INTRODUCTION TO NEURAL NETWORK

DR. SANJAY CHATTERJI

Building Intelligent Machines

- Within a matter of months after birth, infants can recognize the faces of their parents, discern discrete objects from their backgrounds, and even tell apart voices.
- Within a year, they've already developed an intuition for natural physics, can track objects even when they become partially or completely blocked, and can associate sounds with specific meanings.
- And by early childhood, they have a sophisticated understanding of grammar and thousands of words in their vocabularies.

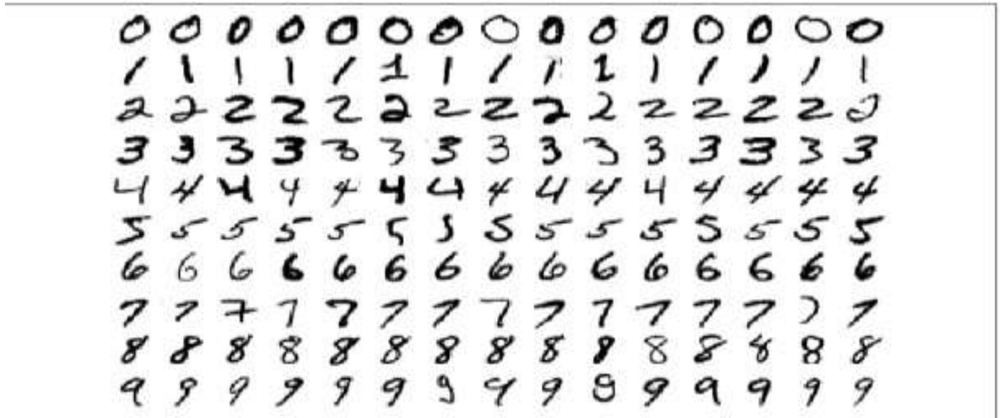
Building Intelligent Machines

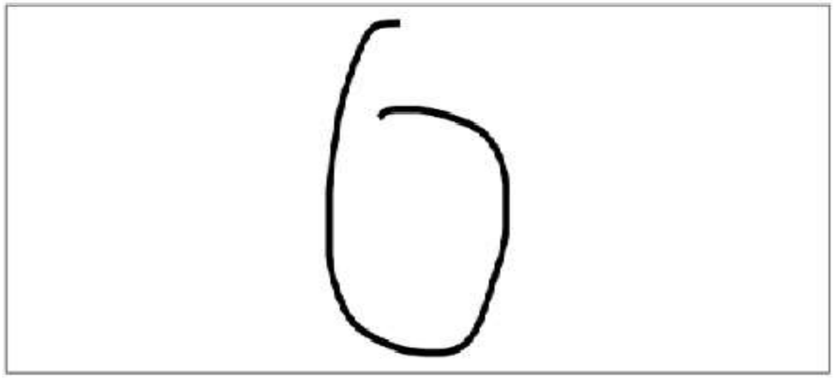
- The brain enables us to
  - store memories
  - experience emotions
  - and even dream
- For decades, Scientists dreamed of building intelligent machines with brains like human to solve problems that our brain solves in manner of microsecond
- This is an extremely active field of artificial computer intelligence often referred to as DEEP LEARNING.

Limits of Traditional Computer Programs

- Traditional computer programs are good at two things.
  1) performing arithmetic really fast
  2) explicitly following a list of instructions

- Write a program to automatically read someone's hand writing.
  - What if someone doesn't perfectly close the loop on their zero?
  - How do you distinguish a messy zero from a six?
  - We can add more and more rules, or features, through careful observation and months of trial and error.



