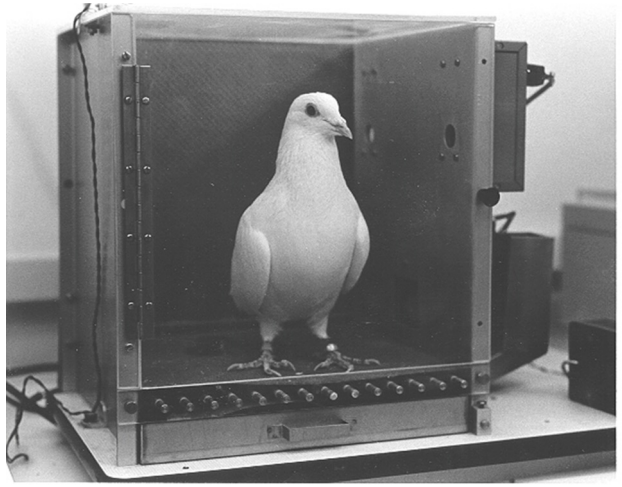
Introduction to Artificial Neural Networks (R)

Contents

Pigeons as Art Experts

- Experiment:
- Pigeon in Skinner box
- Present paintings of two different artists (e.g. Chagall / Van Gogh)
- Reward for pecking when presented a particular artist (e.g. Van Gogh)



- Pigeons were able to discriminate between Van Gogh and Chagall with 95% accuracy (When presented with pictures they had been trained on)
- Discrimination still 85% successful for previously unseen paintings of the artists.

Pigeons as Art Experts

Pigeons do not simply memorize the pictures

They can extract and recognise patterns (the ‘style’)

They generalize from the already seen to make predictions

This is what neural networks (biological and artificial) are good at (unlike conventional computer).

Neural Networks Introduction

- Artificial neural network (ANN) is a machine learning approach that models human brain and consists of a number of artificial neurons.

- Each neuron in ANN receives a number of inputs.

- An activation function is applied to these inputs which results in activation level of neuron (output value of the neuron).

How do our brains work?

The Brain is a massively parallel information processing system. Our brains are a huge network of processing elements. A typical brain contains a network of 10 billion neurons.

Features of Neural Networks

- Massive connectivity
- Nonlinear, Parallel, Robust and Fault Tolerant
- Capability to adapt to surroundings
- Ability to learn and generalize from known examples
- Collective behaviour is different from individual behaviour

- Artificial Neural Networks mimics some of the properties of the biological neural networks

Why Neural Networks?

- There are two basic reasons why we are interested in building artificial neural networks (ANNs):

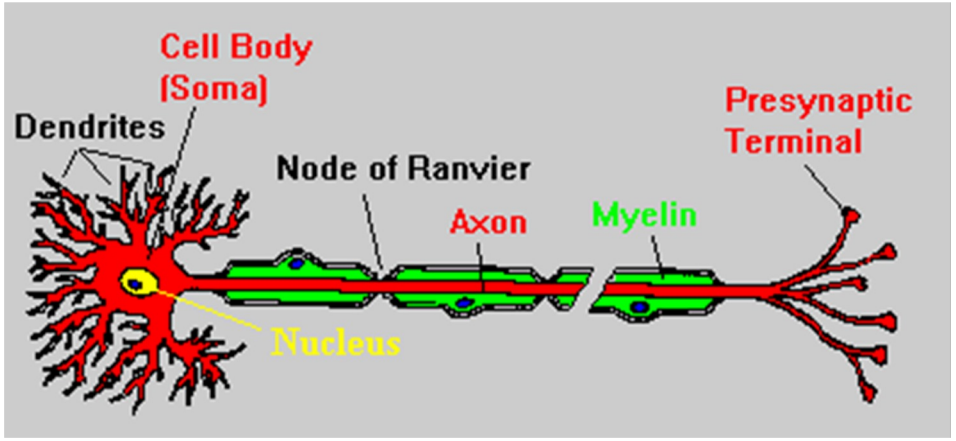
- Technical viewpoint:
- Some problems such as character recognition or the prediction of future states of a system require massively parallel and adaptive processing.

