Start-Tech Academy

Gradient Descent

Error Function

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

Error Function

Assume predicted output = 0.3 , actual output = 0

Distance =0 - 0.3 = -0.3

Error Function 1 = |-0.3|= 0.3

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

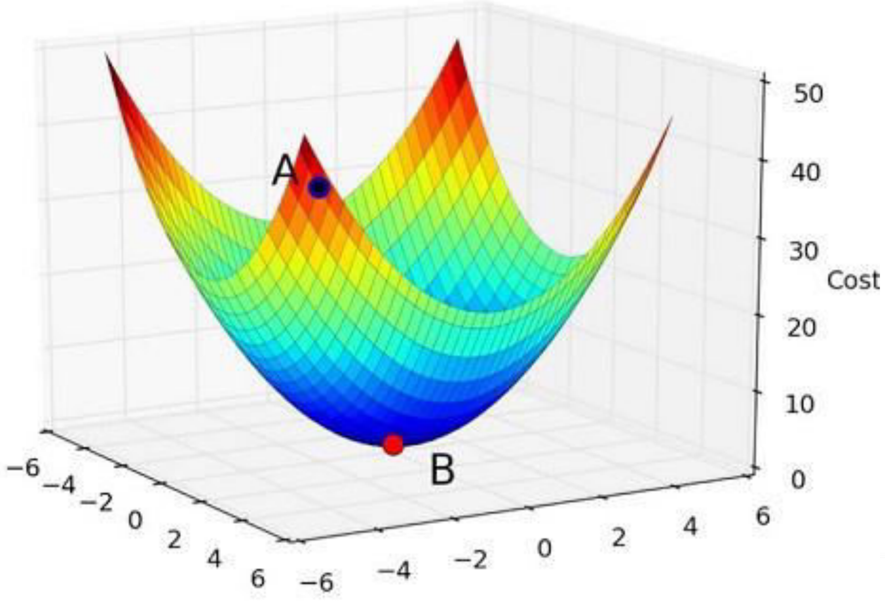
Square function works well with regression but not with classification

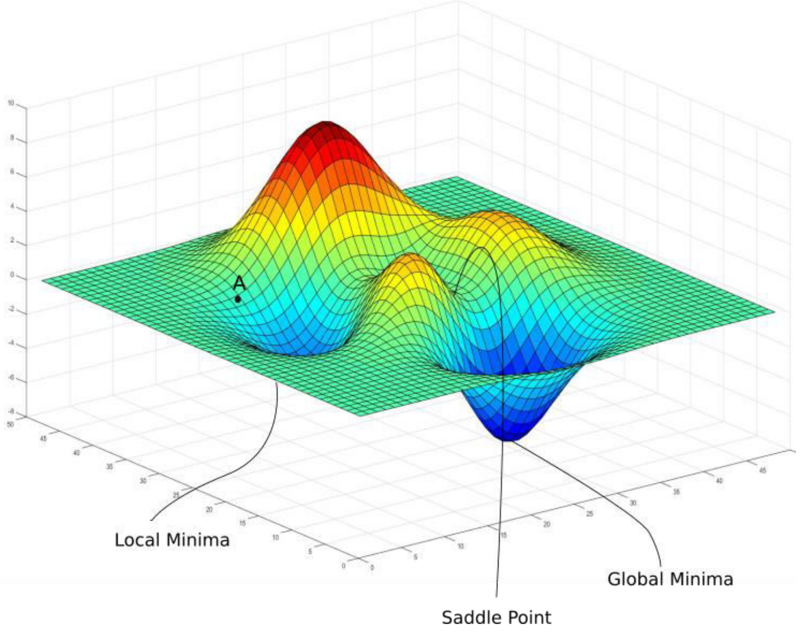
Gradient Descent

Error Function

Cross Entropy Error Function

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





Gradient Descent

Error Function

Cross Entropy Error Function

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

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’)