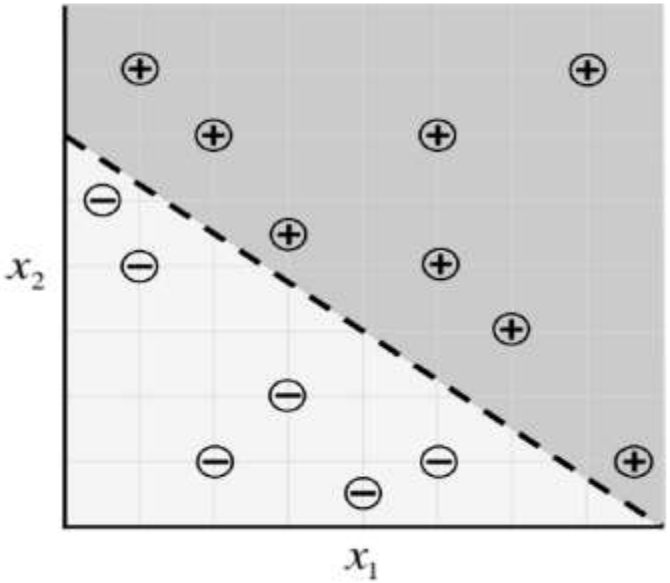


Mechanics of Machine Learning

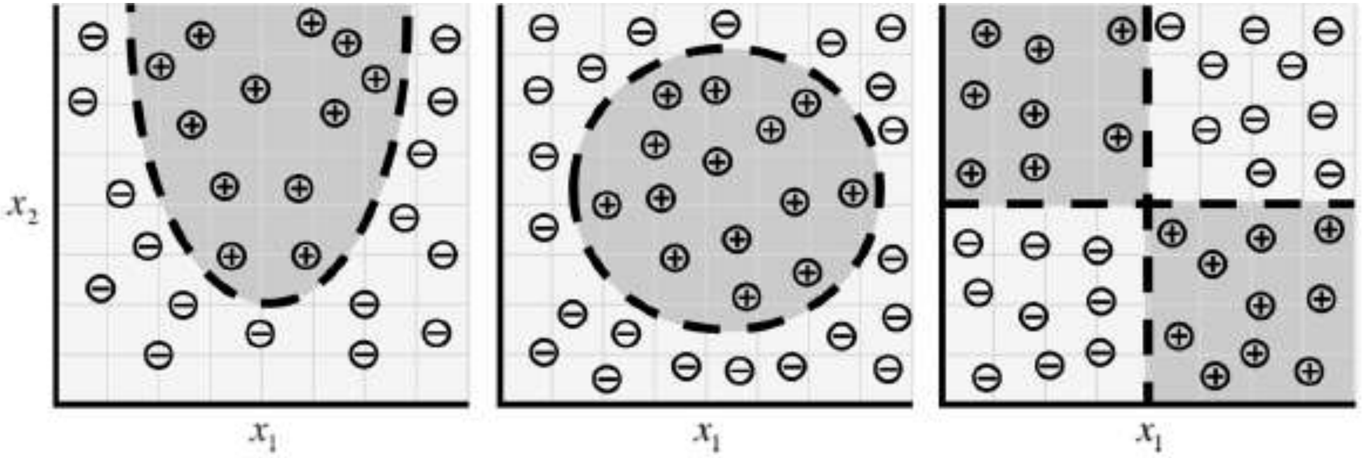
- Two years old initially didn’t recognize a dog. He/She learned to recognize a dog by being shown multiple examples.
- Our brains provided us with a model that described how to visualize the world by taking sensory inputs and make a guess.
- Machine learning that uses the concept of learning by examples, did not use massive list of rules to solve the problem.
- We give it a model with which it can evaluate examples, and a set of instructions to modify the model when it makes mistake.
- In our course we will be discussing on deep learning which is subset of a more general field of artificial intelligence called machine learning.

Mechanics of Machine Learning

$$h(\mathbf{x}, \theta) = \begin{cases} -1 & \text{if } \mathbf{x}^T \cdot \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + \theta_0 < 0 \\ 1 & \text{if } \mathbf{x}^T \cdot \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + \theta_0 \geq 0 \end{cases}$$



$$h(\mathbf{x}, \theta) = \begin{cases} -1 & \text{if } 3x_1 + 4x_2 - 24 < 0 \\ 1 & \text{if } 3x_1 + 4x_2 - 24 \geq 0 \end{cases}$$

