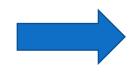


Start-Tech Academy

Human Brain VS Computer

Motivation

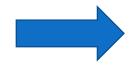






- Human mind Computer
- Good at image recognition, pattern recognition etc
- Good at arithmetic calculations



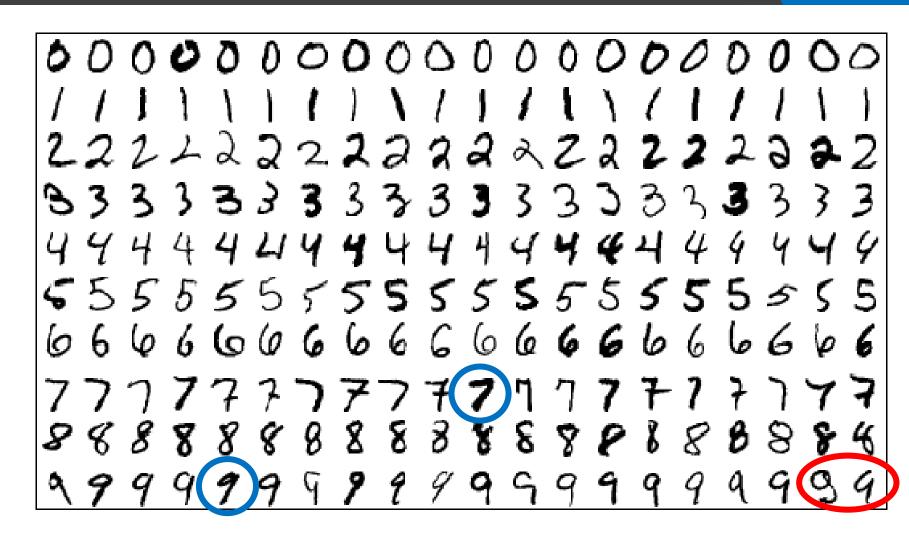


 $2574304 \times e^{354} \div \tan 5.1\pi$



Handwriting recognition

Making precise rules is difficult

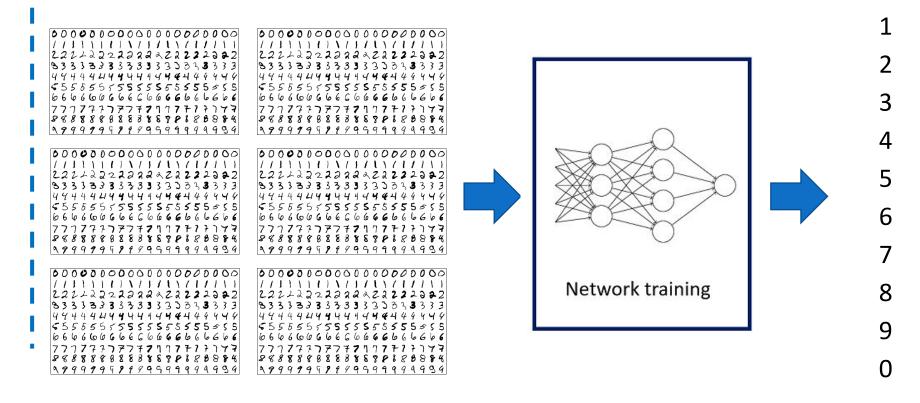




Neural Networks

Neural Networks creates own complex pattern recognition rules

Pattern recognition





Training data

Future Prediction

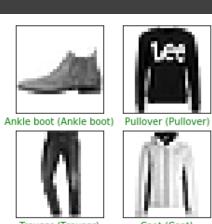
Dataset

Fashion MNIST

We will classify images into 10 fashion items



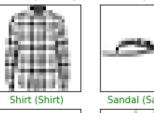
Pullover (Pullover)

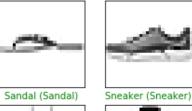


Pullover (Pullover)

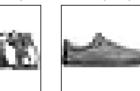
Sandal (Sandal)











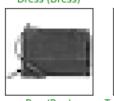






Trouser (Trouser)

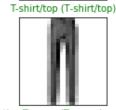






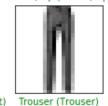
Shirt (Shirt)







