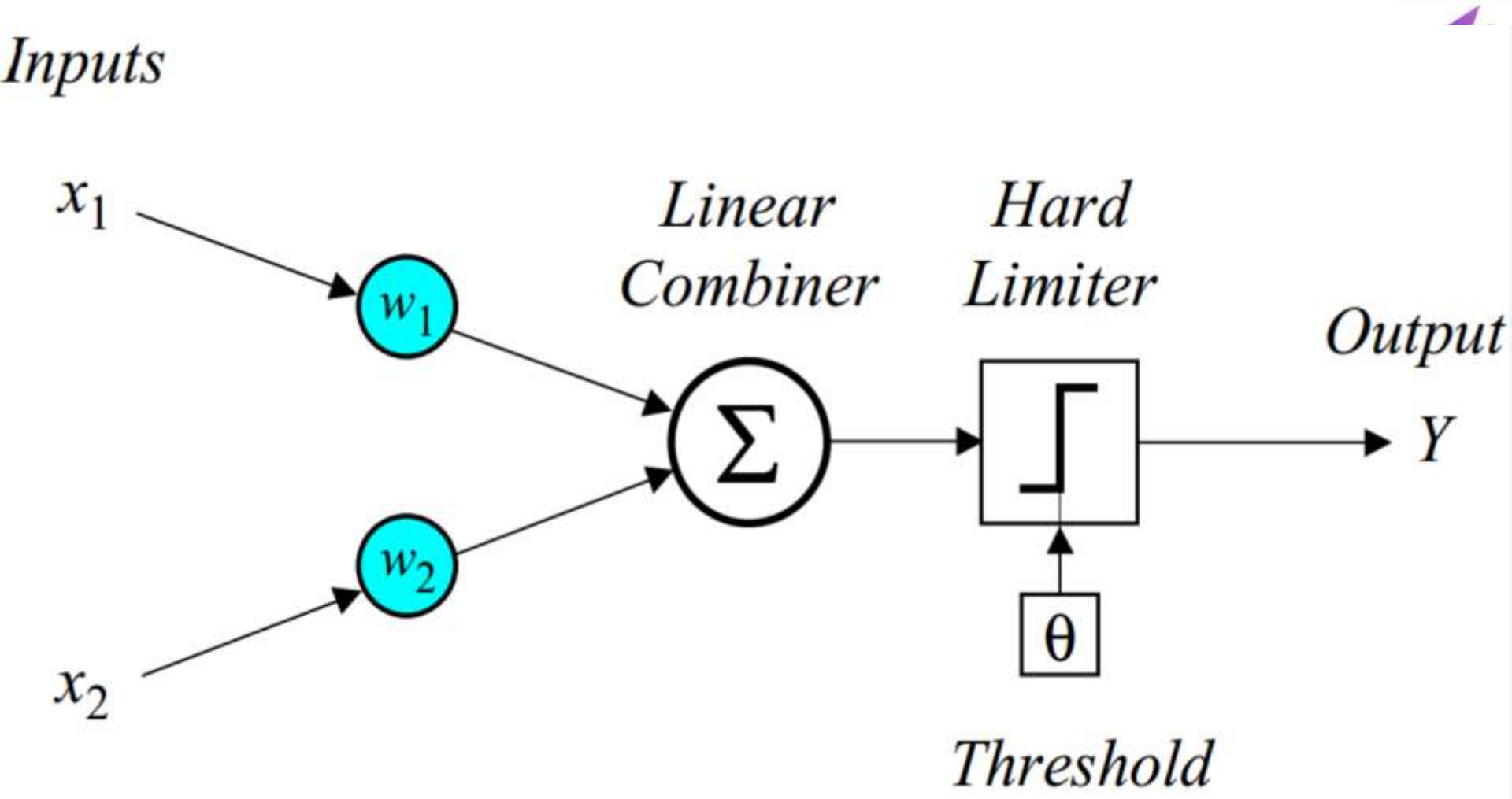


Diagram of a neuron

PANA ACADEMY

Perceptron

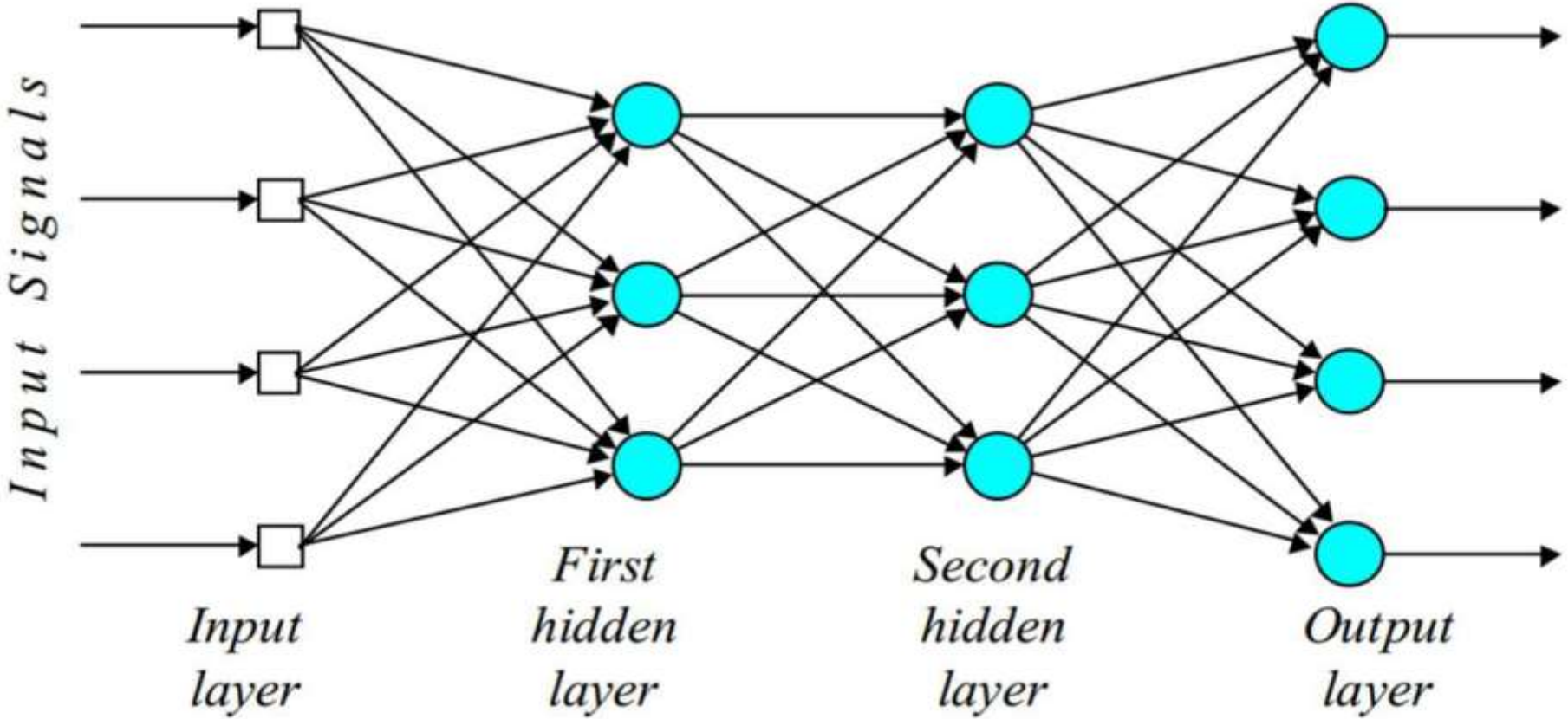
- In 1958, Frank Rosenblatt introduced a training algorithm that provided the first procedure for training a simple ANN : a **perceptron**
- The perceptron is the simplest form of a neural network. It consists of a single neuron with *adjustable* synaptic weights and a **hard limiter**
- The operation of Rosenblatt's perceptron is based on the McCulloch and Pitts neuron model. The model consists of a **linear combiner** followed by a hard limiter
- The weighted sum of the inputs is applied to the hard limiter, which produces an output equal to +1 if its input is positive and -1 if its input is negative


Inputs
x_1
x_2
w_1
w_2
Linear
Combiner
Σ
Hard
Limiter
θ
Threshold
Output
Y

PANA ACADEMY

Multilayer perceptron

- A multi layer perceptron is a feed forward neural network with one or more hidden layers
- The network consists of :
  - Input Layer
  - Hidden Layer
  - Output Layer
- The input signal is propagated in a forward direction in a layer-by-layer basis


Input Signals
Output Signals
Input layer
First hidden layer
Second hidden layer
Output layer