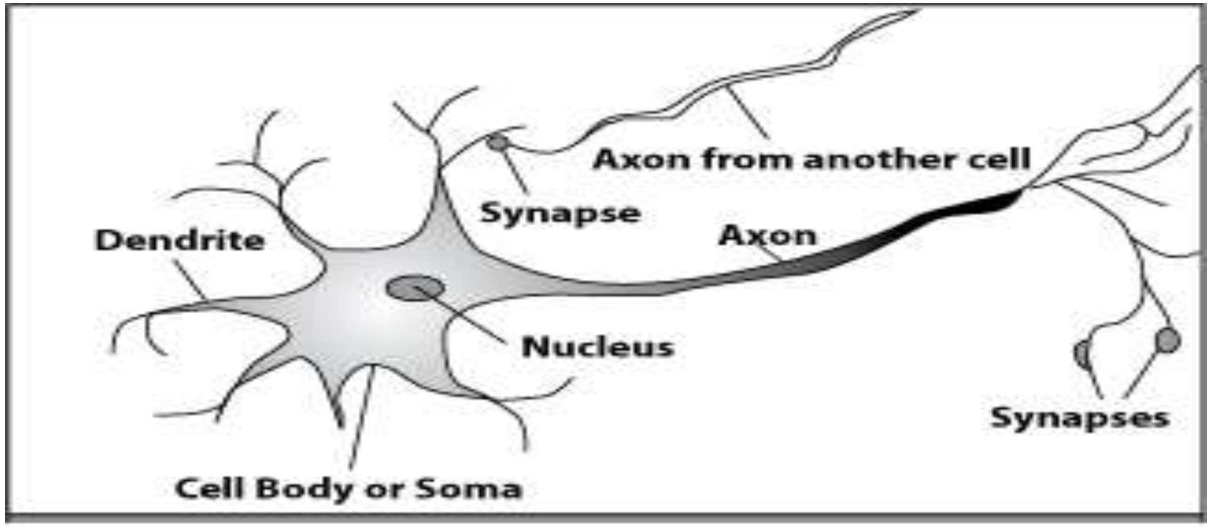


Neuron Model

- Neuron processed signals from *dendrites.*
- Sends out processed signal through an *axon*, which splits into thousands of branches.
- At end of each branch, a *synapses* transform signal into either exciting or inhibiting activity of a dendrite at another neuron.

How do our brains work?