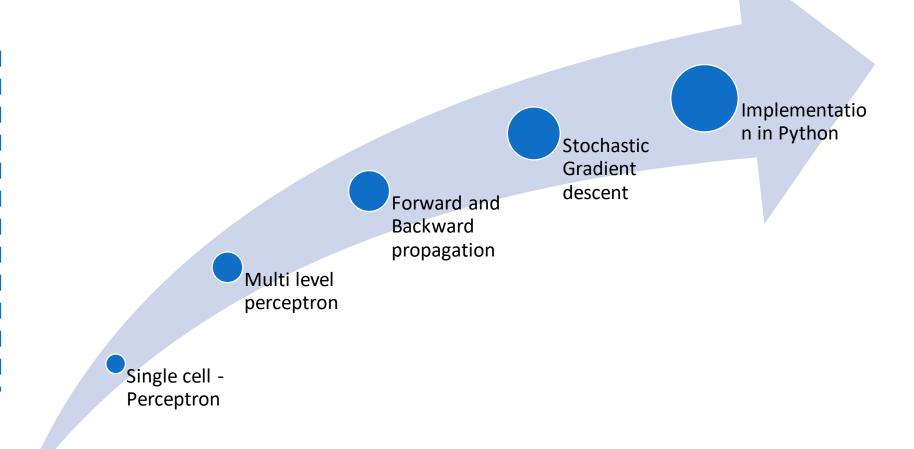






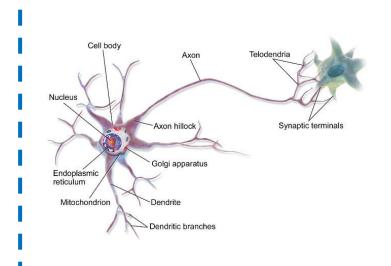
Course Flow

Course Flow

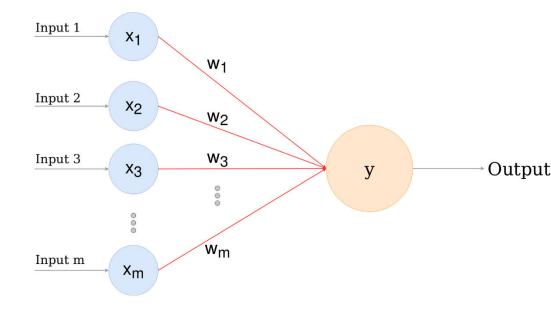




Artificial Neuron



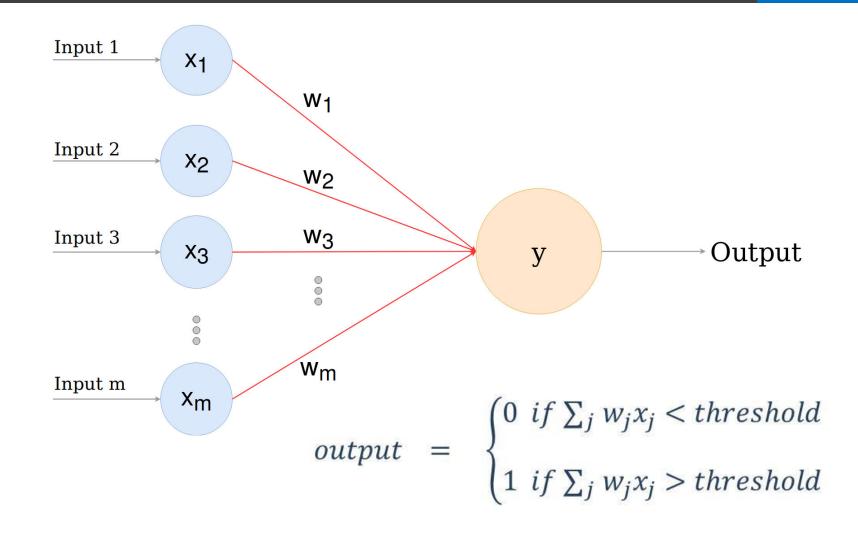
Biological Neuron



Artificial Neuron



Artificial Neuron





Purchasing a Shirt

Color

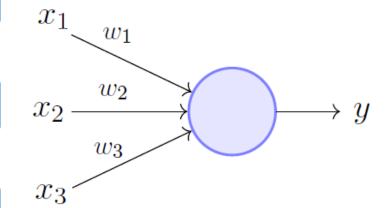
• Blue or Not

Sleeves

• Full or half

Fabric

• Cotton or not





Purchasing a Shirt

Color

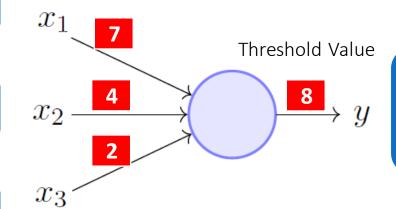
• Blue or Not

Sleeves

• Full or half

Fabric

• Cotton or not





Purchasing a Shirt

Color

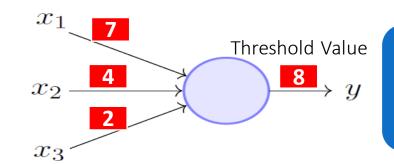
• Blue or Not

Sleeves

• Full or half

Fabric

• Cotton or not



Color	Sleeves	Fabric	Calculated Sum	Threshold	Buy / Not Buy
Blue	Half	Non Cotton	7*1 + 4*0 + 2*0 = 7	8	Not buy
Blue	Full	Non Cotton	11	8	Buy
Not Blue	Full	Cotton	6	8	Not Buy



