NEURAL NETWORKS

Name - Shivam Malviya
Roll No. - 18EC01044

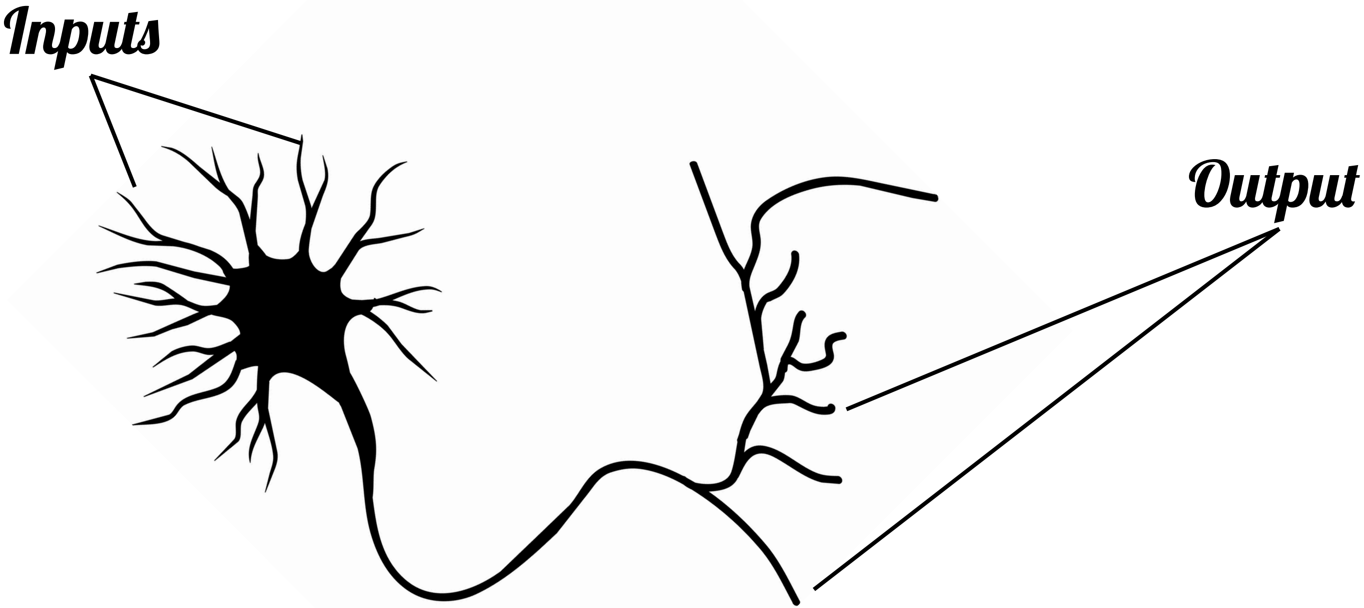
Table Of Content

Inspiration

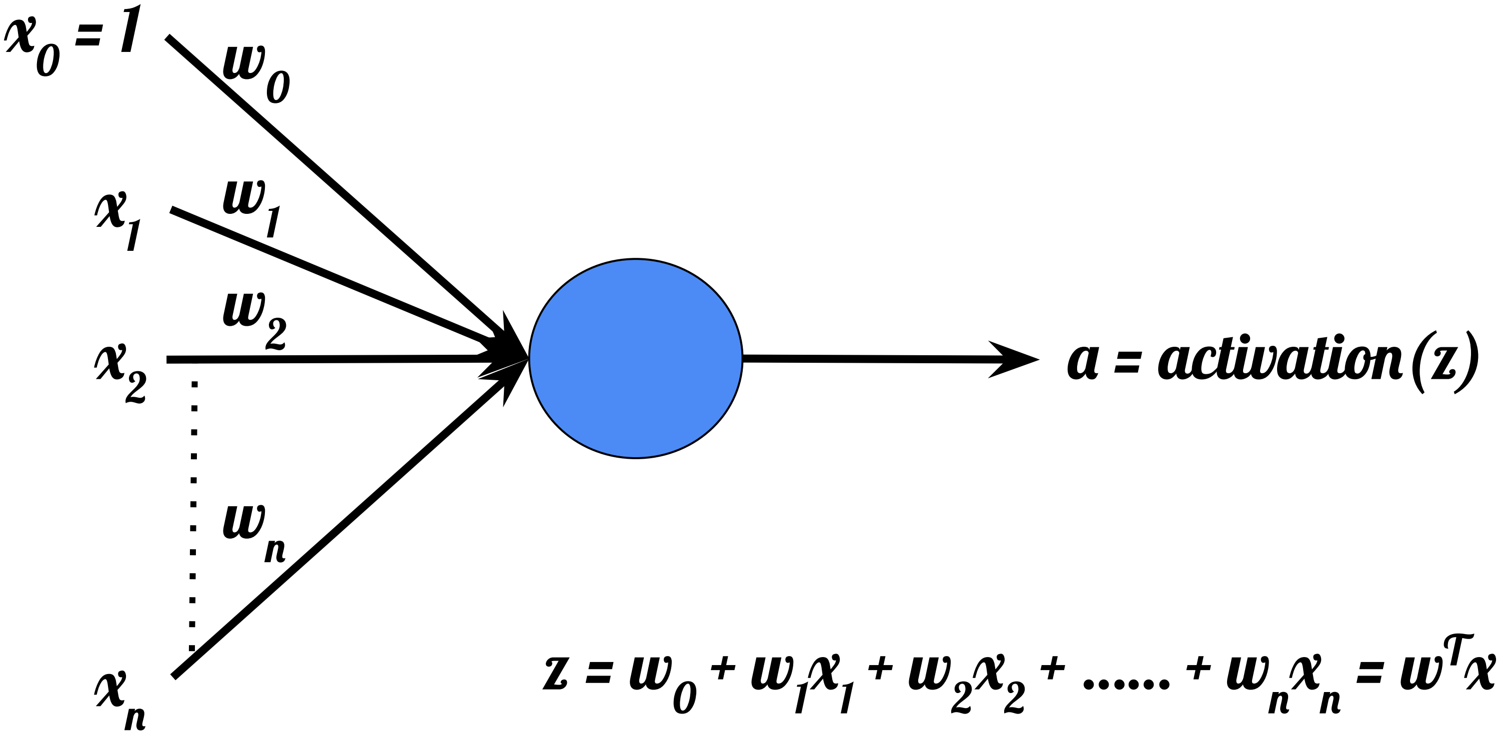
In 1943, Warren McCulloch, a neurophysiologist, and a young mathematician, Walter Pitts, wrote a paper on how neurons might work.