- A processing element

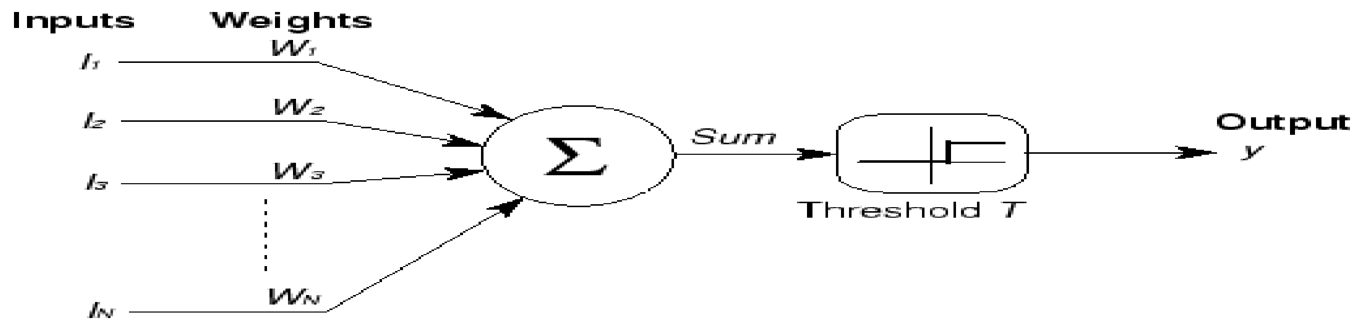


Dendrites: Input
Cell body: Processor
Synaptic: Link
Axon: Output

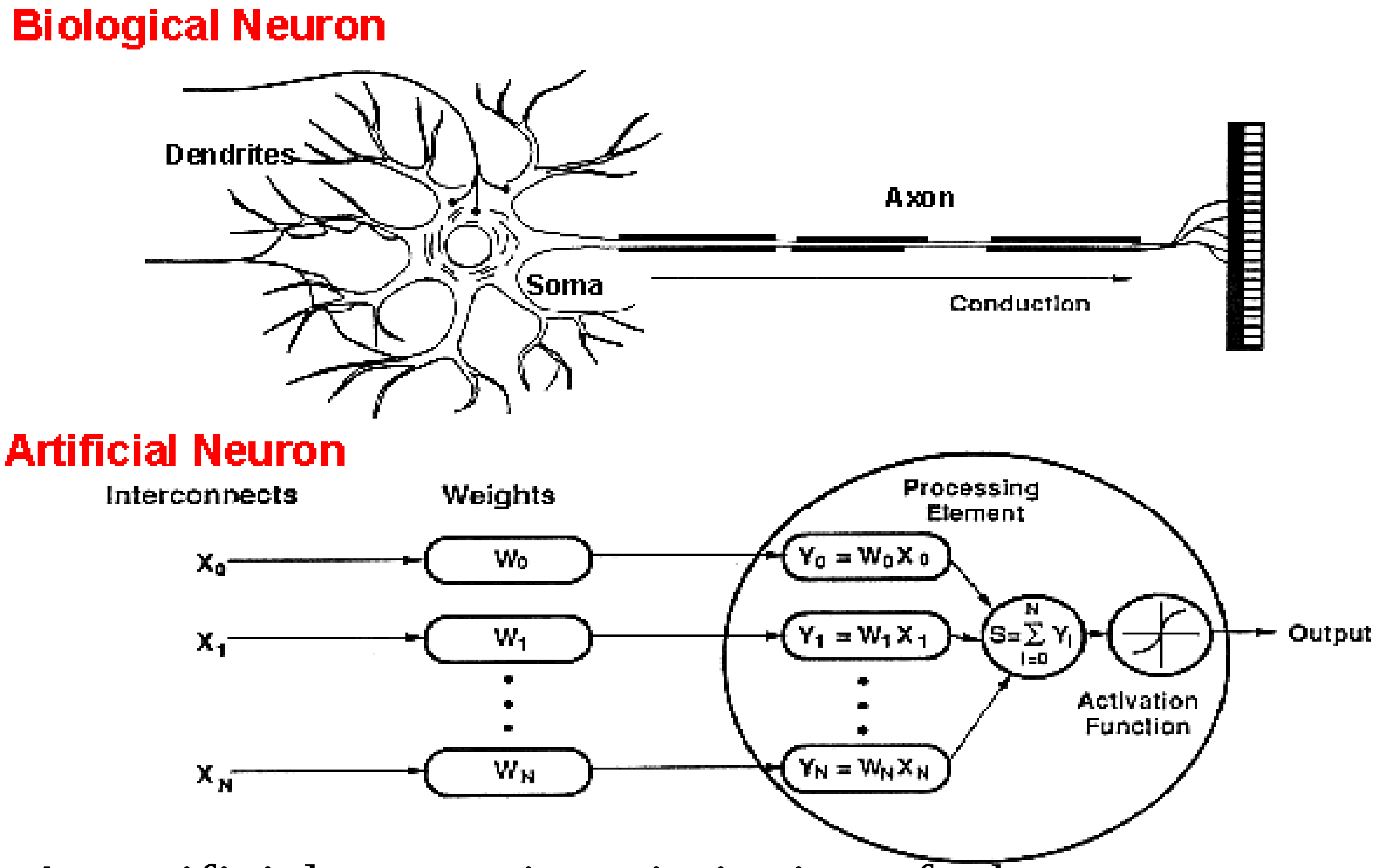
Once input exceeds a critical level, the neuron discharges a spike - an electrical pulse that travels from the body, down the axon, to the next neuron(s)

Continued..

- So we can translate this functional understanding of the neurons in our brain into an artificial model that we can represent on our computer.
- In 1943, neurophysiologist Warren McCulloch and mathematician Walter Pitts published a paper on how neurons might work. In order to demonstrate how neurons in the human brain might function, they develop a simple neural network using electrical circuits.
- Linear neuron takes in inputs, do a weighted sum and produce ‘0’ if below threshold and ‘1’ otherwise.