Purchasing a Shirt

Color

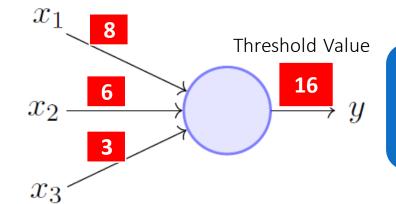
• Blue or Not

Sleeves

• Full or half

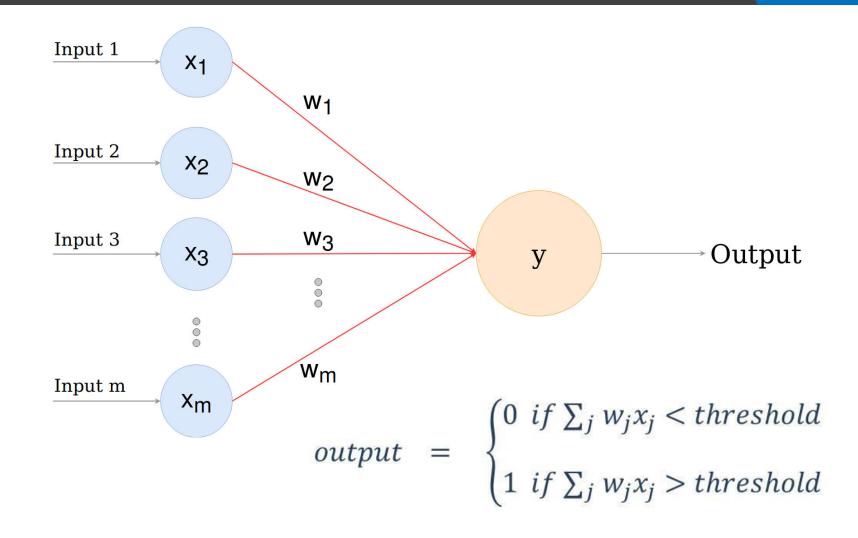
Fabric

• Cotton or not



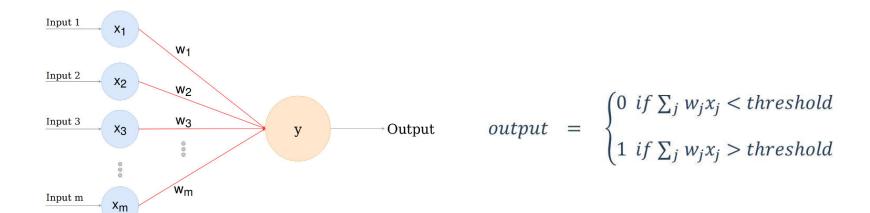


Removing Binary Restriction





Standard Equation

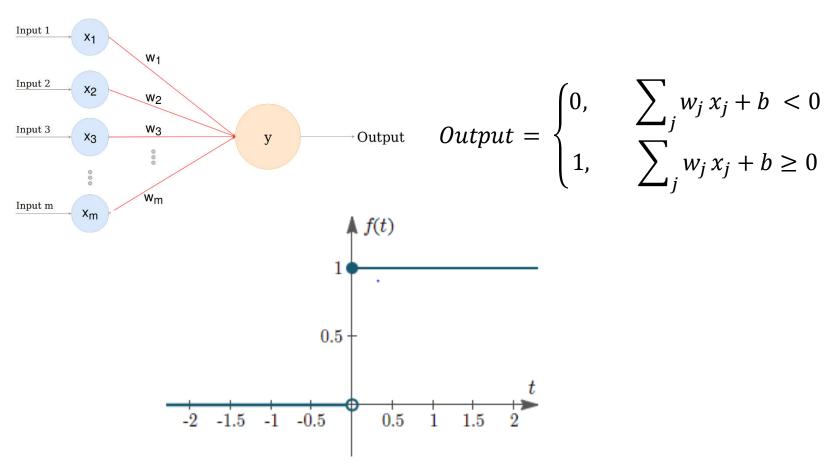


$$Output = \begin{cases} 0, & \sum_{j} w_{j} x_{j} + b < 0 \\ 1, & \sum_{j} w_{j} x_{j} + b \ge 0 \end{cases}$$