The first artificial neural network was invented in 1958 by psychologist Frank Rosenblatt, it is called Perceptron.

Neuron

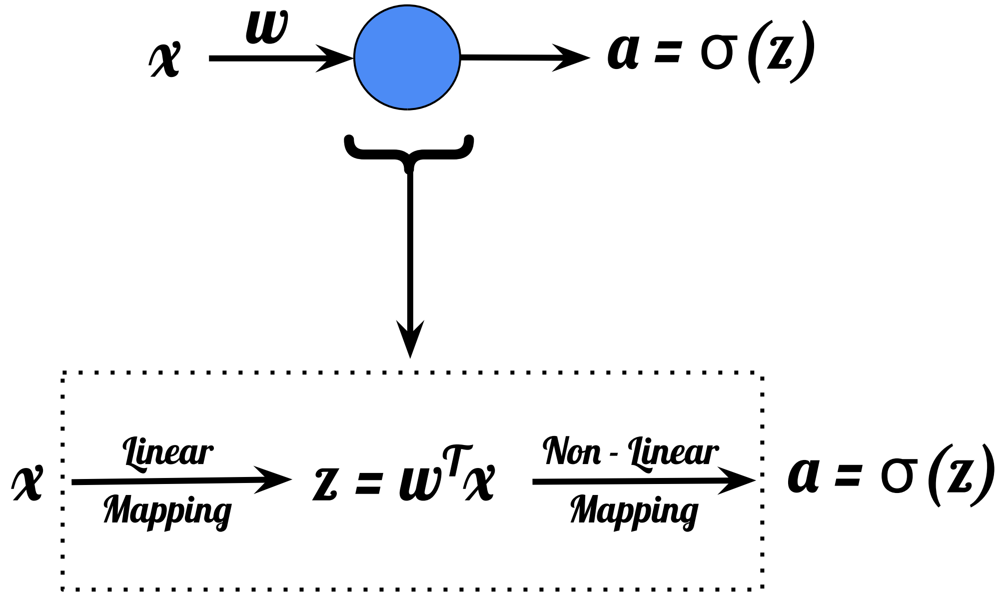


Neuron



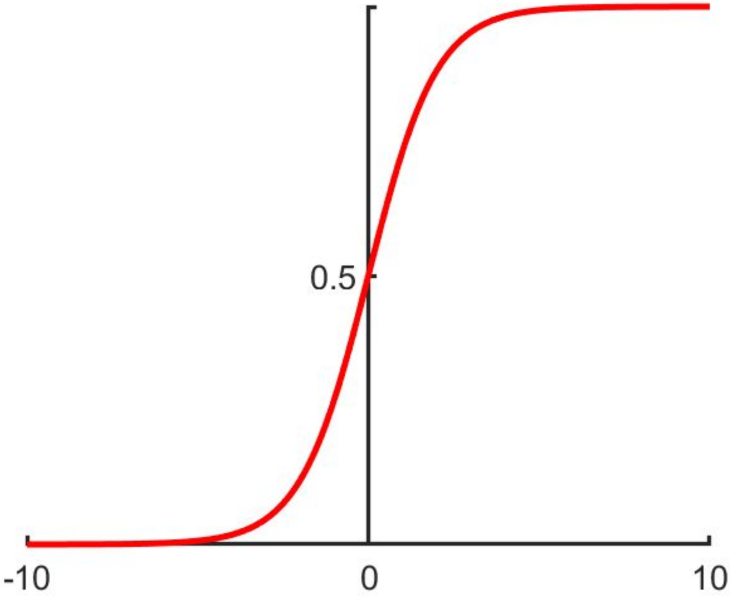
$$z = w_0 + w_1x_1 + w_2x_2 + \ldots\ldots + w_nx_n = w^Tx$$