Back Propagation

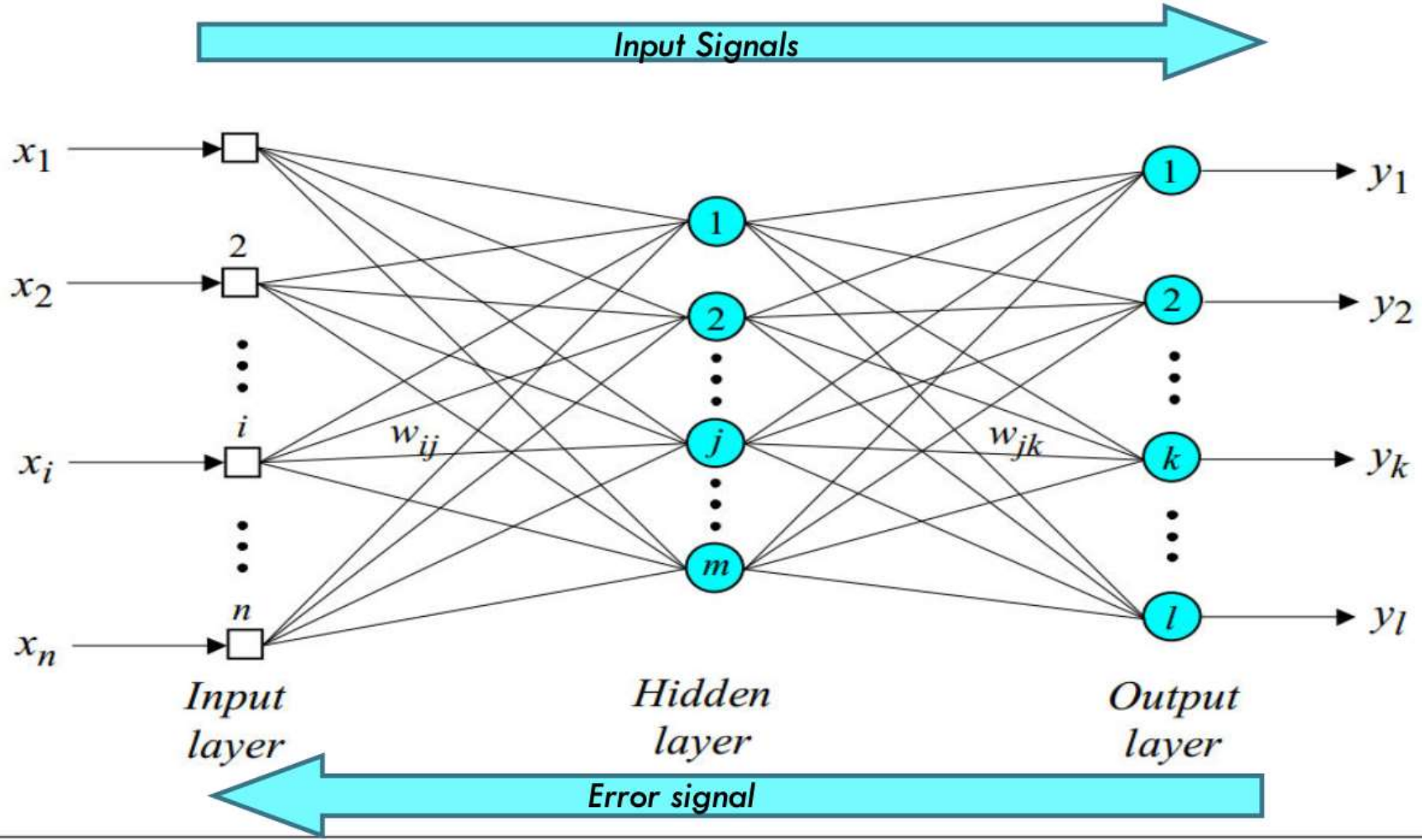
Learning in a multilayer network proceeds the same way as for a perceptron

A training set of input patterns is presented to the network

The network computes its output pattern, and if there is an error –or other word difference between actual and desired output pattern – the weight are adjusted to reduce the error

In a back-propagation neural network, the learning algorithm has two phases

First, a training input pattern is presented to the network input layer. The network propagates the input pattern from layer to layer until the output pattern is generated by the out layer


Input Signals
x_1
x_2
x_i
x_n
1
2
i
n
w_{ij}
1
2
j
m
w_{jk}
1
2
k
l
y_1
y_2
y_k
y_l
Input layer
Hidden layer
Output layer
Error signal

PANA ACADEMY

Learning Rate

The learning rate is a critical hyperparameter in machine learning that influences how quickly or slowly a model updates its parameters during training.

The learning rate controls the step size at which the weights of a neural network are adjusted in response to the gradient of the loss function.

It determines how fast a model learns from the data.

High Learning Rate:

Can cause the model to oscillate or diverge, failing to settle at a minimum.

May skip over minima entirely, leading to erratic behavior during training.

Low Learning Rate:

Results in slow convergence, prolonging training times.

Increases the risk of getting stuck in local minima.

Adaptive Learning Rates:

Algorithms like Adam, RMSprop, and AdaGrad dynamically adjust the learning rate based on historical gradient information, allowing for more efficient training.

Gradient Descent

Gradient descent is a fundamental optimization algorithm widely used in machine learning and deep learning to minimize a cost or loss function

Gradient descent is an iterative first-order optimization algorithm used to find the local minimum of a differentiable function. It works by taking steps proportional to the negative of the gradient (or approximate gradient) of the function at the current point.

The delta rule

The Delta Rule, also known as the Delta Learning Rule, is a fundamental concept in artificial neural networks, particularly in training single-layer perceptrons.

It is a gradient descent learning technique that adjusts the weights of the network based on the error between the expected and actual outputs.

The Delta Rule is derived from the gradient descent optimization method, which aims to minimize the error in predictions by adjusting weights in the opposite direction of the gradient of the loss function.

The delta rule

The Delta Rule is primarily used in supervised learning scenarios where labeled data is available.

