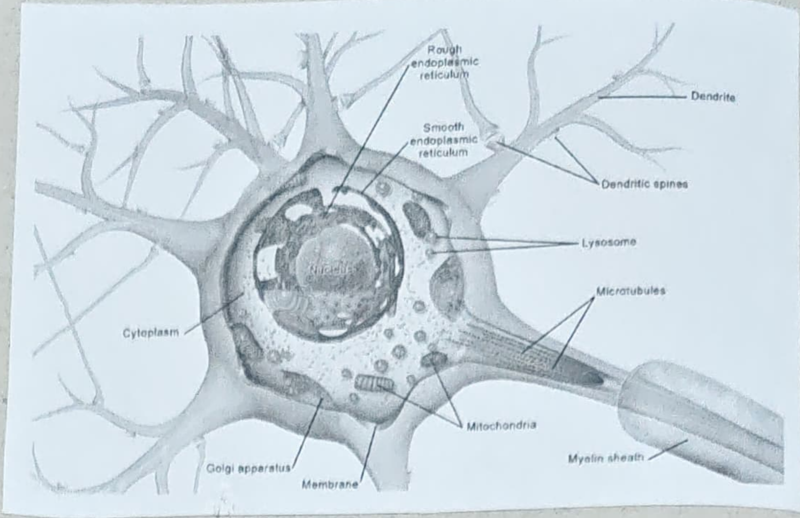
10. Introduction to Neural Networks

Neurons

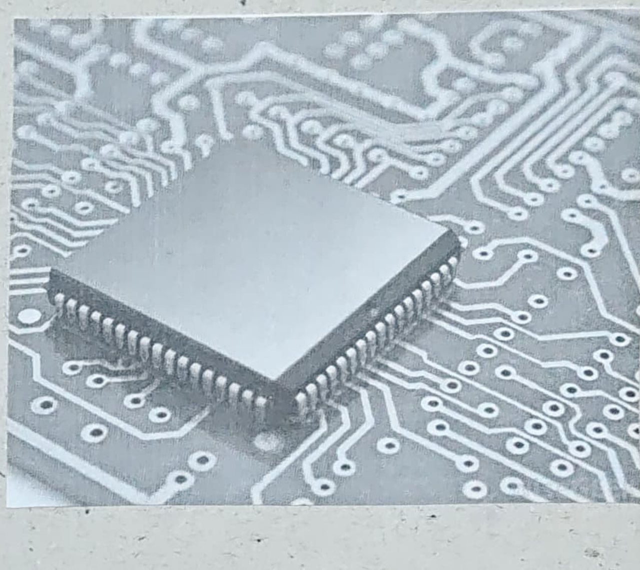
- Neurons are structural constituents of the human brain.



> **Note:**
> Size of a typical neuron: 4 - 100 μm

- Neurons are much slower than Silicon logic gates!

> **Note:**
> What are the durations of operations on a Silicon chip and neurons?

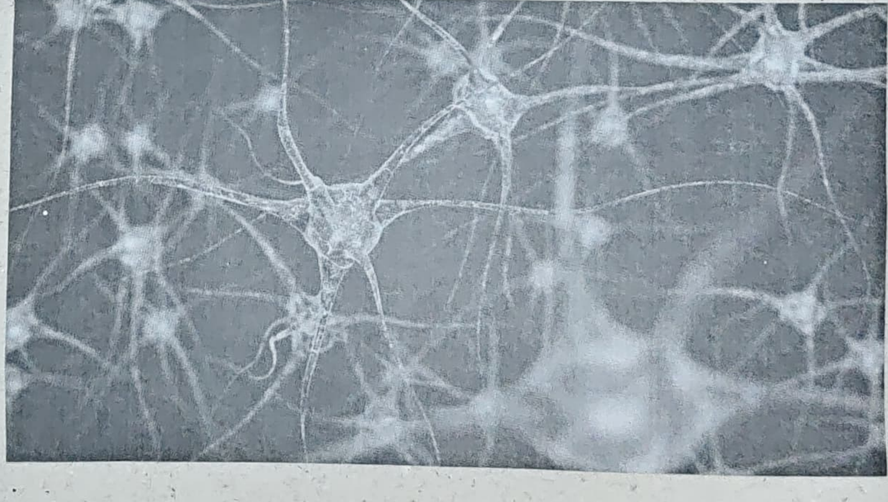


- Operations on a Silicon chip occur over nano seconds (ns) 10^{-9} duration, whereas Neural events span milliseconds (ms) 10^{-3}.