Neuron

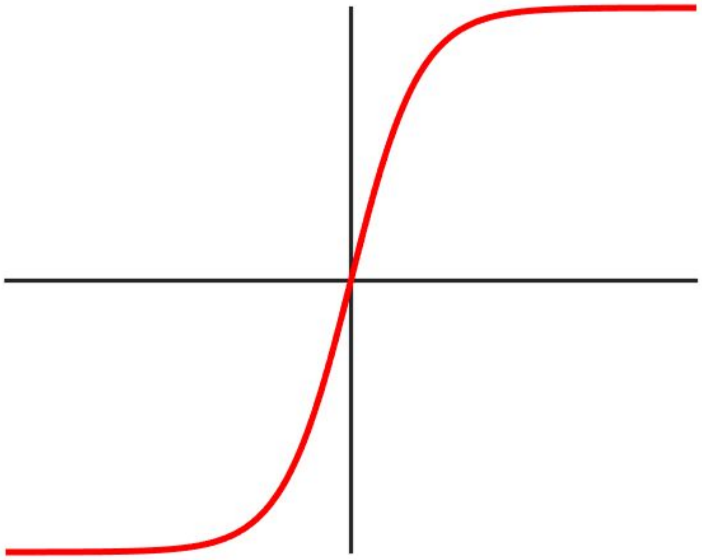


Activation Functions

Sigmoid

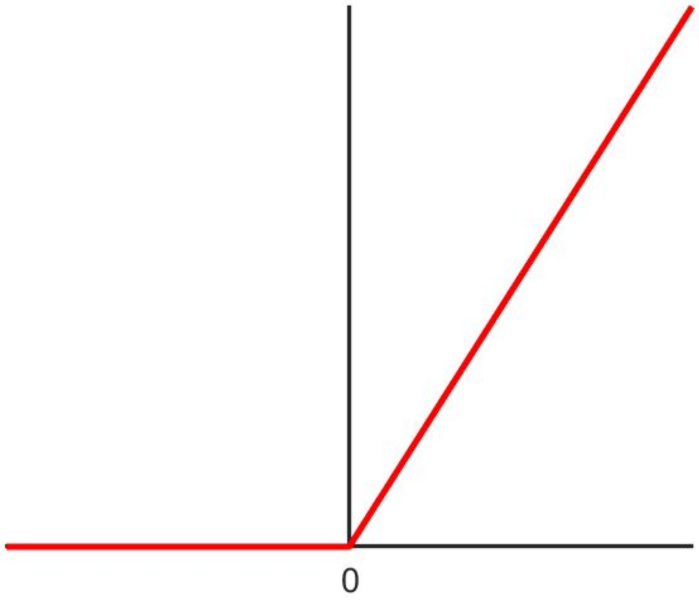


Tanh

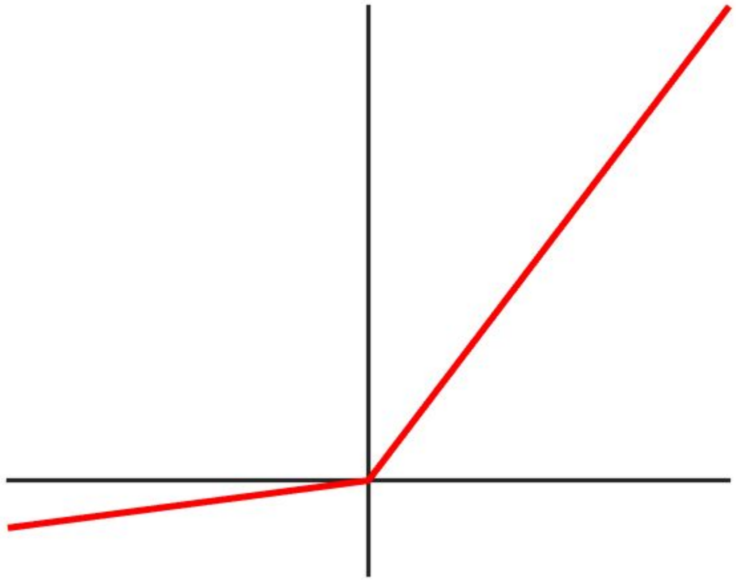


Activation Functions

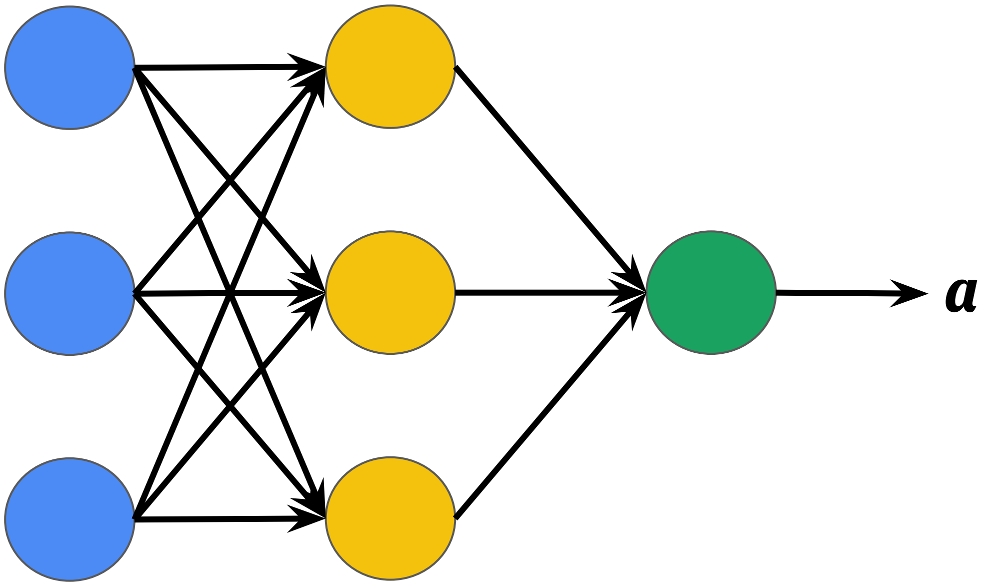
ReLU - Rectified Linear Unit



Leaky ReLU



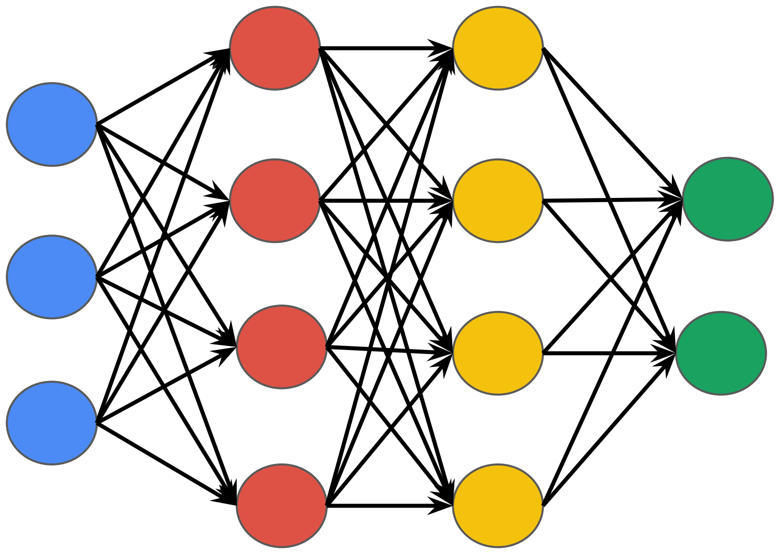