It is effective for linearly separable problems, such as basic classification tasks.

The choice of activation function affects how the Delta Rule is applied. For instance, if a linear activation function is used, the rule simplifies to:

$$\Delta w_{ji} = \alpha(t_j - y_j)x_i$$

Hebbian Learning

Hebbian learning is a fundamental principle in neuroscience and artificial intelligence that explains how neural connections strengthen through repeated activation.

The phrase often associated with this theory is "cells that fire together, wire together," which encapsulates the core idea of how learning occurs at the synaptic level.

when two neurons are activated simultaneously, the synaptic connection between them strengthens. This process enhances the efficiency of communication between these neurons

Hebbian Learning

Learning occurs through the growth of synaptic knobs (or terminals) at the axon ends, which improves the transmission efficiency between neurons.

Repeated co-activation leads to lasting changes in the strength of synapses.

Mathematical formula:

$$\Delta w_{ij} = \eta \cdot x_i \cdot y_j$$

Δw_{ij} is the change in weight between neuron

η is the learning rate.

x_i is the activation of neuron

y_j is the activation of neuron

Adaline Network

Adaline (Adaptive Linear Neuron) is a type of artificial neural network introduced by Bernard Widrow and Ted Hoff in 1960.

It is primarily used for supervised learning tasks, particularly in regression and binary classification problems.

Adaline consists of a single linear unit that processes inputs and produces an output. The architecture typically includes:

Input Layer: Receives input data.

Weight Adjustment Unit: Applies weights to the inputs.

Output Layer: Produces the final output.

Adaline Network

Unlike other neural networks that may use non-linear activation functions, Adaline employs a linear activation function.

This means the output is a linear combination of the inputs.

Computationally efficient due to its simple architecture and linearity.

Training involves straightforward calculations with fewer parameters.

PANA ACADEMY

Hopfield network

Hopfield neural network was invented by Dr. John J. Hopfield in 1982.

It consists of a single layer which contains one or more fully connected recurrent neurons.

The Hopfield network is commonly used for auto-association and optimization tasks.

Hopfield network

- neural networks with feedback – Hopfield networks presence of such loops has a profound impact on the

- learning capability of the network

- After applying a new input, the network output is calculated and feedback to adjust the input. Then the output is calculated again, and the process is repeated until the output becomes constant.[Working mechanism of Recurrent network]

Discrete hopfield network

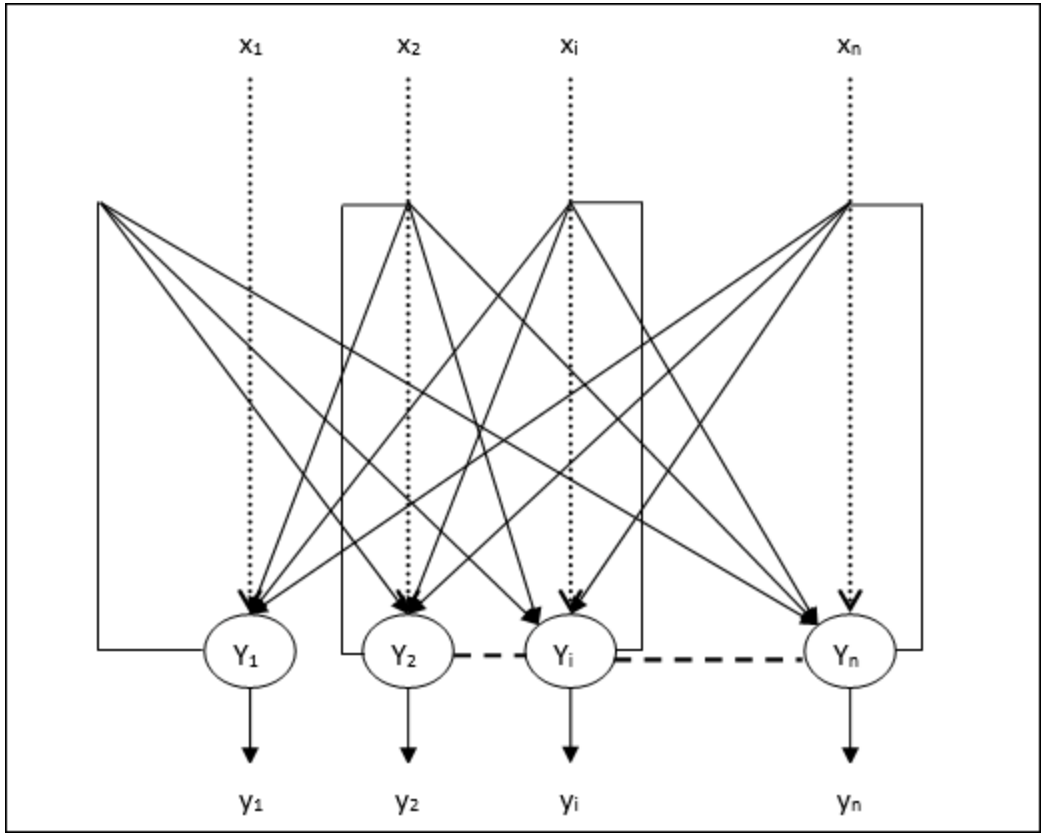
A Hopfield network which operates in a discrete line fashion or in other words, it can be said the input and output patterns are discrete vector, which can be either binary 0,1 or bipolar +1,−1 in nature.