b is called Bias



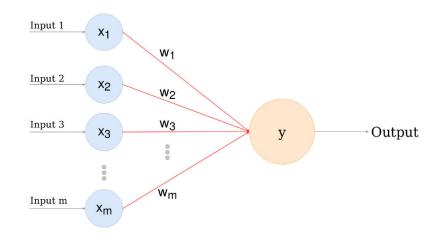
Graphical Representation

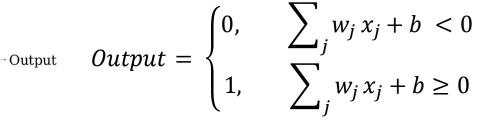


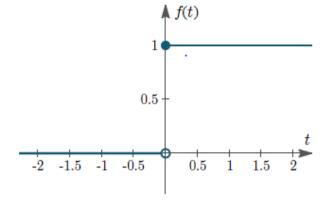
Step Activation function



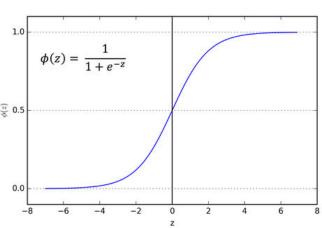
Sigmoid Activation







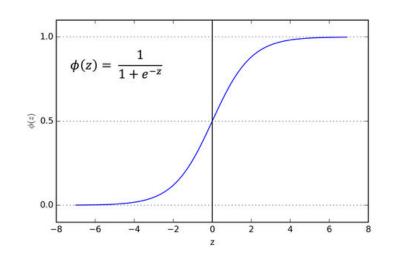




Sigmoid Activation function



Sigmoid Activation



Sigmoid Activation function

- Sigmoid is better because it is less sensitive to individual observation
- Artificial neuron with sigmoid activation is called sigmoid or logistic neuron

$$\sigma(z) \equiv rac{1}{1+e^{-z}} . \hspace{1.5cm} extit{Output} = \hspace{0.1cm} rac{1}{1+\exp(-\sum_{j}w_{j}x_{j}-b)} .$$



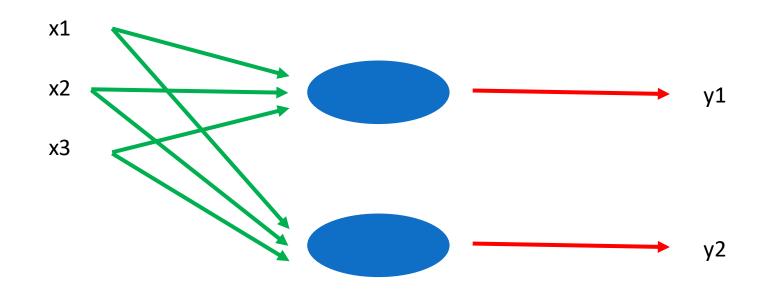
Two types of Stacking

Parallel

Sequential



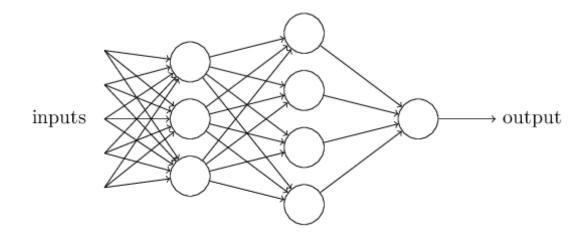
Parallel Stacking



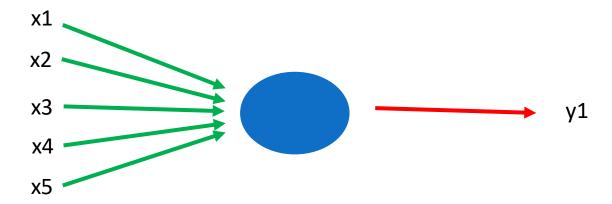
With parallel stacking we can get multiple outputs with the same input