Neural Network



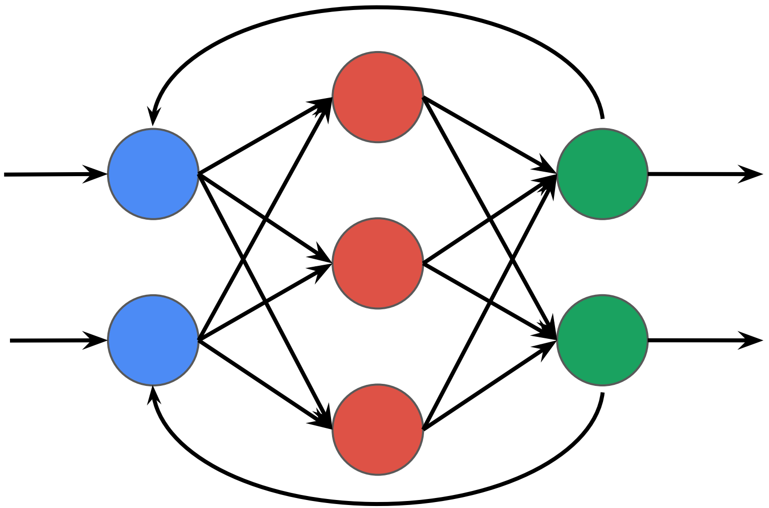
Classification

Based on connection patterns

Feedforward NN



Feedback NN



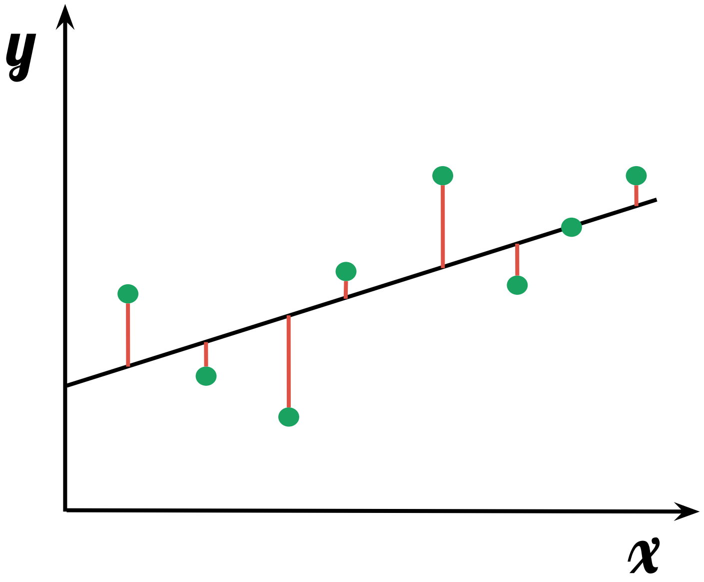
Training

Loss Function - The output of loss function tells us how well our neural network model the given dataset.