The network has symmetrical weights with no self-connections i.e., $w_{ij} = w_{ji}$ and $w_{ii} = 0$.

PANA ACADEMY

Architecture

- This model consists of neurons with one inverting and one non-inverting output.
- The output of each neuron should be the input of other neurons but not the input of self.
- Weight/connection strength is represented by w_{ij}.
- Connections can be excitatory as well as inhibitory. It would be excitatory, if the output of the neuron is same as the input, otherwise inhibitory.
- Weights should be symmetrical, i.e. $w_{ij} = w_{ji}$



Regression, CNN,RNN,LSTM-Supervised Learning

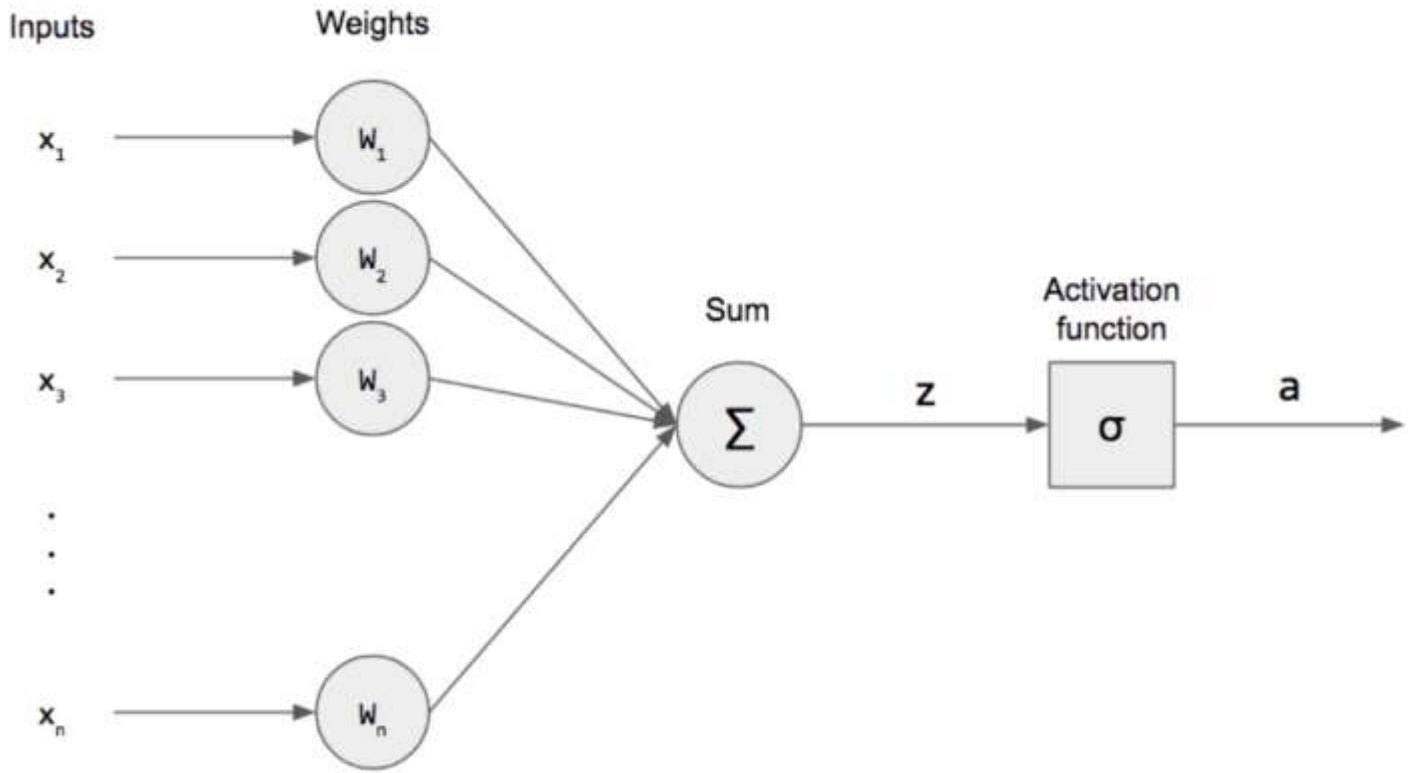
PANA ACADEMY

- We have Inputs, outputs, train the model till we get a desired output.
- Our model generates output, then we have comparison with the actual output.
- Then we decide, whether model has given the correct output(close to correct output) or not.
- If Predicted output not equal to correct output, we use gradient descent(weight adjustment)

The model is directed.