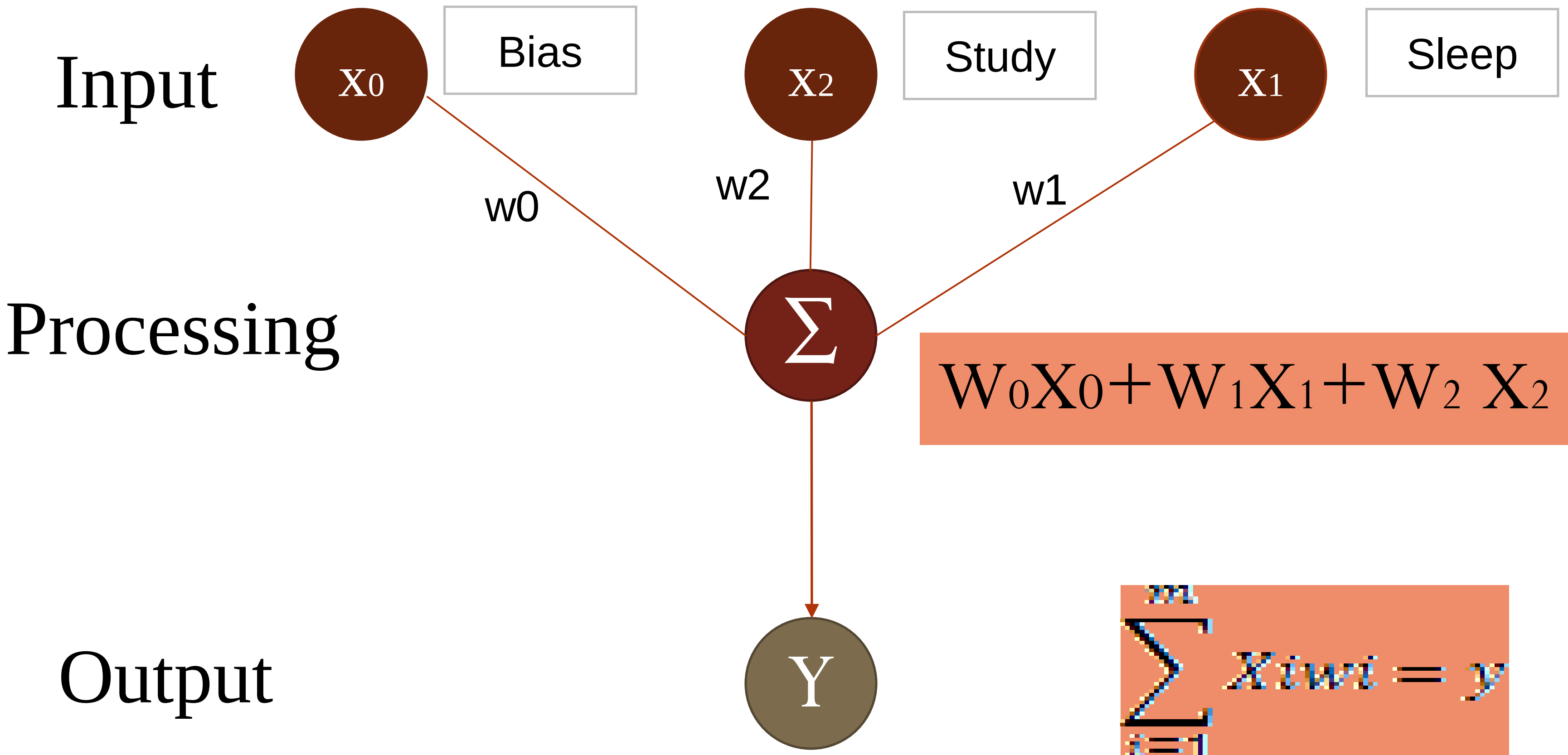


Mapping from Biological Neuron to ANN



An artificial neuron is an imitation of a human neuron

Simple ANN Model



$$W_0X_0+W_1X_1+W_2\ X_2\ =Y$$

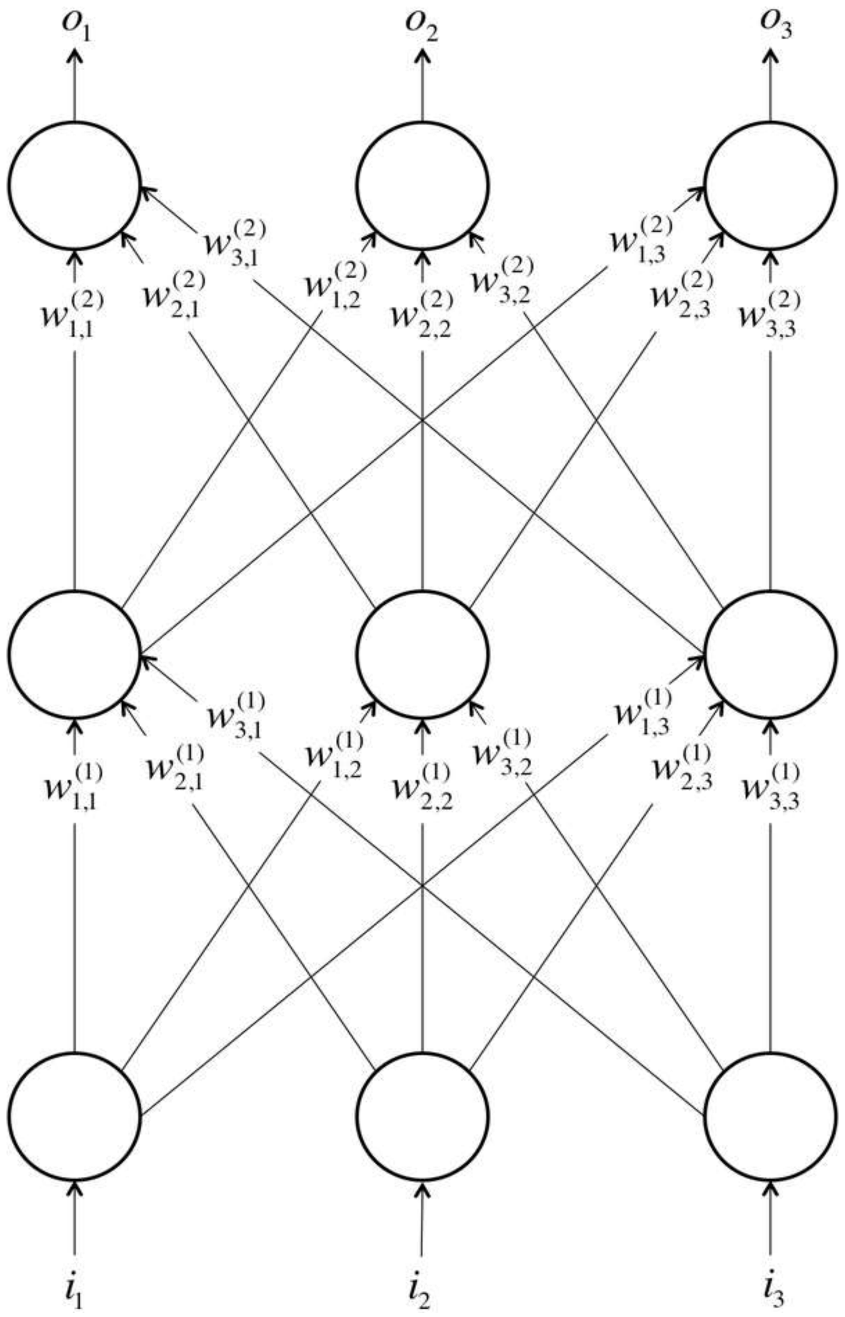
$$\sum_{i=1}^{m} Xiwi = y$$

Feed-Forward Neural Networks

- Although single neurons are more powerful than linear perceptrons, they’re not nearly expressive enough to solve complicated learning problems.
- The neurons in the human brain are organized in layers.
- The human cerebral cortex (the structure responsible for most of human intelligence) is made up of six layers.
- Information flows from one layer to another until sensory input is converted into conceptual understanding.

a feed-forward neural network

- Hidden layers identify useful features automatically.
- Connections only traverse from a lower layer to a higher layer.
- They are the simplest to analyze.
- Hidden layers have fewer neurons than input layer.
- Selecting which neurons to connect to which neurons in the next layer is an art that comes from experience.
- The inputs and outputs are *vectorized* representation.