Loss Function in Regression

$$\mathcal{L} = \frac{\sum (errors)^2}{m}$$

m = No. of data points

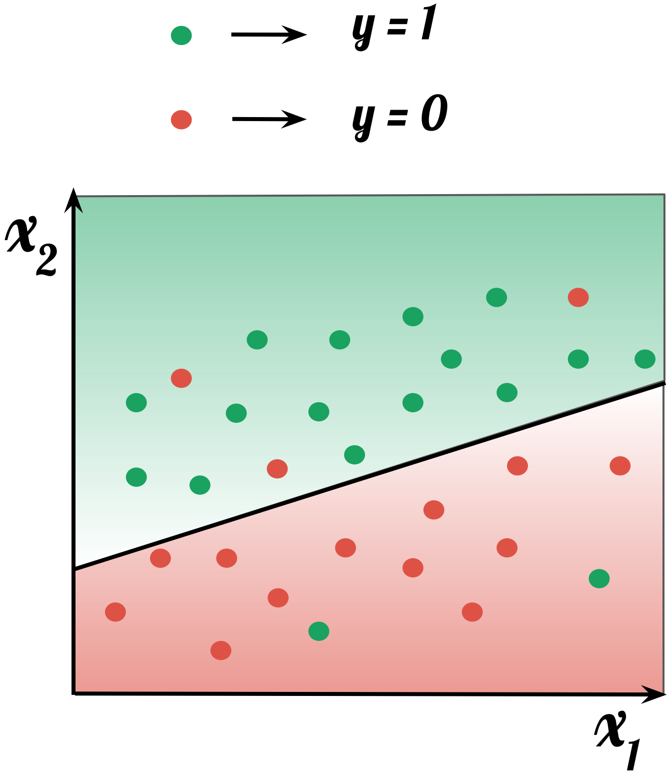


Training

Loss Function in Classification

$$\mathcal{L} = \frac{-\sum (y\log(p) + (1-y)\log(1-p))}{m}$$

m = No. of data points

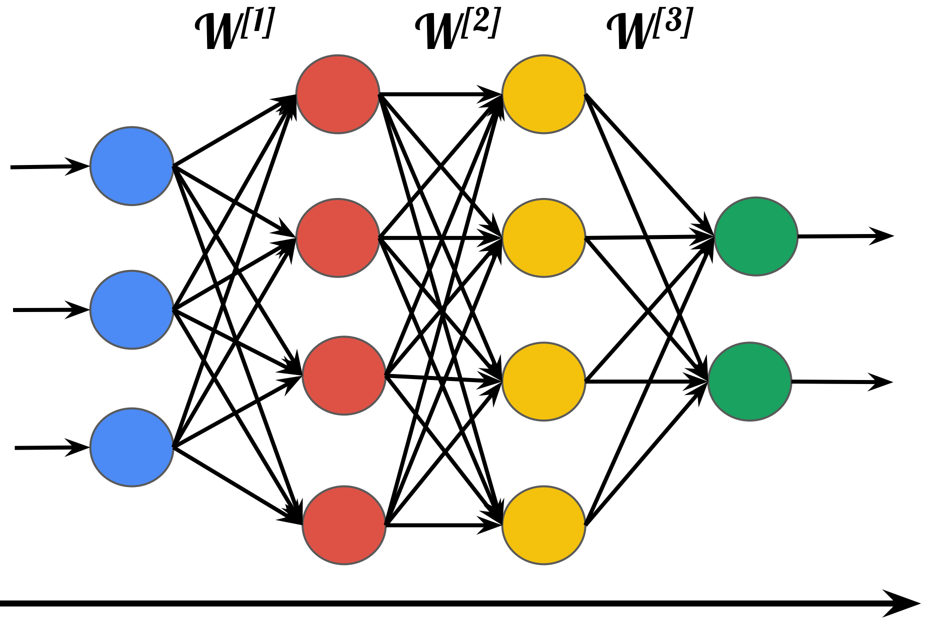


Training

Forward Propagation

$a^{[l]}$ = Activation of L - layer

$$a^{[0]} \longrightarrow a^{[1]} \longrightarrow a^{[2]} \longrightarrow a^{[3]}$$