- Input-> layer 1 ->Layer 2 -> output.

Gate realization using neural network



PANA ACADEMY

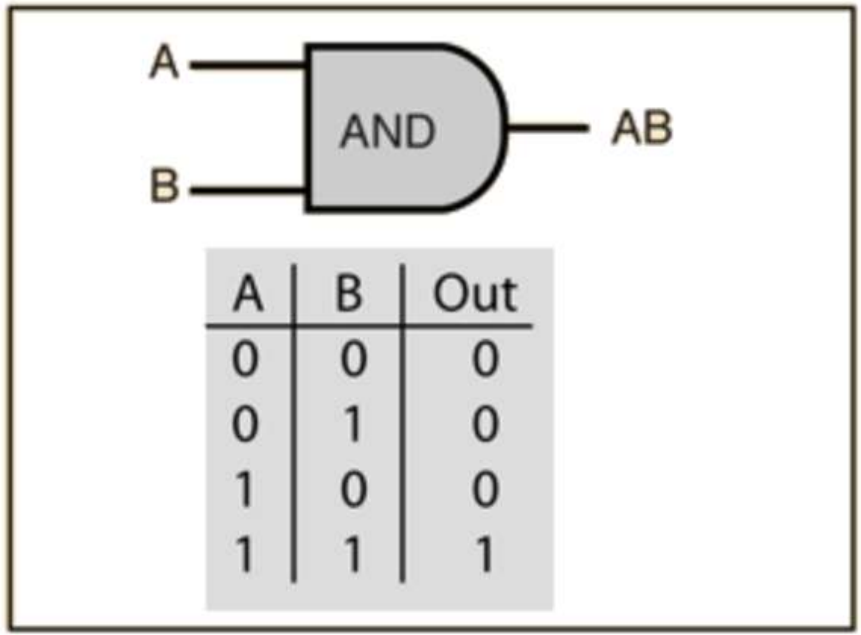
Steps to show how neural network learn

- Initialize weight values and bias
- Forward Propagate
- Check the error
- Backpropagate and Adjust weights and bias
- Repeat for all training examples

AND gate

The question is, what are the weights and bias for the AND perceptron?

First, we need to understand that the output of an AND gate is 1 only if both inputs (in this case, x1 and x2) are 1. So, following the steps listed above;



A	B	Out
0	0	0
0	1	0
1	0	0
1	1	1

AND Gate

Row 1

- From w1*x1+w2*x2+b, initializing w1, w2, as 1 and b as –1, we get;

x1(1)+x2(1)–1

- Passing the first row of the AND logic table (x1=0, x2=0), we get;

0+0–1 = –1

- From the Perceptron rule, if Wx+b≤0, then y`=0. Therefore, this row is correct, and no need for Backpropagation.

Row 2

- Passing (x1=0 and x2=1), we get;

$0+1-1 = 0$

- From the Perceptron rule, if $Wx+b \leq 0$, then $y` = 0$. This row is correct, as the output is 0 for the AND gate.
- From the Perceptron rule, this works (for both row 1, row 2 and 3).

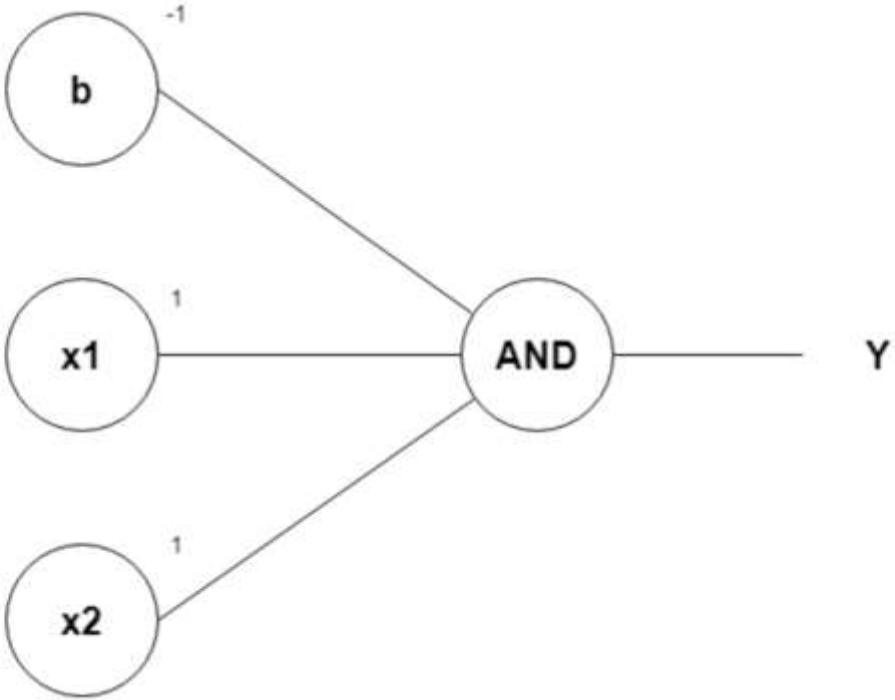
Row 4

- Passing (x1=1 and x2=1), we get;