- Biological viewpoint:
- ANNs can be used to replicate and simulate components of the human (or animal) brain, thereby giving us insight into natural information processing.

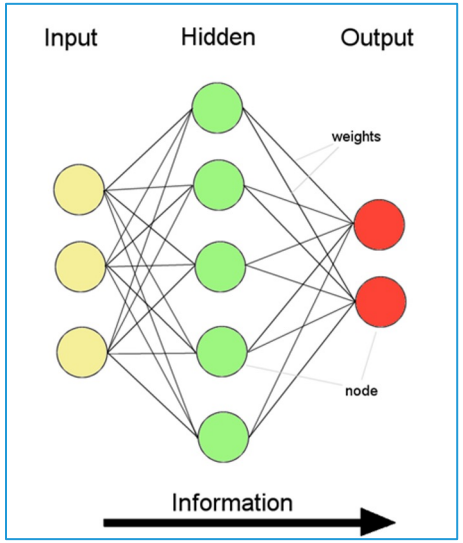
Biological Neuron vs Artificial Neuron (Perceptron)

- Information flow is unidirectional
- Data is presented to Input layer
- Passed on to Hidden Layer
- Passed on to Output layer
- Information is distributed
- Information processing is parallel

Biological Neuron



Network function f: R3 {0, 1}2



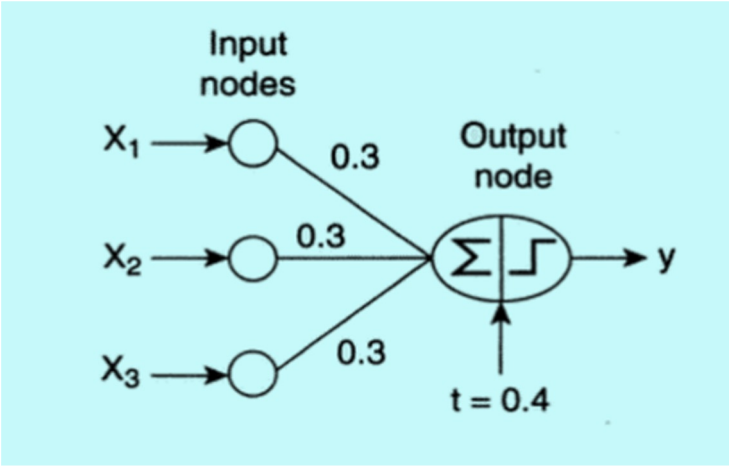
Perceptron Model

- Consider data with 3 Boolean variables X1, X2, X3 and an output variable y

Data

X1	X2	X3	Y
1	0	0	-1
1	0	1	1
1	1	0	1
1	1	1	1
0	0	0	-1
0	0	1	-1
0	1	0	-1
0	1	1	1

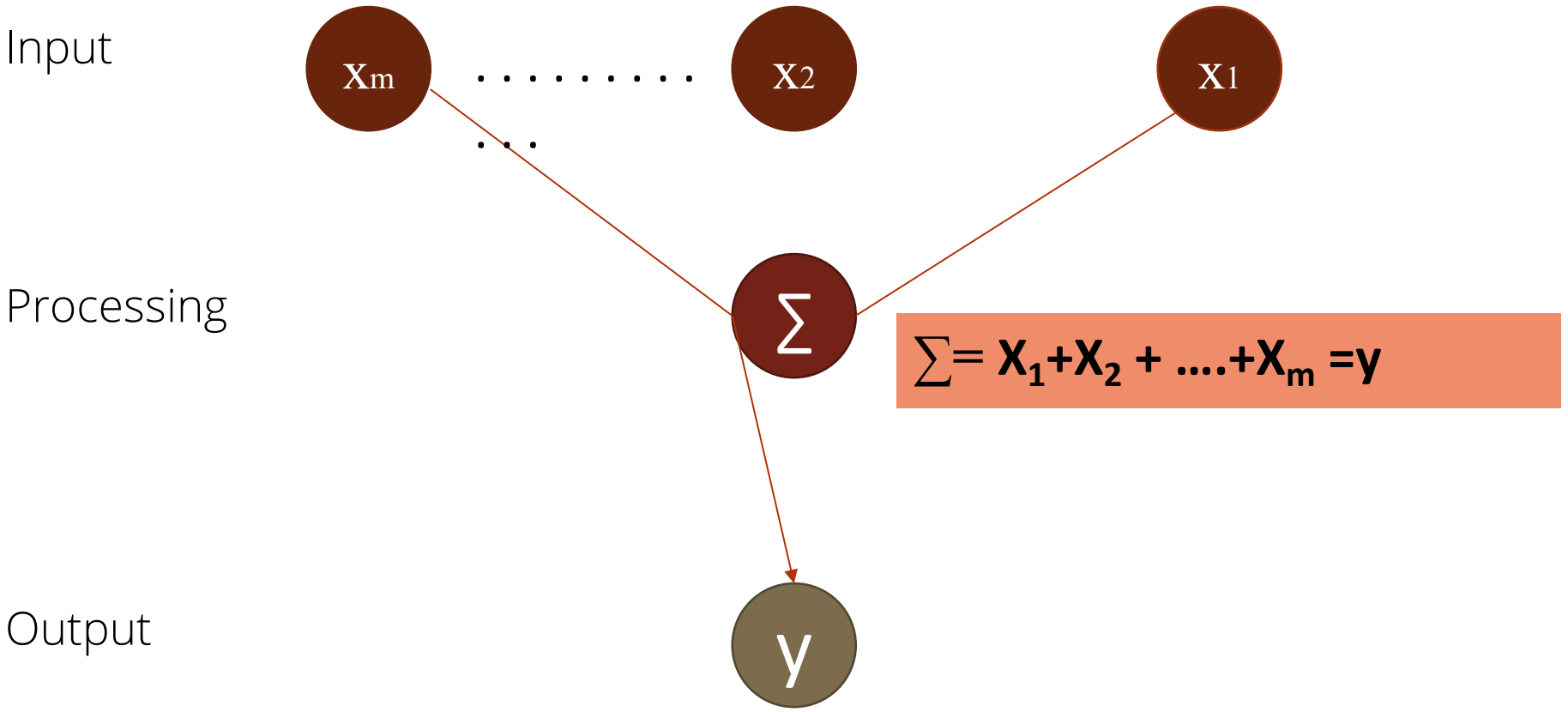
Perceptron



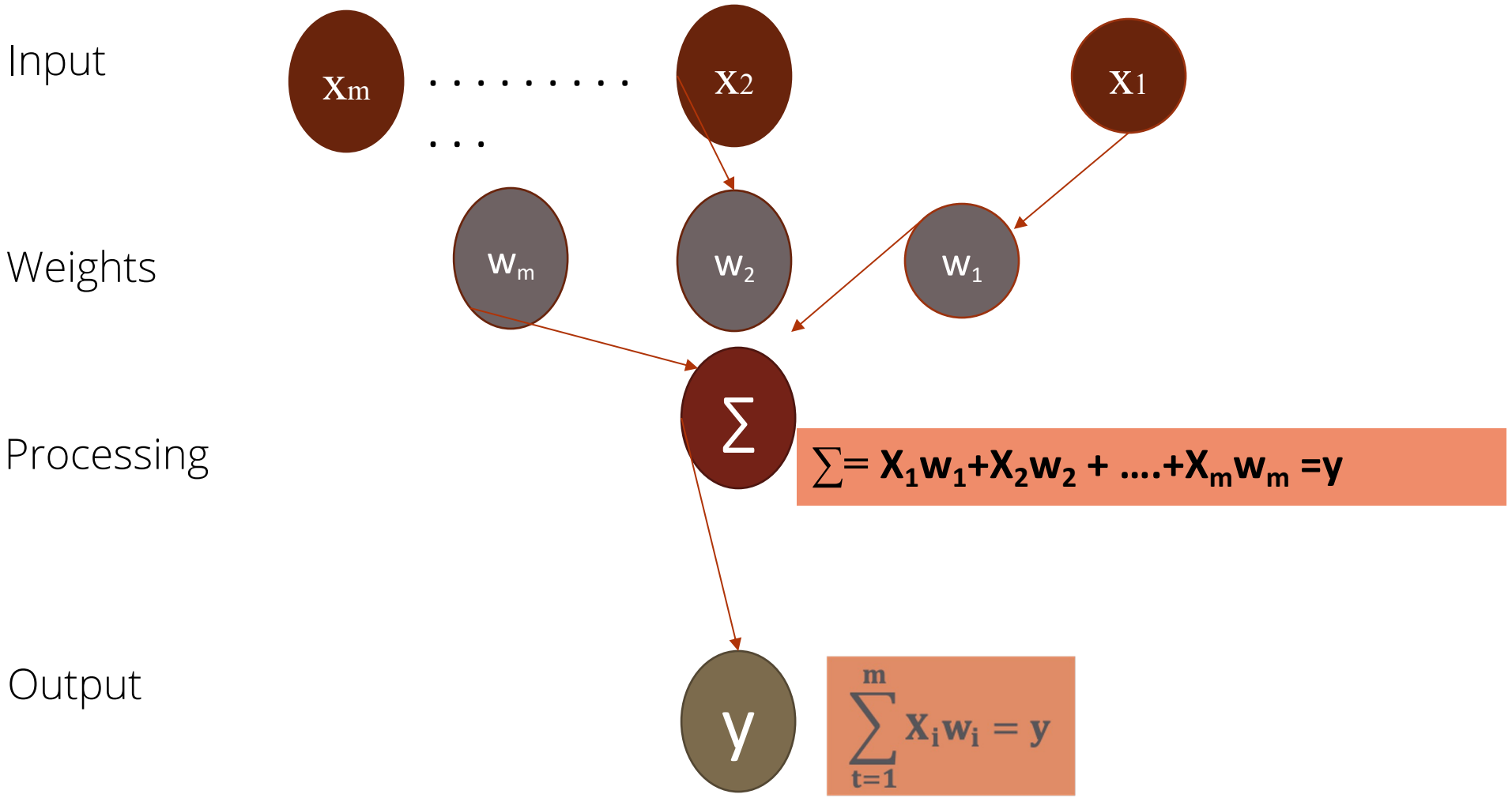
$$y \begin{cases} -1, & \text{if } 0.3\,X1 + 0.3\,X2 + 0.3\,X3 - 0.4 > 0; \\ 1, & \text{if } 0.3\,X1 + 0.3\,X2 + 0.3\,X3 - 0.4 < 0. \end{cases}$$

- Perceptron consists of two types of nodes: input nodes which are used to represent the attributes, and the output node which is used to represent the model output.
- Perceptron computes its output value, by performing a weighted sum on its inputs, subtracting a bias factor t from the sum, and then examining the sign of the result.