Direction of flow of calculation

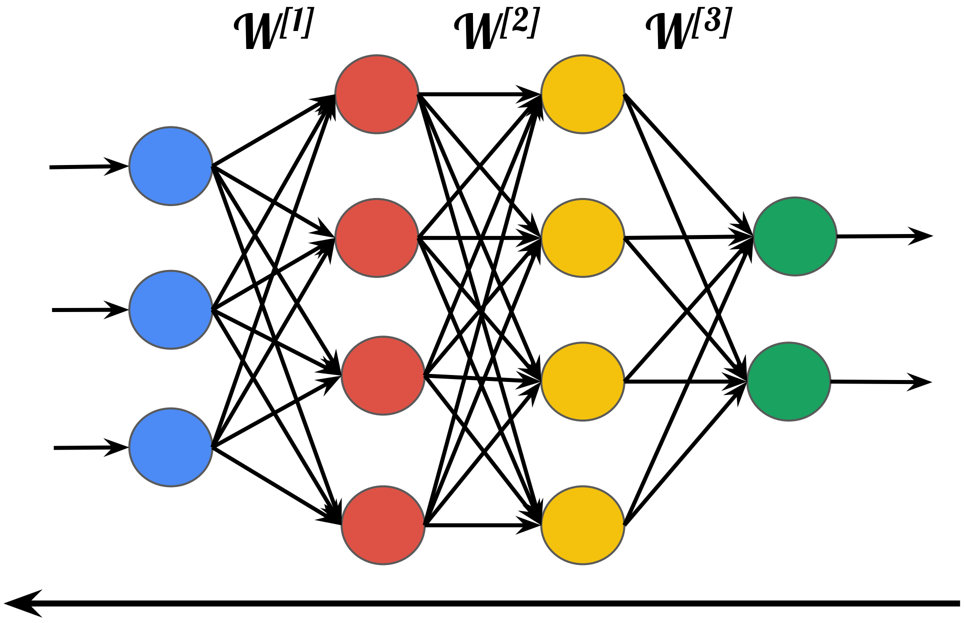
Training

Backward Propagation

$$\frac{\partial L}{\partial a^{[3]}} \rightarrow \frac{\partial L}{\partial W^{[3]}} \rightarrow \frac{\partial L}{\partial a^{[2]}}$$

$$\downarrow$$

$$\frac{\partial L}{\partial W^{[1]}} \leftarrow \frac{\partial L}{\partial a^{[1]}} \leftarrow \frac{\partial L}{\partial W^{[2]}}$$