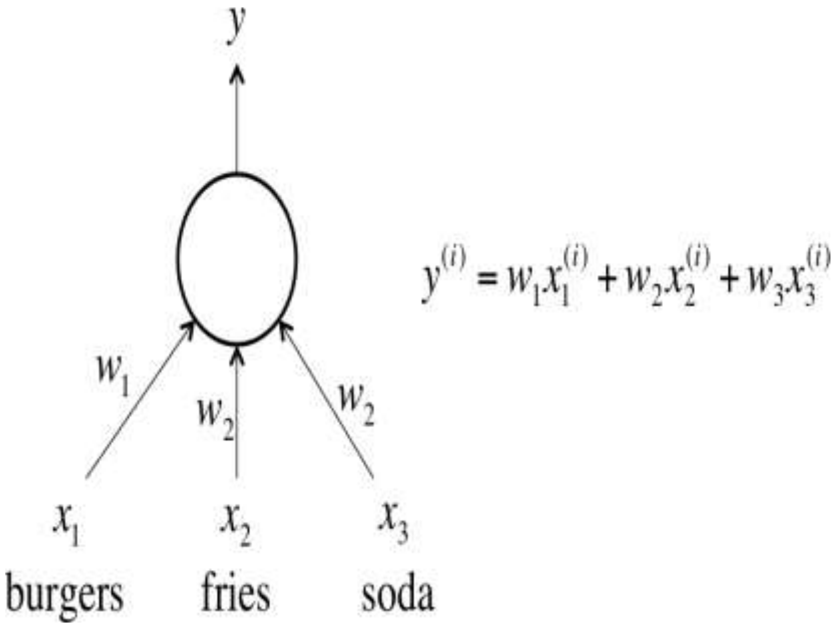


Expressing neural network as a series of vector and matrix operations

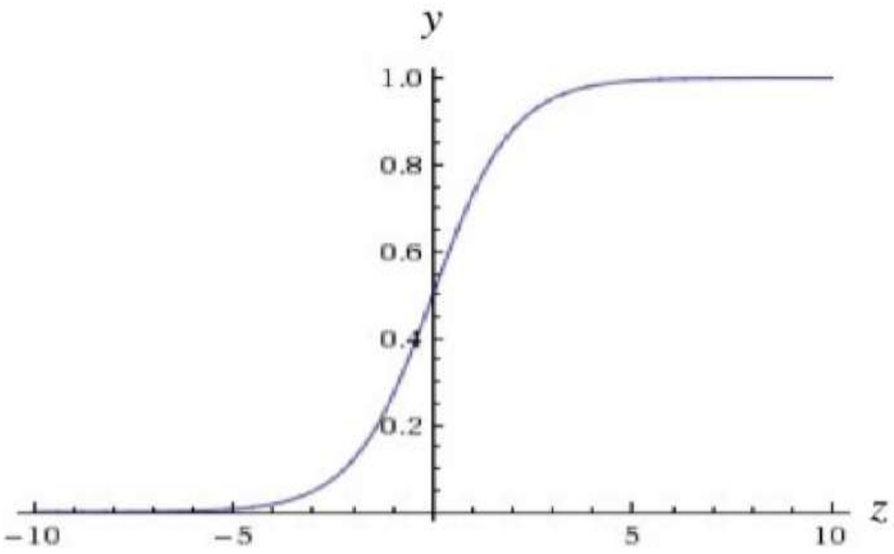
- input to the *ith* layer of the network **x** = [x1 x2 ... xn]
- vector produced by propagating the input through the neurons **y** = [y1 y2 ... ym]
- weight matrix of size *n* × *m* and a bias vector of size *m*.
- *jth* element of a column corresponds to the weight of the connection pulling in the *jth* element of the input.
- **y**= *ƒ*(**WTx** + **b**) (the transformation function) is applied to the vector elementwise.
- This reformulation will become all the more critical as we begin to implement these networks in software.