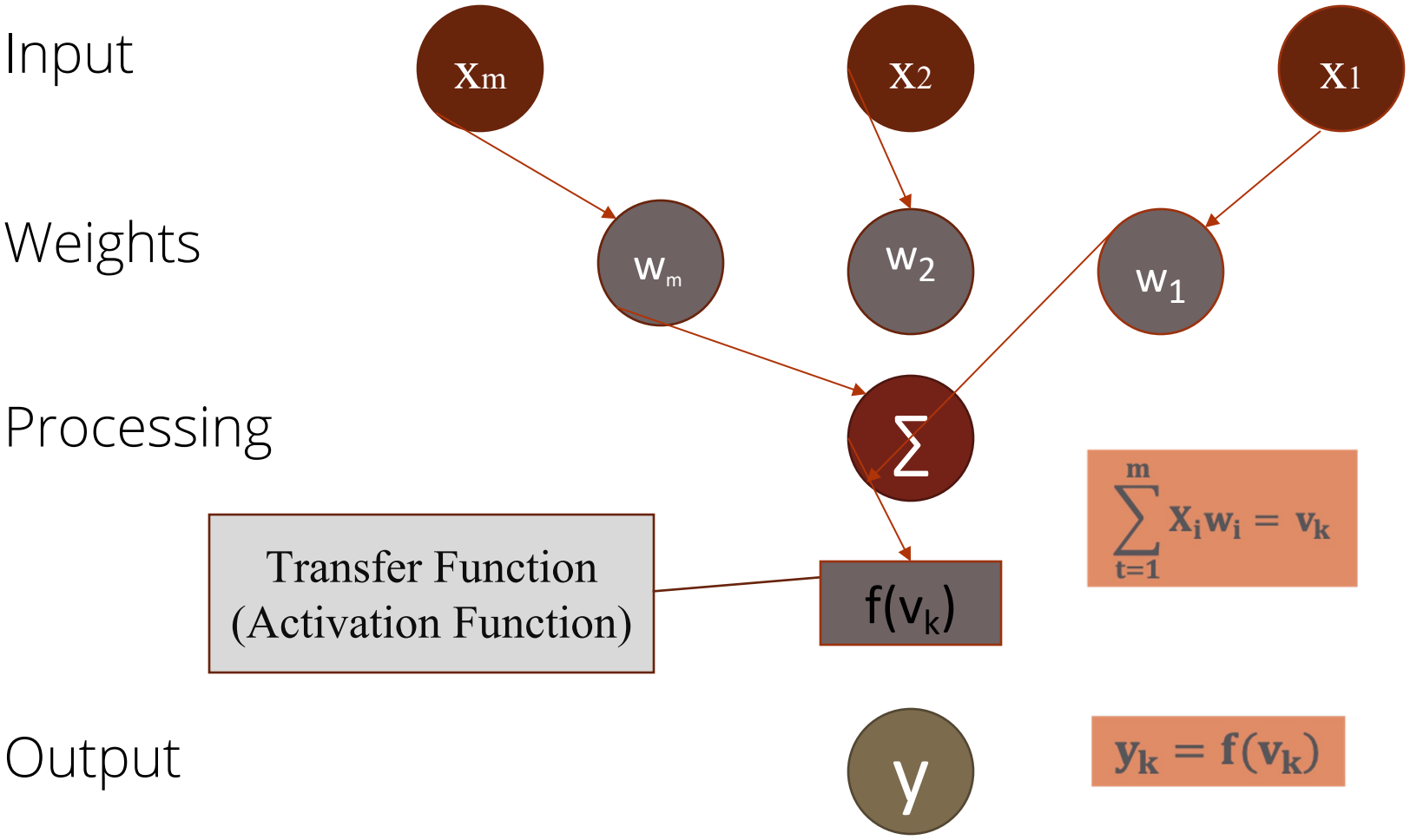
A Single Artificial Neuron…



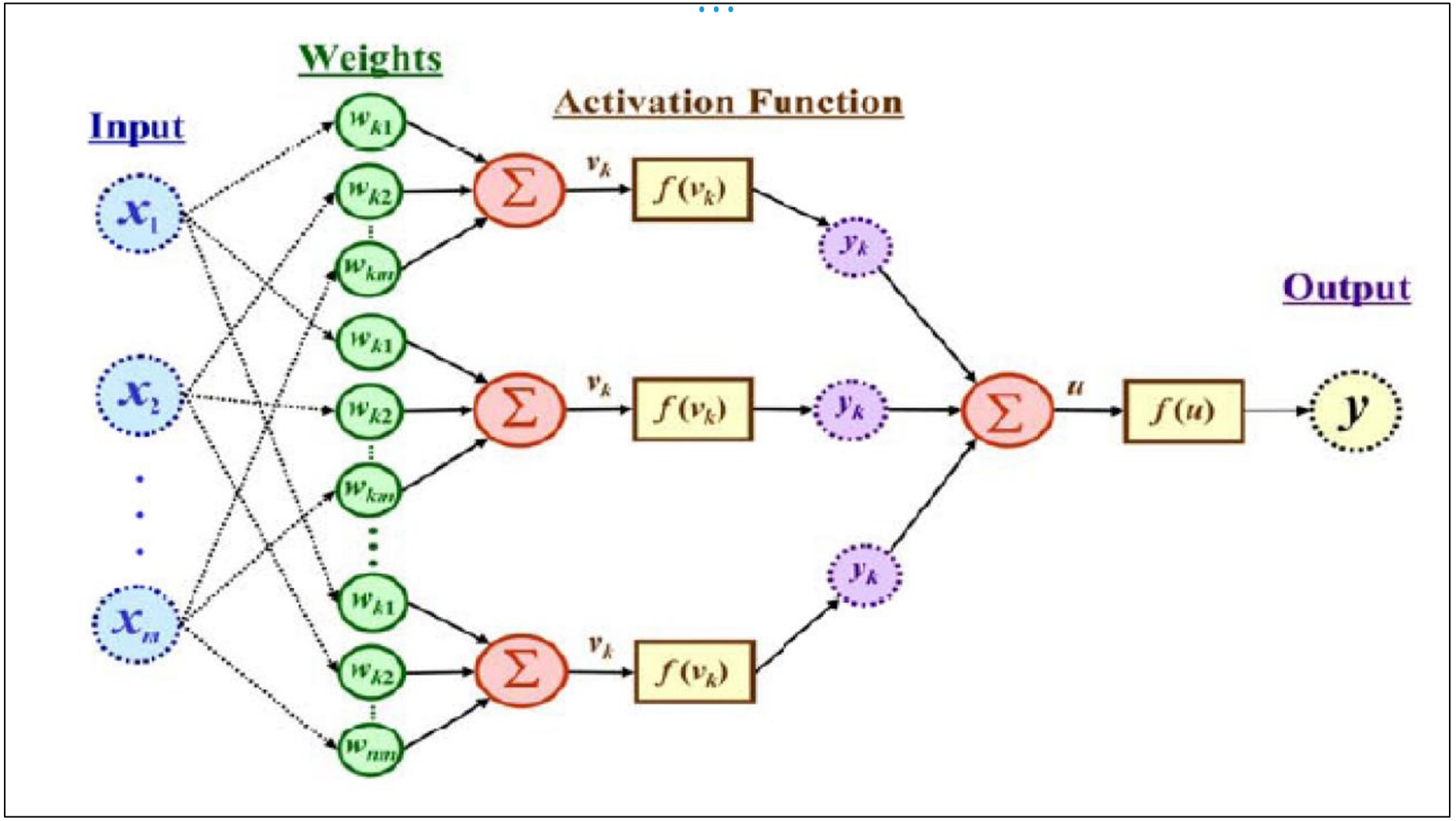
A Single Artificial Neuron…



A Single Artificial Neuron…