(i.e) Million times higher

Brain

The brain overcomes this relatively slow rate of operation of a neuron, by having an enormous number of neurons, with massive interconnections between them, termed as - Synapses.

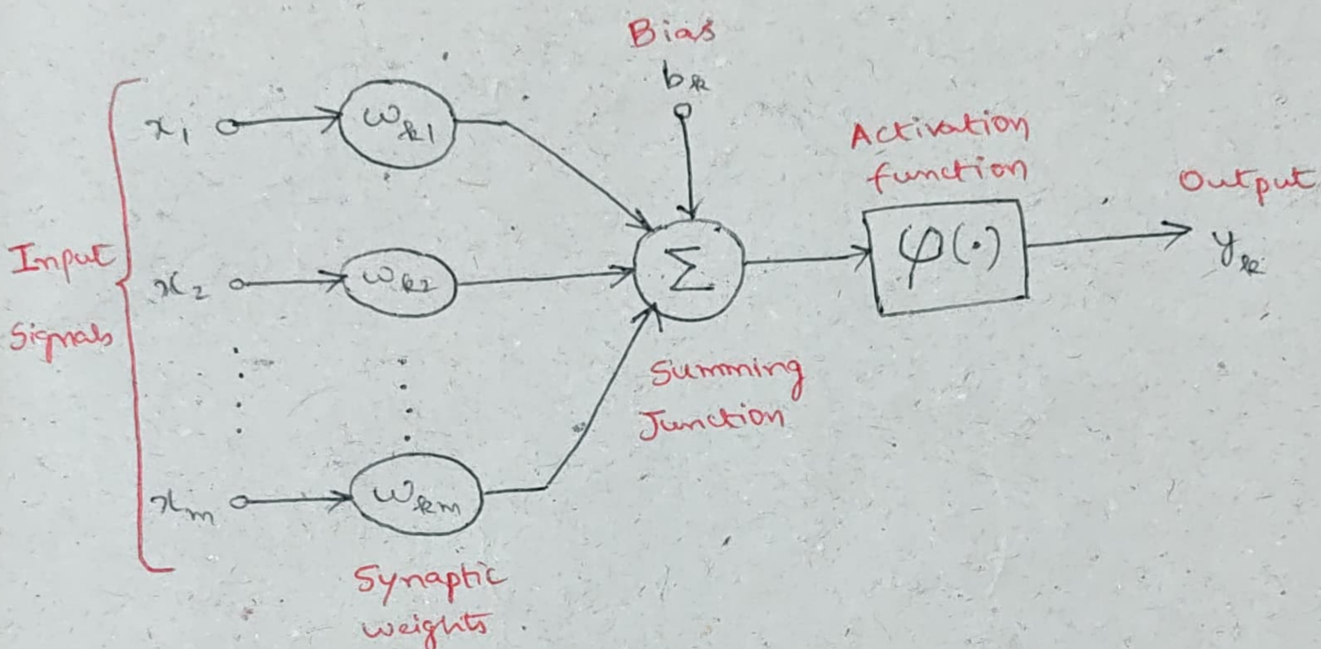


Note: Approximatily how many neurons are there in the human brain?
What is the number of synapses?

- There are approximately 10 billion neurons in the human brain and 60 trillion synapses or connections!
This makes the brain an enormously efficient structure!

Model of a neural net.

⊙ A neuron is an basic data processing unit, which is a fundamental building block of a Neural network.



⊙ A neuron has 3 basic components.

- Synapses
- Adder
- Activation function

(i) Synapses :

These are connecting links, each characterized by a weight.

(ii) Adder :

Sums the weighted input signals that are input to the synapses. This is therefore LINEAR.

(iii) Activation function :

Activation function limits the amplitude of the output of a neuron. Also referred to as a squashing function.