$1+1-1 = 1$

Therefore, we can conclude that the model to achieve an AND gate, using the Perceptron algorithm is;

$x1+x2-1$



PANA ACADEMY

MCQ

1. What is the primary function of biological neural networks?

   A) Data storage

   B) Information processing and transmission

   C) Image recognition

   D) Data compression

   Answer: B

2. Which of the following best describes an artificial neural network (ANN)?

A) A collection of biological neurons

B) A computational model inspired by biological neural networks

C) A type of software for data analysis

D) A hardware device for processing information

Answer: B

PANA ACADEMY

3. What does a McCulloch-Pitts neuron model represent?

A) A biological neuron

B) A simple computational model for binary decision making

C) An advanced neural network architecture

D) A learning algorithm

Answer: B

PANA ACADEMY

In the context of ANNs, what does the activation function do?

A) Determines the input to the network

B) Produces the final output of the neuron

C) Adjusts the weights of connections

D) Initializes the network

Answer: B

Which activation function is commonly used in binary classification tasks?

A) ReLU

B) Sigmoid

C) Tanh

D) Softmax

Answer: B

What is a key characteristic of feedforward neural networks?

A) They have feedback loops.

B) Information flows in one direction only.

C) They can process sequences of data.

D) They use recurrent connections.

Answer: B

What type of neural network is a perceptron?

A) Multi-layer feedforward network

B) Single-layer feedforward network

C) Recurrent network

D) Convolutional network

Answer: B

What does the learning rate control in an ANN?

A) The number of hidden layers

B) The speed at which weights are updated during training

C) The size of the input data

D) The type of activation function used

Answer: B

PANA ACADEMY

Which optimization method is commonly used to minimize loss in neural networks?

A) Gradient Descent

B) Genetic Algorithm

C) Simulated Annealing

D) Particle Swarm Optimization

Answer: A

The Delta Rule is primarily associated with which type of learning?

A) Supervised learning

B) Unsupervised learning

C) Reinforcement learning

D) Semi-supervised learning

Answer: A

Hebbian learning is often summarized by which phrase?

A) "What fires together, wires together."

B) "Learning from mistakes."

C) "Less is more."

D) "Practice makes perfect."

Answer: A

What is the primary purpose of the Adaline network?

A) To classify images

B) To perform linear regression or binary classification tasks

C) To generate new data samples

D) To recognize patterns in time series data

Answer: B

Which architecture allows for multiple layers and can learn complex patterns?

A) Single-layer perceptron

B) Multilayer perceptron (MLP)

C) Hopfield network

D) Radial basis function network

Answer: B

PANA ACADEMY

What is backpropagation primarily used for in neural networks?

A) Data preprocessing

B) Weight adjustment during training

C) Feature extraction

D) Model evaluation

Answer: B

In a Hopfield network, what type of memory does it exhibit?

A) Short-term memory only

B) Long-term memory only

C) Associative memory

D) Procedural memory

Answer: C

Which type of neural network architecture is characterized by feedback connections?

A) Feedforward Neural Network

B) Recurrent Neural Network (RNN)

C) Convolutional Neural Network (CNN)

D) Single-layer Perceptron

Answer: B

What does the term "overfitting" refer to in machine learning?

A) When a model performs well on training data but poorly on unseen data.

B) When a model performs poorly on both training and testing data.

C) When a model learns too quickly.

D) When a model has too few parameters.

Answer: A

Which activation function allows for non-linear transformations and helps prevent vanishing gradients in deep networks?

A) Sigmoid

B) Tanh

C) ReLU (Rectified Linear Unit)

D) Softmax

Answer: C

In gradient descent, what does the term "learning rate" signify?

A) The number of epochs in training.

B) The step size at each iteration while moving toward a minimum.

C) The total number of parameters in the model.

D) The complexity of the model architecture.

Answer: B

What type of problem can a single-layer perceptron solve effectively?

A) Non-linear problems

B) Linearly separable problems

C) Time-series forecasting

D) Image recognition problems

Answer: B

Which neural network architecture is specifically designed for image processing tasks?

A) Radial Basis Function Network

B) Convolutional Neural Network (CNN)

C) Multilayer Perceptron (MLP)

D) Hopfield Network

Answer: B

In which scenario would you use a recurrent neural network (RNN)?

A) For static image classification

B) For time series prediction

C) For simple linear regression

D) For clustering tasks

Answer: B

What does backpropagation use to update weights in a neural network?

A) Forward pass errors

B) Gradient descent

C) Hebbian learning

D) Genetic algorithms

Answer: B

In an Adaline network, what type of activation function is typically used?

– A) Step function

– B) Linear function

– C) Sigmoid function

– D) Tanh function

Answer: B

Which algorithm is commonly associated with optimizing weight adjustments in ANNs?