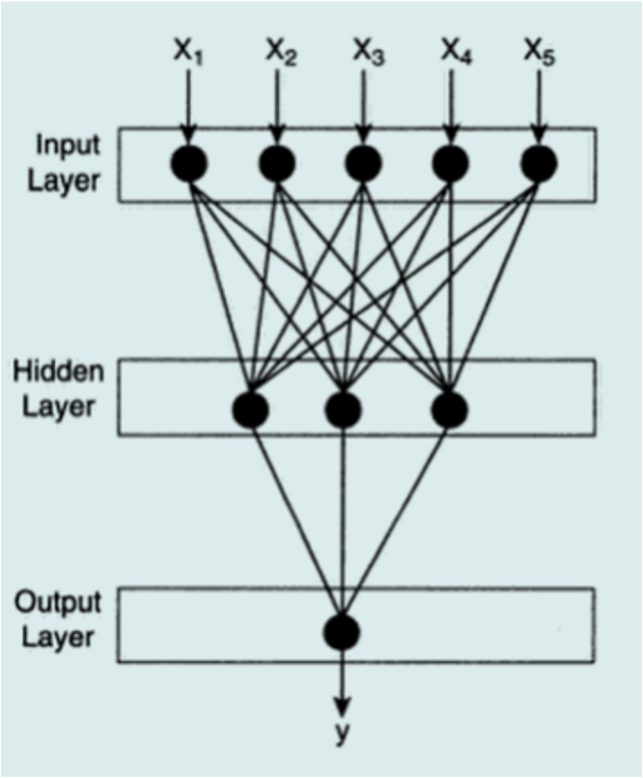


The output is a function of the input, that is affected by the weights, and the transfer functions