i/p range : $-\infty$ to ∞

o/p range : -1 to 1

- Output of the NN

The output of the NN can be modeled as:

$$O_k = \psi(v_k)$$

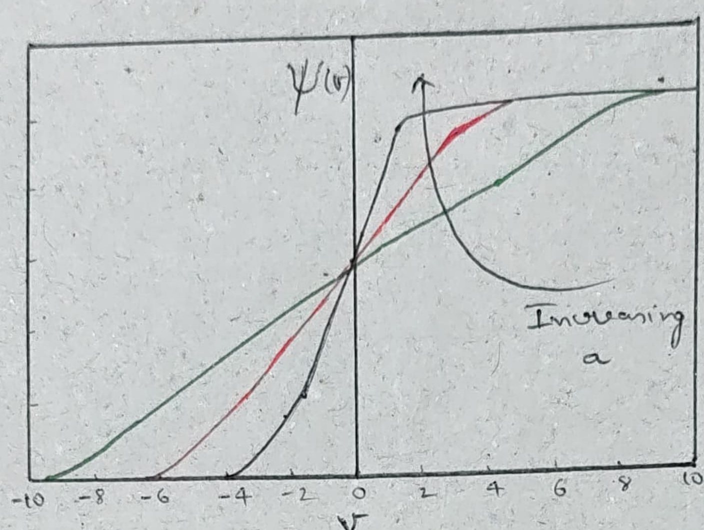
$$= \psi\left(\sum_{j=1}^{m} w_{kj}\, x_j + b_k\right)$$

Sigmoid activation

- Most popular activation function used in NN.
- Exhits a balance between linear and nonlinear behavior.

$$\psi(v) = \frac{1}{1 + e^{-av}}$$

controls the slope.



- a is the slope parameter of the Sigmoid
- Note that, the sigmoid function is differentiable, which is an important feature in neural networks.

Backpropagation

- Backpropagation is a widely used algorithm for training neural networks.
- Backpropagation employs the gradient of the loss function w.r.t the weight vector.
- Gradient descent is an iterative optimization algorithm for finding a local minimum of a differentiable function.
- $f(\bar{x})$ decreases fastest at $\bar{x}$, when one travels in the direction of negative gradient of $f(\bar{x})$.

$$\bar{x}_{n+1} = \bar{x}_n - \gamma_n \nabla f(\bar{x}_n)$$

Problem:

What is the gradient of

$$f(\bar{\omega}^T \bar{x}) = \cos(\bar{\omega}^T \bar{x}) = \cos\left(\sum_{i=1}^{n} w_i x_i\right).$$

The gradient is

$$f'(\bar{\omega}^T \bar{x})(\nabla \bar{\omega}^T \bar{x}) = \sin(\bar{\omega}^T \bar{x}) \begin{bmatrix} \frac{\partial}{\partial x_1} \sum_{i=1}^{n} w_i x_i \\ \frac{\partial}{\partial x_2} \sum_{i=1}^{n} w_i x_i \\ \vdots \\ \frac{\partial}{\partial x_n} \sum_{i=1}^{n} w_i x_i \end{bmatrix}$$

$$= -\sin(\bar{\omega}^T \bar{x}) \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

$$= -\sin(\bar{\omega}^T \bar{x})\, \bar{\omega}$$

$$\frac{\partial}{\partial w_i} \cos\left(\sum_{i=1}^{n} w_i x_i\right) = -x_i \sin\left(\sum_{i=1}^{n} w_i x_i\right)$$

$$\nabla \cos\left(\sum_{i=1}^{n} w_i x_i\right) = \begin{bmatrix} -x_1 \sin\left(\sum_{i=1}^{n} w_i x_i\right) \\ -x_2 \sin\left(\sum_{i=1}^{n} w_i x_i\right) \\ \vdots \\ -x_n \sin\left(\sum_{i=1}^{n} w_i x_i\right) \end{bmatrix}$$

$$= -\sin\left(\sum_{i=1}^{n} w_i x_i\right) \bar{x}$$

Loss function

- The Loss function can be defined as

$$L(\bar{w}_k) = \frac{1}{2}\left(y_k(n) - \psi\left(\underbrace{\sum_{j=1}^{m} w_{kj} x_j(n) + b_k}_{v_k(n)}\right)\right)^2$$

- To minimize the loss, update the weights as

$$\bar{w}_k(n) = \bar{w}_k(n-1) - \eta_k \nabla L(\bar{w}_k) \Big|_{\bar{w}_k = \bar{w}_k(n-1)}$$

- The quantity $\nabla L(\bar{w}_k)$ is given as

$$\nabla L(\bar{w}_k) = -\left(y - \psi(v_k(n))\right) \psi'(v_k(n)) \bar{x}$$

- Therefore, the update rule is given as

$$\bar{w}_k(n) = \bar{w}_k(n-1) + \eta_k e_k(n) \psi'(v_k(n)) \bar{x}$$

$$e_k(n) = y_k - \psi(v_k(n)) \Big|_{\bar{w}_k = \bar{w}_k(n-1)}$$