– A) K-means clustering

– B) Genetic algorithm

– C) Gradient descent

– D) Principal component analysis

What is one limitation of using Hebbian learning alone in neural networks?

– A) It requires labeled data

– B) It cannot handle negative weights

– C) It doesn’t account for inhibitory signals

– D) It can only be applied to single-layer networks

Answer: C

PANA ACADEMY

The primary goal of an artificial neural network (ANN)?

– A) To replicate human brain functions exactly

– B) To learn from data and make predictions or decisions

– C) To store large amounts of information

– D) To perform arithmetic calculations quickly

Answer: B

Which type of neural network uses convolutional layers to detect features in input data?

– A) Recurrent Neural Network

– B) Convolutional Neural Network (CNN)

– C) Multilayer Perceptron (MLP)

– D) Radial Basis Function Network

Answer: B

What does a Hopfield Network primarily serve as?

– A) Classification tool

– B) Associative memory system

– C) Regression tool

– D) Clustering algorithm

Answer: B

In which type of learning does an ANN adjust its weights based on feedback from its output compared to expected results?

– A) Supervised learning

– B) Unsupervised learning

– C) Reinforcement learning

– D) Semi-supervised learning

Answer: A

Which mathematical operation forms the basis for calculating outputs in an artificial neuron?

– A) Addition only

– B) Weighted sum followed by an activation function

– C) Multiplication only

– D) Division only

Answer: B

What is one advantage of using multilayer perceptrons over single-layer perceptrons?

– A) They are easier to train

– B) They can solve non-linear problems

– C) They require less computational power

– D) They have fewer parameters to tune

Answer: B

When using gradient descent, what happens if the learning rate is set too high?

A）The model converges quickly to an optimal solution

B）The model may oscillate or diverge

C）The training process becomes slower

D）The model will always find a global minimum

Answer：B

Which layer in a neural network performs feature extraction from inputs？

A）Input layer

B）Hidden layer

C）Output layer

D）Bias layer

Answer：B

PANA ACADEMY

In backpropagation, which direction do gradients flow during weight updates？

—A）From output to input

—B）From input to output

—C）Randomly throughout the network

—D）Only through hidden layers

Answer：A

PANA ACADEMY

What role does bias play in an artificial neuron？

—A）It adjusts the output independently from inputs

—B）It prevents overfitting

—C）It increases computational complexity

—D）It serves as an additional input layer

Answer：A

How do convolutional layers differ from fully connected layers？

—A）Convolutional layers connect every neuron to every other neuron

—B）Convolutional layers apply filters to detect features while reducing dimensionality

—C）Fully connected layers are used only for classification tasks

—D）Convolutional layers cannot learn features from images

Answer：B

Which type of activation function outputs values between −1 and +1？

—A）Sigmoid

—B）ReLU

—C）Tanh

—D）Softmax

Answer：C

What distinguishes recurrent neural networks from feedforward networks？

—A）RNNs have no hidden layers

—B）RNNs can maintain memory across time steps

—C）RNNs are simpler than feedforward networks

—D）RNNs do not use weights for connections

Answer：B

PANA ACADEMY

What technique helps prevent overfitting in deep learning models？

—A）Increasing complexity

—B）Regularization techniques

—C）Decreasing dataset size

—D）Using fewer epochs

Answer：B

Which algorithm adjusts weights based on both current and previous errors？

—A）Gradient descent

—B）Hebbian learning

—C）Backpropagation

—D）Reinforcement learning

Answer：C

In Hebbian learning, what happens when two neurons are activated simultaneously？

—A）Their connection weakens

—B）Their connection strengthens

—C）No change occurs

—D）They become unresponsive

Answer：B

Which architecture allows for feedback loops within neurons？

—A）Feedforward Neural Networks

—B）Radial Basis Function Networks

—C）Recurrent Neural Networks

—D）Single-layer Perceptrons

Answer：C

What is one key feature of multilayer perceptrons (MLPs)?

—A）They have no hidden layers

—B）They can learn complex non-linear functions

—C）They use only linear activation functions

—D）They do not require training data

Answer：B

Which method is typically used to optimize weight adjustments during training？

--A)Stochastic gradient descent

--B)K-means clustering

--C)Support vector machines

--D)Principal component analysis

Answer:A

In which scenario would you prefer using a Hopfield Network？

--A)For image classification

--B)For associative memory tasks

--C)For time series prediction

--D)For regression analysis

Answer:B

How does an Adaline network differ from a perceptron？

--A)Adaline uses linear activation while perceptron uses step activation

--B)Adaline cannot learn from errors while perceptron can

--C)Adaline has more layers than perceptron

--D)Adaline requires more computational resources than perceptron

Answer:A

What defines supervised learning within ANNs？

--A)Learning without labeled data

--B)Learning with labeled input-output pairs

--C)Learning through trial and error

--D)Learning without any feedback

Answer:B