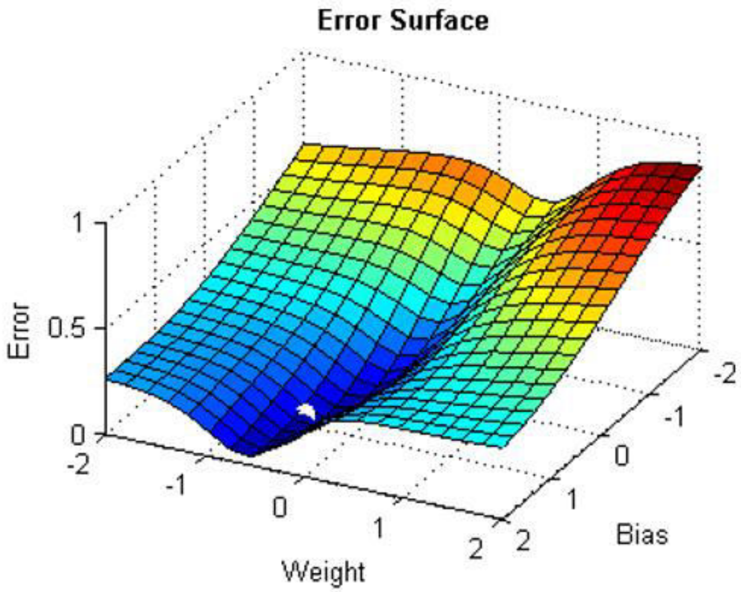
$\Rightarrow$ Maximize y’

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

Gradient Descent

Back Propagation

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



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

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

α is learning rate, Δw *and* Δb are unit steps

Alpha determines number of steps we take in downward direction

Gradient Descent

Back Propagation

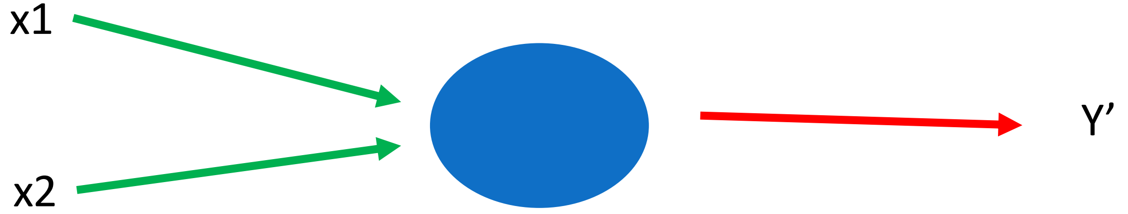
$w = w - \alpha \Delta w$

$b = b - \alpha \Delta b$

To find $\Delta w \; and \; \Delta b$

We do back propagation

Example



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

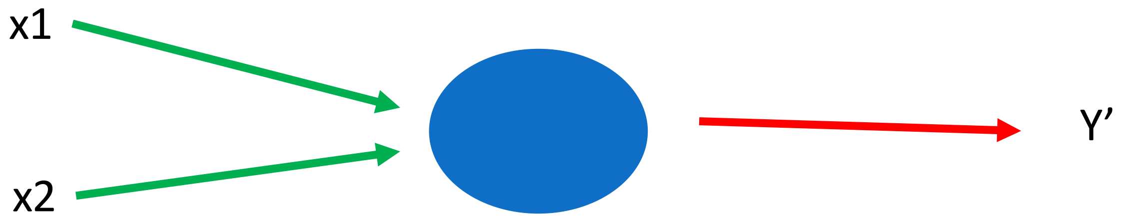
$$\sigma(z) = \frac{1}{1+e^{-z}}.$$

Error Function

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

Gradient Descent

Back Propagation



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

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

Error Function

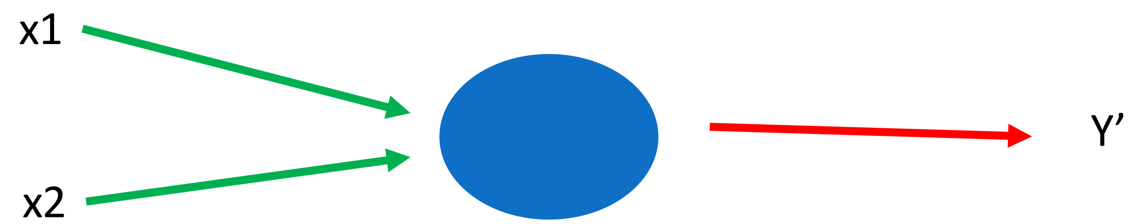
$$= \quad -y \log(y') - (1-y) \log(1-y')$$