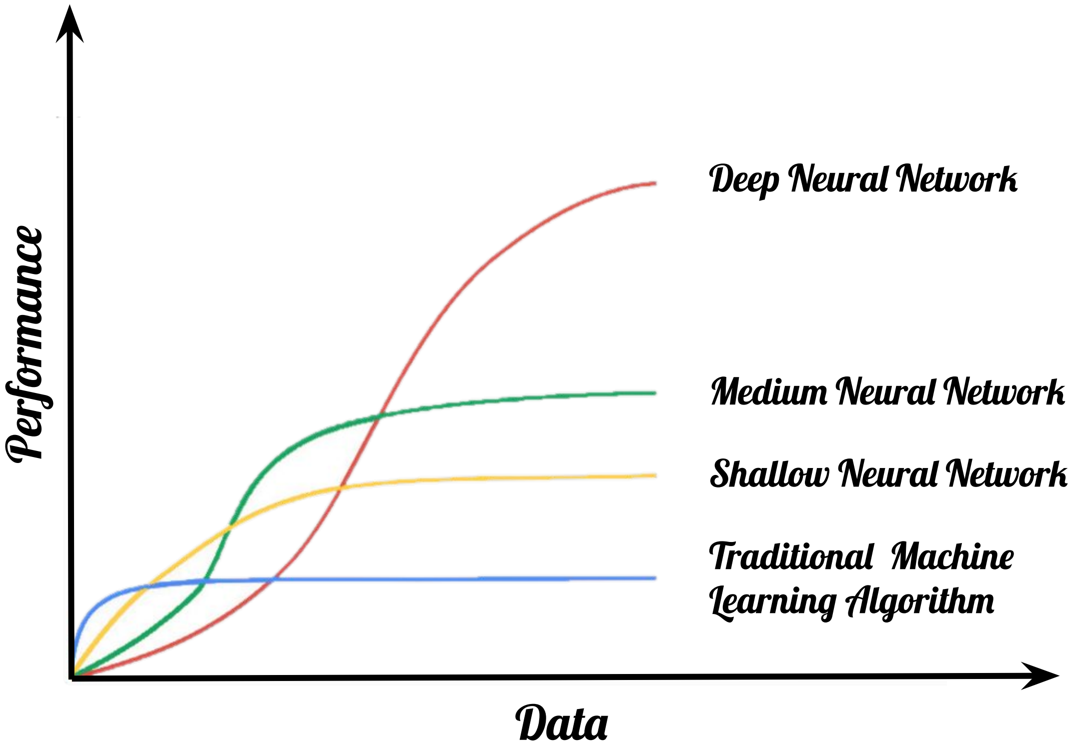


Direction of flow of calculation

Advantages

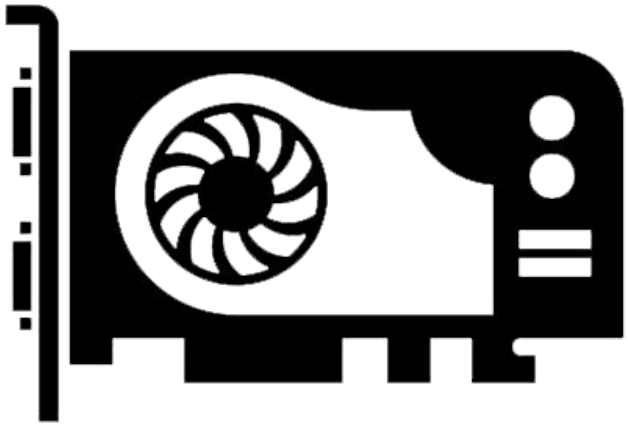
- It can capture very complex patterns

- Universal Approximation Theorem



Advantages

- Parallel processing capabilities



GPU

Advantages

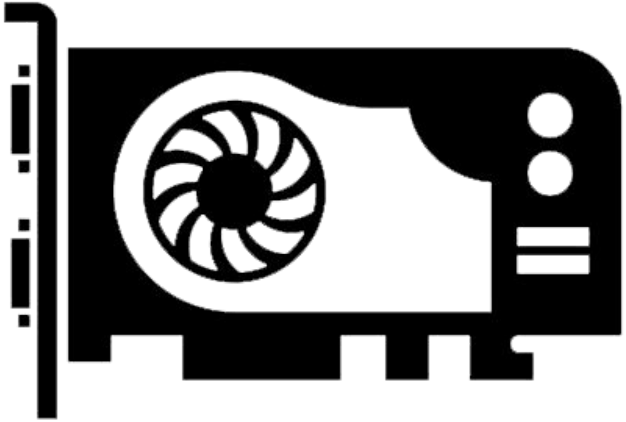
- High tolerance to noisy data
- Ability to work with incomplete knowledge

Disadvantages

- Requires a lot of data

Disadvantages

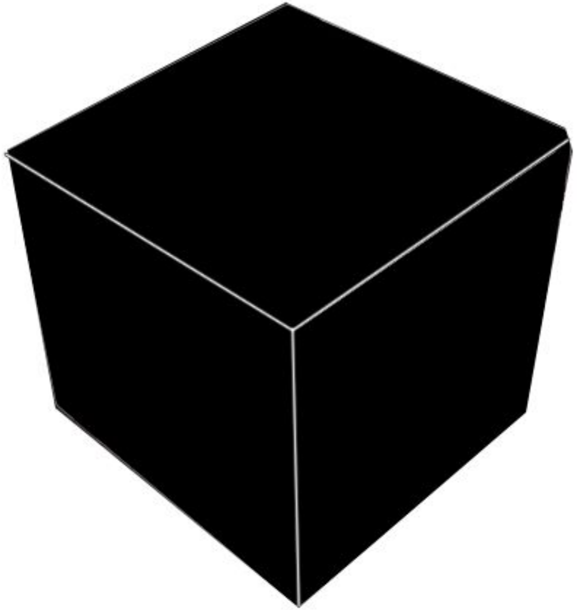
- Require a huge computational power
- Takes a lot of time to train



GPU

Disadvantages

- Unexplained behaviour of a neural network



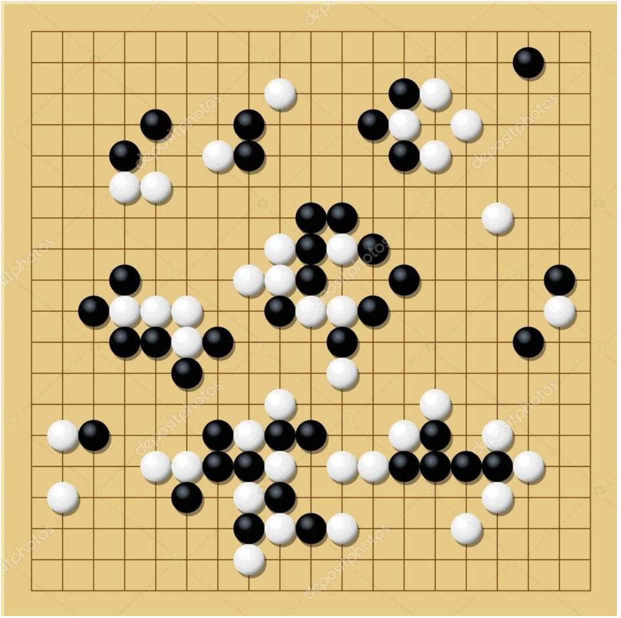
Applications

- Deepmind's AlphaGo Zero

➢ No. of atoms in the observable universe $= 10^{80}$

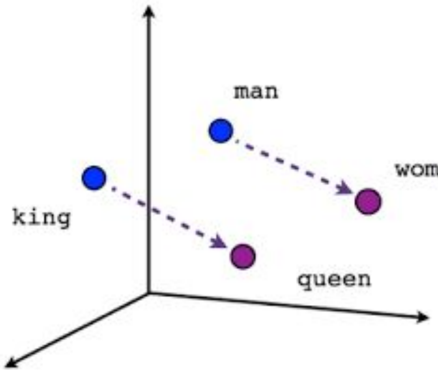
➢ No. of board positions in the chess $= 10^{120}$