Sequential Stacking

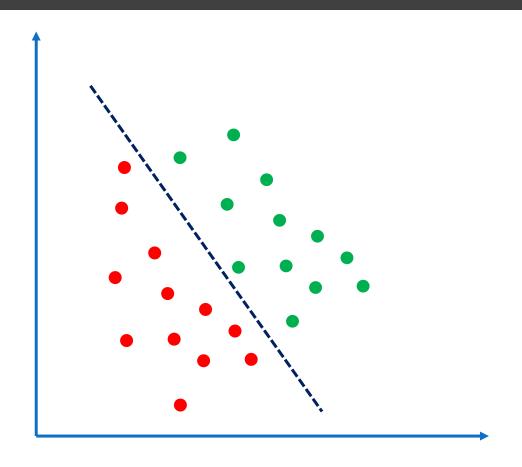


Why not use a single neuron





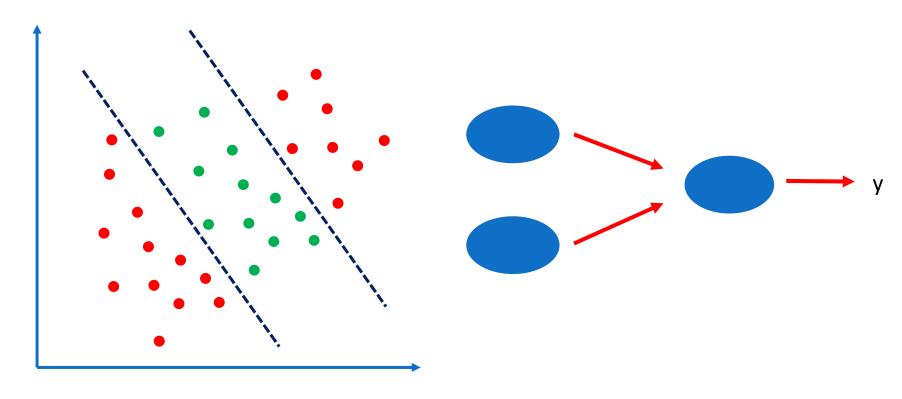
Sequential Stacking



Single neuron can handle such linear classification problem



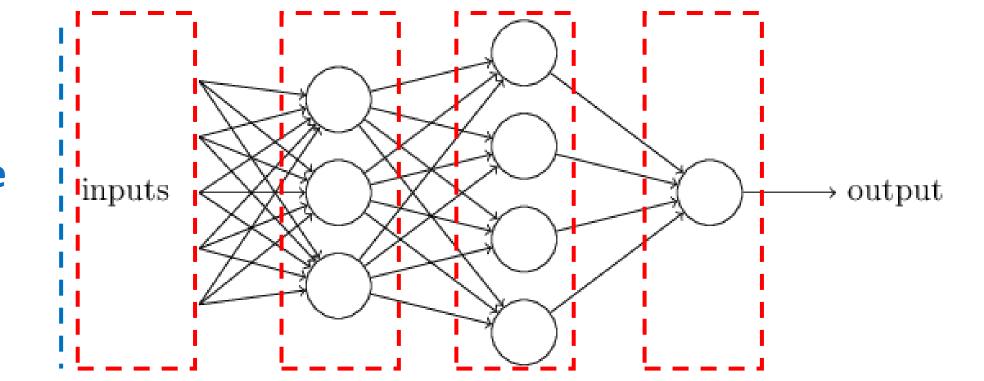
Sequential Stacking



Each neuron can focus on the particular features of the object instead of the final outcome



Input Layer



Hidden Layer 2

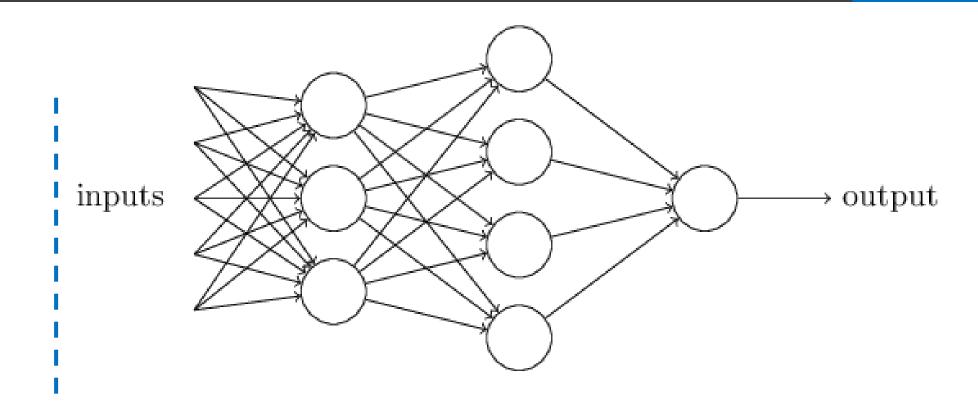
Output Layer

Nomenclature



Hidden Layer 1

Nomenclature



Feed Forward Network — One directional processing

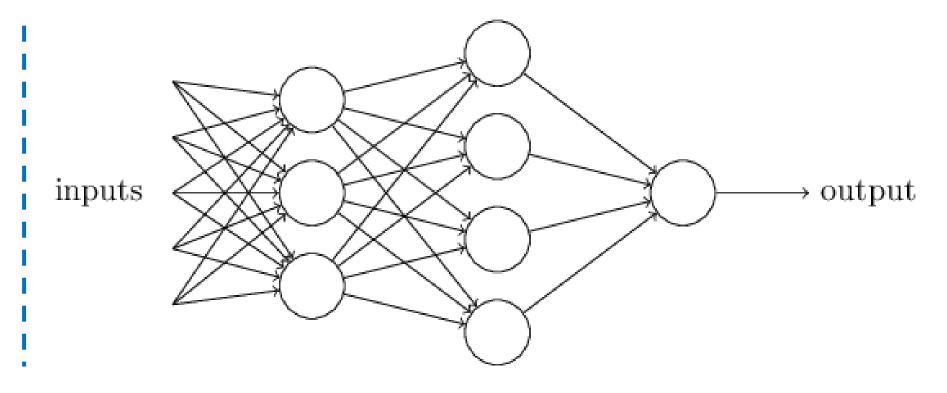
Fully connected network — Output from a neuron goes to all neurons of next layer