$Step\ 1$ – Initialization

W1	W2	B
2	3	-4

Gradient Descent

Back Propagation



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

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

Error Function

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

$Step\ 2$ – Forward propagation

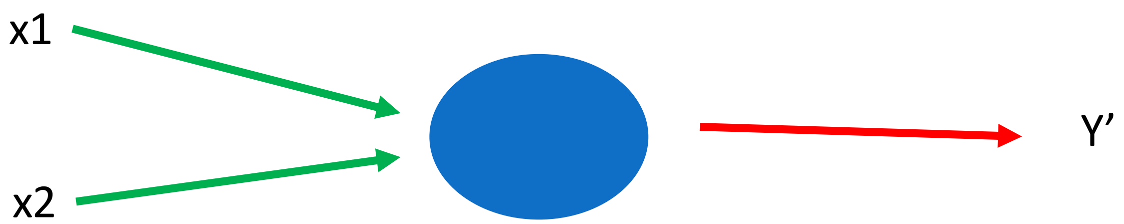
x1	x2	y
10	-4	1

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

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

Gradient Descent

Back Propagation



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

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

Error Function

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

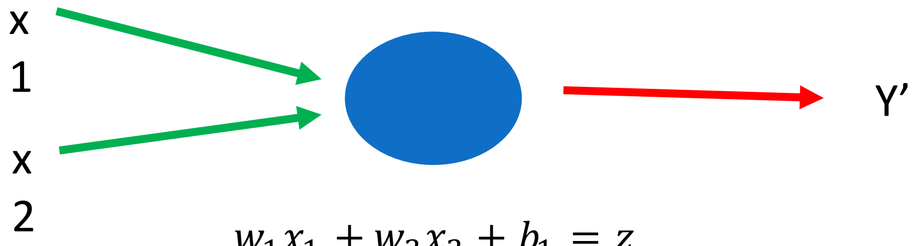
$Step$ 3 – Error calculation $= -y \log(y') - (1 - y) \log(1 - y')$

Y'	y
0.982	1

$$E = 0.0079$$

Gradient Descent

Back Propagation



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

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

Error Function

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

Step 4 – Back Propagation

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

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

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

Gradient Descent

Back Propagation

$Step\ 4$ – Back Propagation

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

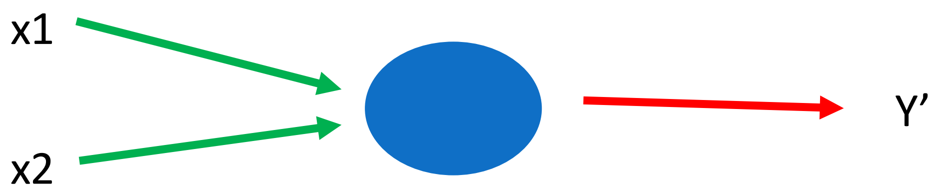
$$\frac{\partial y'}{\partial z} = \text{slope of activation function wrt z} = \frac{e^{-z}}{(1+e^{-z})^2}$$

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

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

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

Gradient Descent



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

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

Error Function

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

Back Propagation

$Step\ 5$ – Updating w and b

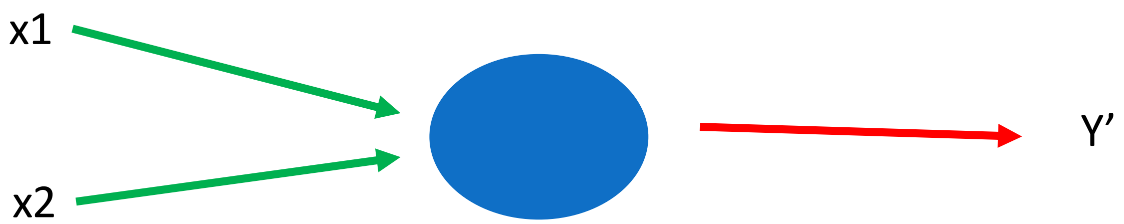
$$w1 = w1 - \alpha \Delta w1 = 2 - 5 \times -0.186 = 2.93$$

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

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

Gradient Descent

Back Propagation



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

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

Error Function

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

$Repeat\ Step\ 2\ -$

x1	x2	y
10	-4	1

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

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