Artificial Neural Network (ANN)

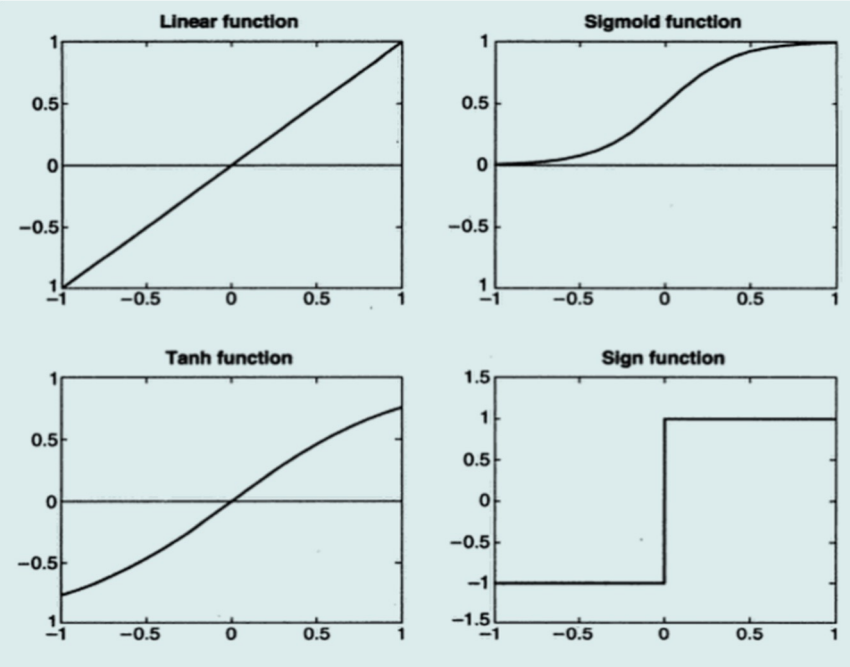
- A neural network is more complex than a perceptron model in multiple ways.
- Network may contain several intermediary layers between its input and output layers which are called hidden layers and the nodes embedded in these layers are called hidden nodes. This restructuring is known as a multilayer neural network.
- In a feed forward network, nodes in one layer are connected to nodes in the next layer.
- In a recurrent network, the links may connect nodes within the same layer or from one layer to the previous layers.



Types of ANN

- The artificial neural network may use types of activation functions other than sign function. Examples of other activation function include linear, sigmoid, and hyperbolic tangent functions.

- These additional complexities allow multilayer neural networks to model more complex relationships between the input and output variables.

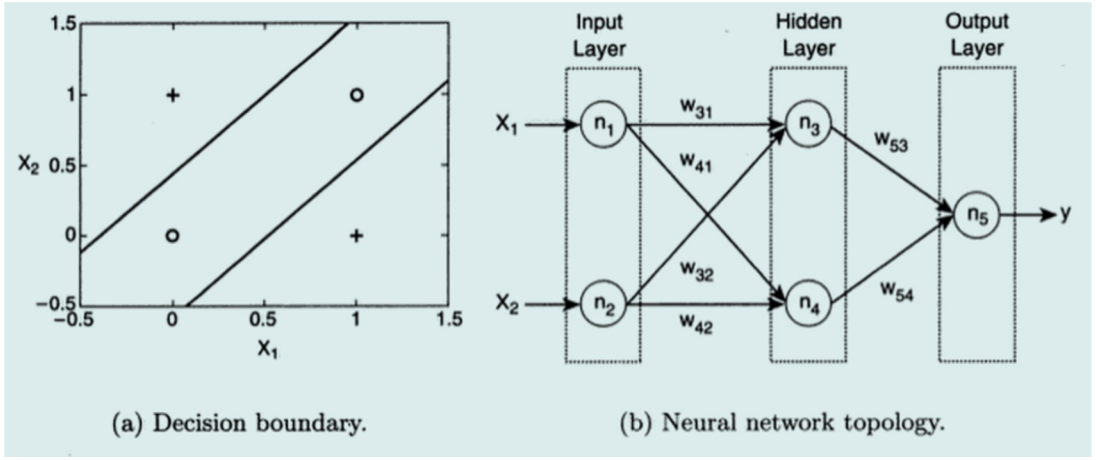


Learning ANN

- In ANN, we determine set of weights w that minimize total sum of squared errors.