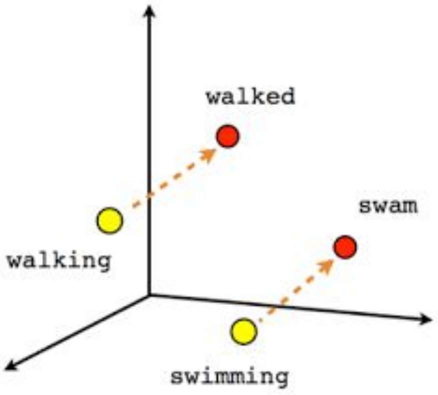
➢ No. of board positions in the go $= 10^{170}$

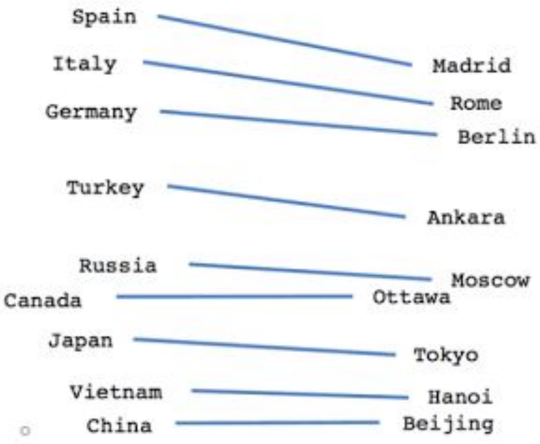


Applications

- Word Embedding

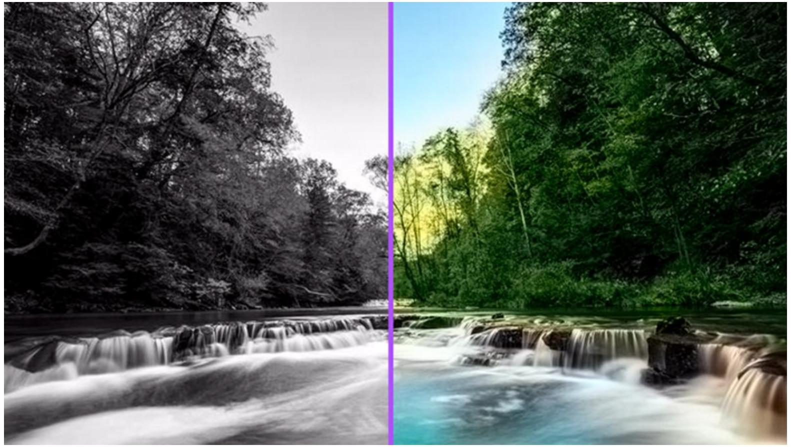






Applications

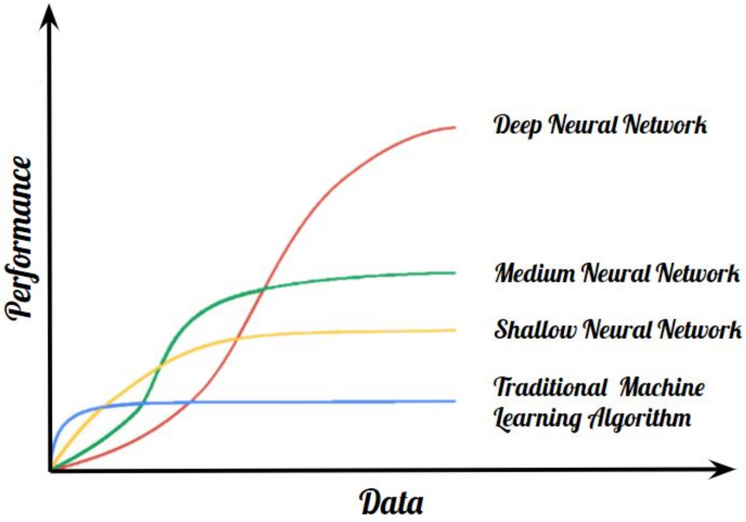
- Automatic Colorization



Conclusion

- Neural Networks have ability to perform tasks at which humans are good and computers are bad.

- Other algorithms may perform better in easy tasks.



References

- Graph illustrating the impact of data available on performance of traditional machine learning algorithms. - Research Gate. By : Benoit Gallix

- Image Colorization with Deep Convolutional Neural Networks
  By : Jeff Hwang and You Zhou

- Efficient Estimation of Word Representations in Vector Space.
  By : Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean

- Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm. By : David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, Demis Hassabis

THANK YOU