Deep Learning

Such artificial neural networks primarily constitutes deep learning

Deep Learning





More number of layers => Deeper network => More complex relationships

Neural Network

How it works

Covered till Now

What is a neural network

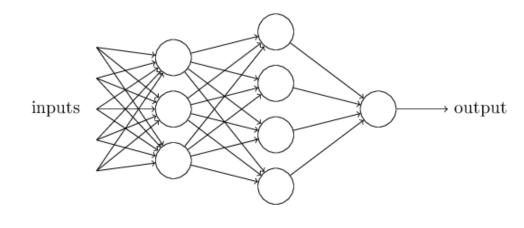
Now we are going to learn

How does a neural network works



Problem Statement

Quick Recap



$$\sigma(z) \equiv rac{1}{1+e^{-z}}.$$

$$Output = \frac{1}{1 + \exp(-\sum_{j} w_{j} x_{j} - b)}.$$

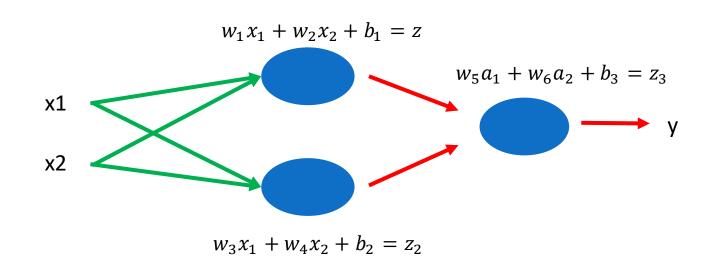
Problem Statement

 Establish the values of weights and biases so that predicted output is as close to actual output as possible



Problem Statement

Example

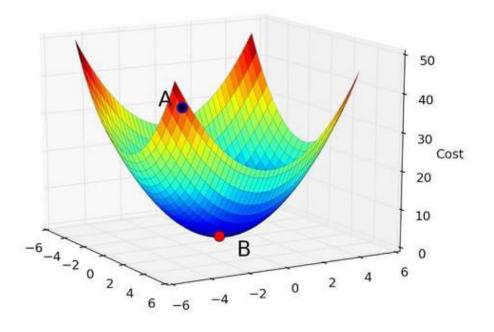


Variables to be established in this neural network

- Weights W1, W2.....W6
- Biases B1, B2, B3

Total - 9 variables

Neural Network



- GD is an optimization technique to find minimum of a function
- Better than other technique such as OLS when we have large number of features and complex relationships



 Assign random W and B values Step 1 Calculate final output using these values Step 2 Estimate error using error function Step 3 • Find those W and B which can reduce this error Step 4 Update W and B and repeat from step 2 Step 5

Initialization

Forward

Propagation

Backward

Propagation

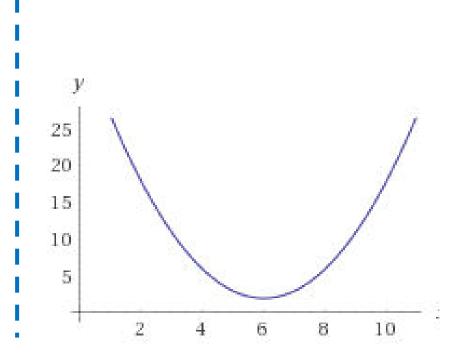
Implementati

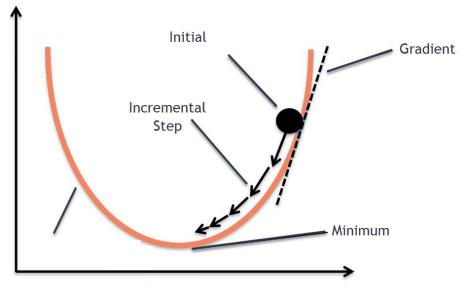
on of GD



Process

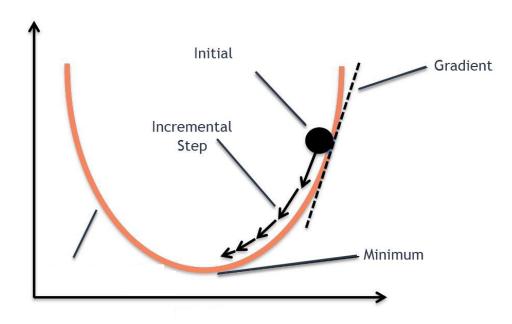
Neural Network







Neural Network



- 1. Start at a random point
- 2. Find out the **instantaneous slope** at that point
- 3. Slightly move in the direction of steepest slope
- 4. Reiterate