$$E(w) = \frac{1}{2}\sum_{i-1}^{N}(y_i - \hat{y}_i)^2$$

- Minimum solution can be obtained by replacing $\hat{y}=w.x$ in the above equation.
- In most cases, output of ANN is non linear function of its parameters, because of the choice of its activation.
- To arrive at a global optimal solution in non linear models, we use gradient descent method to efficiently solve optimization problem.



(a) Decision boundary. (b) Neural network topology.

Gradient Descent Method

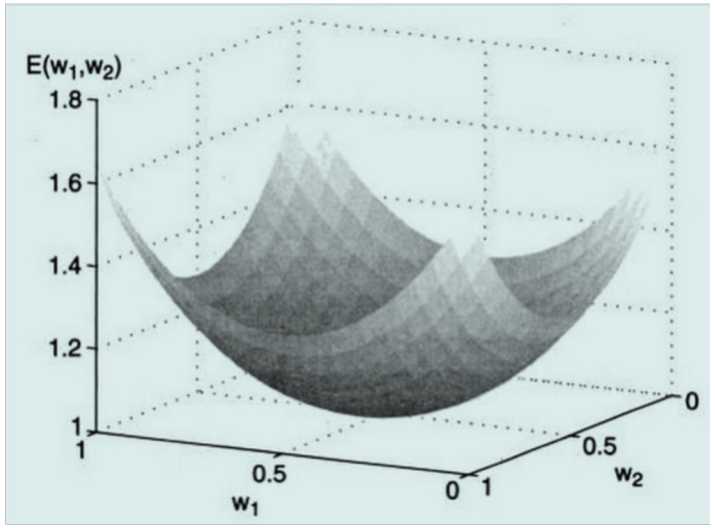
- The gradient descent algorithm can be written as:

$$w_j \leftarrow w_j - \lambda \frac{\partial E(w)}{\partial w_j}$$

- λ is the learning rate, while the second term states that the weight should be increased in a direction that reduces the overall error term

- Each w is initialized to random small value and the iterations are performed until local minima is obtained

- Converging to a local minimum can be very slow or fast depending on the size of learning rate. The weights of the output and hidden



Backpropagation

- In case of hidden nodes, computation becomes complex and error term cannot be assessed without prior information of weights. Back propagation method can be used which works in 2 phases – forward and backward.
- During forward phase, we compute output value of each neuron in the network using previous iteration weight. The computation progresses in the forward direction, outputs at the neuron k are computed prior to computing the outputs at level k+1.
- In backward phase, weight update formula is applied in the reverse direction, i.e. weights at level k+1 are updated before weights at level k.
- This backward propagation approach allows us to use errors for neurons at layer k+1 to estimate the errors for neurons at layer k.

Resilient Backpropagation

- Compared to the traditional back propagation algorithm, the Resilient propagation (rprop) algorithm offers faster convergence and is usually more capable of escaping from local minima.
- Rprop is a first-order algorithm and its time and memory requirement scales linearly with the number of parameters.
- Resilient propagation does not take into account the value of the partial derivative (error gradient), but rather considers only the sign of the error gradient to indicate the direction of the weight update.
- In practice, Rprop is easier to implement than BPNN.

Quick Recap

Neural Networks

- Artificial neural network (ANN) is a machine learning approach that models human brain and consists of a number of artificial neurons.
- In a feed forward network, nodes in one layer are connected to nodes in the next layer.
- In a recurrent network, the links may connect nodes within the same layer or from one layer to the previous layers.
- In ANN, we determine set of weights w that minimize total sum of squared errors.
- To arrive at a global optimal solution in non linear models, we use gradient descent method to efficiently solve optimization problem.
- In case of hidden nodes, we use backpropagation. The Resilient propagation (rprop) algorithm offers faster convergence than traditional backpropagation.