Linear Neurons and Their Limitations

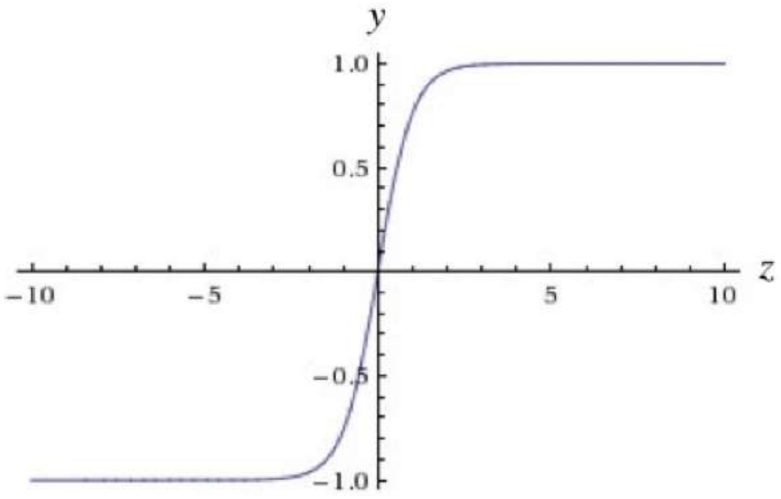
- Linear neurons are easy to compute with, but they run into serious limitations.
- A feed-forward neural network consisting of only linear neurons can be expressed as a network with no hidden layers.
- In order to learn complex relationships, we need to use neurons that employ some sort of nonlinearity.



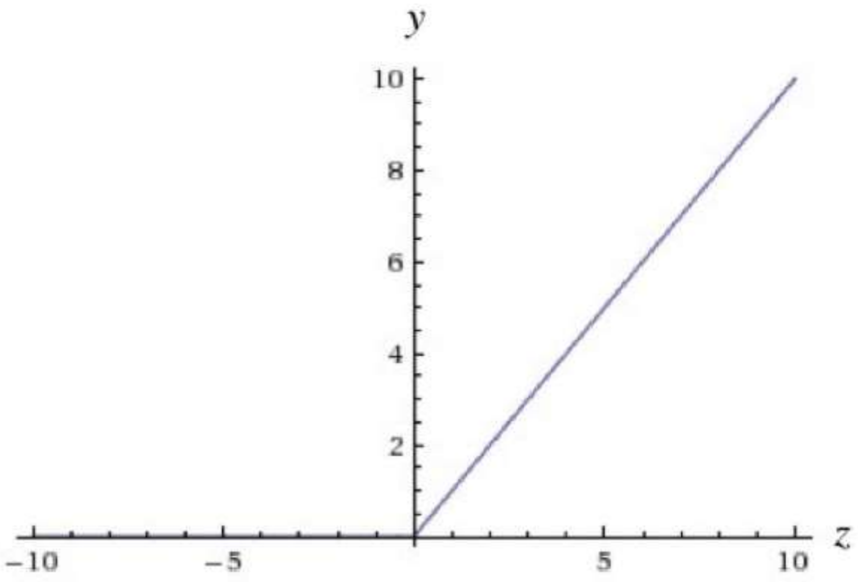
Sigmoid, Tanh, and ReLU Neurons



$$f(z) = \frac{1}{1 + e^{-z}}$$



$$f(z) = \tanh(z)$$



$$f(z) = \max(0, z)$$

Softmax Output Layers

- Oftentimes, we want our output vector to be a probability distribution over a set of mutually exclusive labels.
- For example, let's say we want to build a neural network to recognize handwritten digits. $[p_0 \; p_1 \; p_2 \; p_3 \; \cdots \; p_9]$ where $\sum_{i=0}^{9} p_i = 1$
- This is achieved by using a special output layer called a *softmax layer*.
- The output of a neuron in a softmax layer depends on the outputs of all the other neurons in its layer.

Softmax Output Layers

- Letting z_i be the logit of the i^{th} softmax neuron, we can achieve this normalization by setting its output to:

$$y_i = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

- A strong prediction would have a single entry in the vector close to 1, while the remaining entries were close to 0.
- A weak prediction would have multiple possible labels that are more or less equally likely.

Looking Forward

- Here we’ve talked about the basic structure of a neuron, how feed-forward neural networks work, and the importance of nonlinearity in tackling complex learning problems.
- Next chapter, we will build the mathematical background necessary to train a neural network to solve problems.
- Specifically, we will talk about finding optimal parameter vectors, best practices while training neural networks, and major challenges.

Thank You