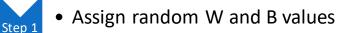




Step

Step 5

Error Function

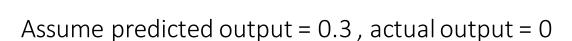


• Calculate final output using these values

• Estimate error using error function

Find those W and B which can reduce this error

Update W and B and repeat from step 2



Distance = 0 - 0.3 = -0.3

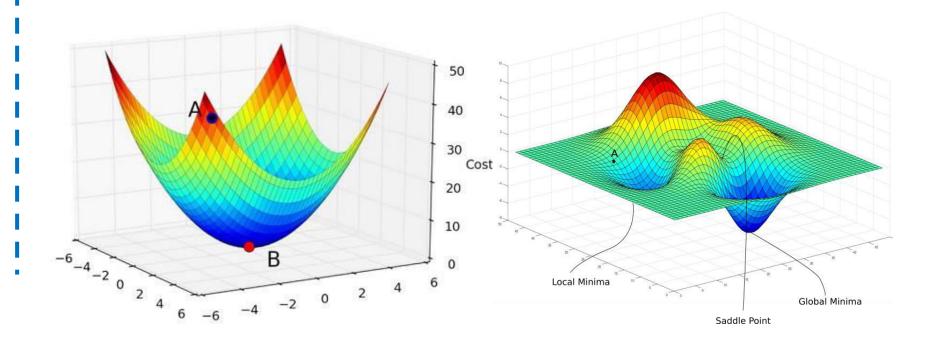
Error Function $_1 = |-0.3| = 0.3$

Error Function $_2 = (-0.3)^2 = 0.09$

Square function works well with regression but not with classification

Cross Entropy Error Function

$$= -y \log(y') - (1-y) \log(1-y')$$





Cross Entropy Error Function

$$= -y \log(y') - (1-y) \log(1-y')$$

Assume actual output = y = 1,

Error = -
$$[1(\log(y')) + (1-1)(\log(1-y'))]$$

Error =
$$-[log(y')]$$

To minimize error, we have to minimize $-\log(y')$

i.e. maximize log(y')

 \Rightarrow Maximize y'

Since y' lies between 0 and 1, y' should be as close to 1 as possible



Back Propagation

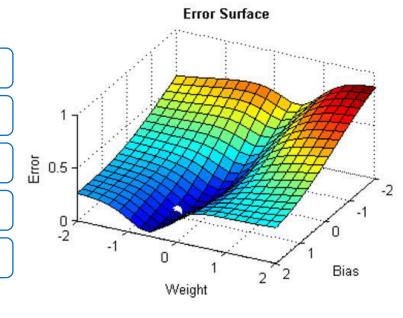


Calculate final output using these values

Estimate error using error function

• Find those W and B which can reduce this error

• Update W and B and repeat from step 2



$$w = w - \alpha \Delta w$$

Step

$$b = b - \alpha \Delta b$$

lpha is learning rate, Δw and Δb are unit steps

Alpha determines number of steps we take in downward direction



Back Propagation

$$w = w - \alpha \Delta w$$

$$b = b - \alpha \Delta b$$

To find Δw and Δb

We do back propagation

Example



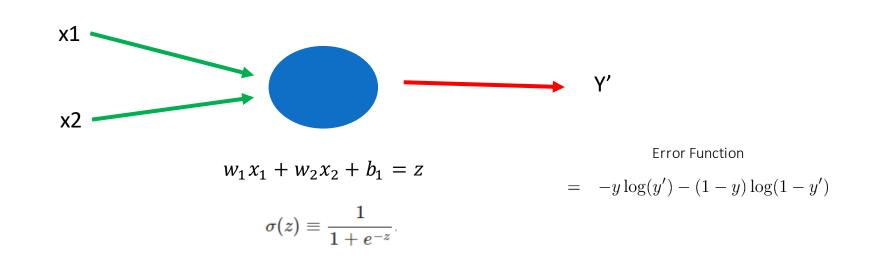
$$w_1 x_1 + w_2 x_2 + b_1 = z$$

$$\sigma(z) \equiv rac{1}{1+e^{-z}}.$$

$$= -y \log(y') - (1-y) \log(1-y')$$



Back Propagation

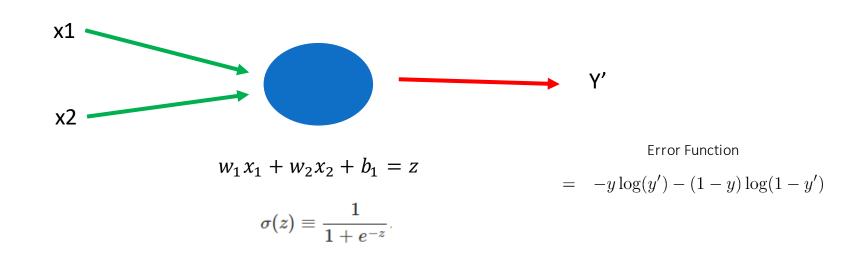


Step 1 — Initialization

W1	W2	В
2	3	-4



Back Propagation



Step 2 — Forward propagation

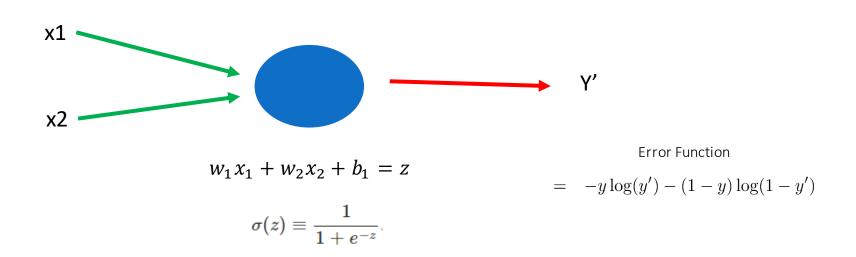
x1	x2	у
10	-4	1

$$z = 2 \times 10 + 3 \times -4 + (-4) = 4$$

Applying activation function $\sigma(z) = 0.982$



Back Propagation



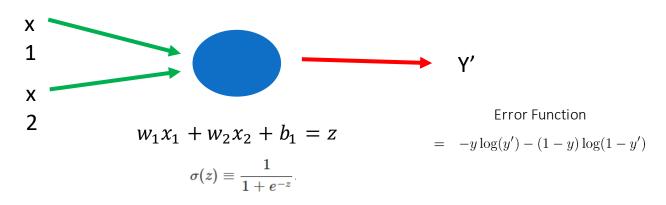
$$\textit{Step 3} - \textit{Error calculation} = -y \log(y') - (1-y) \log(1-y')$$

γ'	у
0.982	1

$$E = 0.0079$$



Back Propagation



Step 4 — Back Propagation

$$\frac{\partial E}{\partial y'}$$
 = slope of error wrt $y' = \frac{\partial (-1 \times \log(y'))}{\partial y'} = -\frac{1}{y'}$

$$\frac{\partial y'}{\partial z}$$
 = slope of activation function wrt z = $\frac{e^{-z}}{(1 + e^{-z})^2}$

$$\frac{\partial z}{\partial w_1} = x_1 = 10 \qquad \frac{\partial z}{\partial w_2} = x_2 = -4 \qquad \frac{\partial z}{\partial b} = 1$$



Back Propagation

Step 4 − Back Propagation

$$\frac{\partial E}{\partial y'} = \text{slope of error wrt } y' = \frac{\partial (-1 \times \log(y'))}{\partial y'} = -\frac{1}{y'}$$

$$\frac{\partial y'}{\partial z} = \text{slope of activation function wrt } z = \frac{e^{-z}}{(1 + e^{-z})^2}$$

$$\frac{\partial z}{\partial w_1} = x_1 = 10 \qquad \frac{\partial z}{\partial w_2} = x_2 = -4 \qquad \frac{\partial z}{\partial b} = 1$$

To
$$get \frac{\partial E}{\partial w_1}$$
 i.e. Δw_1 we apply chain rule $\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial y'} \times \frac{\partial y'}{\partial z} \times \frac{\partial z}{\partial w_1} = -0.186$

Similarly
$$\frac{\partial E}{\partial w_2} = 0.0746$$
 $\frac{\partial E}{\partial b} = -0.0186$

Back Propagation

x1
$$w_1x_1 + w_2x_2 + b_1 = z$$
 Error Function
$$\sigma(z) \equiv \frac{1}{1 + e^{-z}}$$

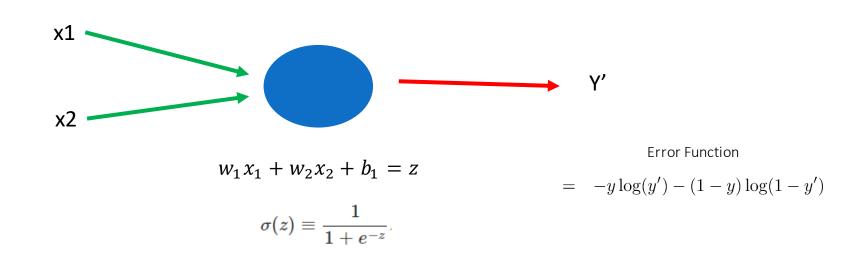
$$w1 = w1 - \alpha \Delta w1 = 2 - 5 \times -0.186 = 2.93$$

$$w2 = w2 - \alpha \Delta w2 = 3 - 5 \times 0.0746 = 2.627$$

$$b = b - \alpha \Delta b = -4 - 5 \times -0.0186 = -3.907$$



Back Propagation



Repeat Step 2 -

x1	x2	у
10	-4	1

$$z = 2.9 \times 10 + 2.6 \times -4 + (-3.9) = 14.7$$

Applying activation function $\sigma(z) = 0.999$

