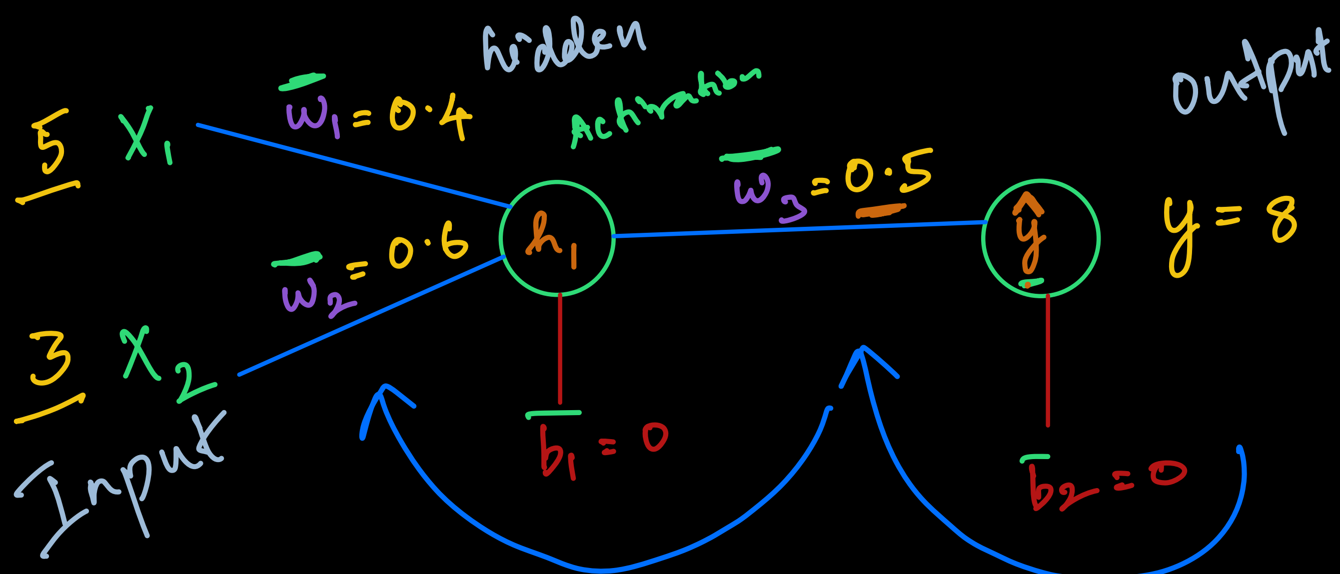
Hyperparameters

1. Iteration
2. Epoch
3. Batch size
4. Optimizers
5. Loss

- Nos of Hidden layers
- Nos of Neurons
- Activation function
- Network Topology

Back Propagation



$$h_1 = x_1 * w_1 + x_2 * w_2 + b_1$$
$$= 5 * 0.4 + 3 * 0.6 + 0$$
$$= 2.0 + 1.8 = 3.8$$

$$\boxed{\hat{y} = h_1 \cdot w_3 + b_2}$$

$$= 3.8 * 0.5 + 0$$

$$= 1.9$$

$$\text{Error} = \boxed{\frac{(\hat{y} - y)^2}{2}}$$

$$= \frac{(1.9 - 8)^2}{2} = \frac{(-6.1)^2}{2} = \frac{37.82}{2}$$

$$\boxed{= 18.6}$$

Ignore 'b'

Goal

Update the value of w_1, w_2, w_3

So that the error is less than 18.6

Back Propagation

ML - Linear

$$\frac{(\hat{y}-y)^2}{n}$$

$$\frac{((mx+b)-y)^2}{n}$$

$$\frac{m^2x^2+b^2+y^2-2y(mx+b)}{n}$$

$$e = \frac{m^2x^2+b^2+y^2-2ymx-2yb}{n}$$

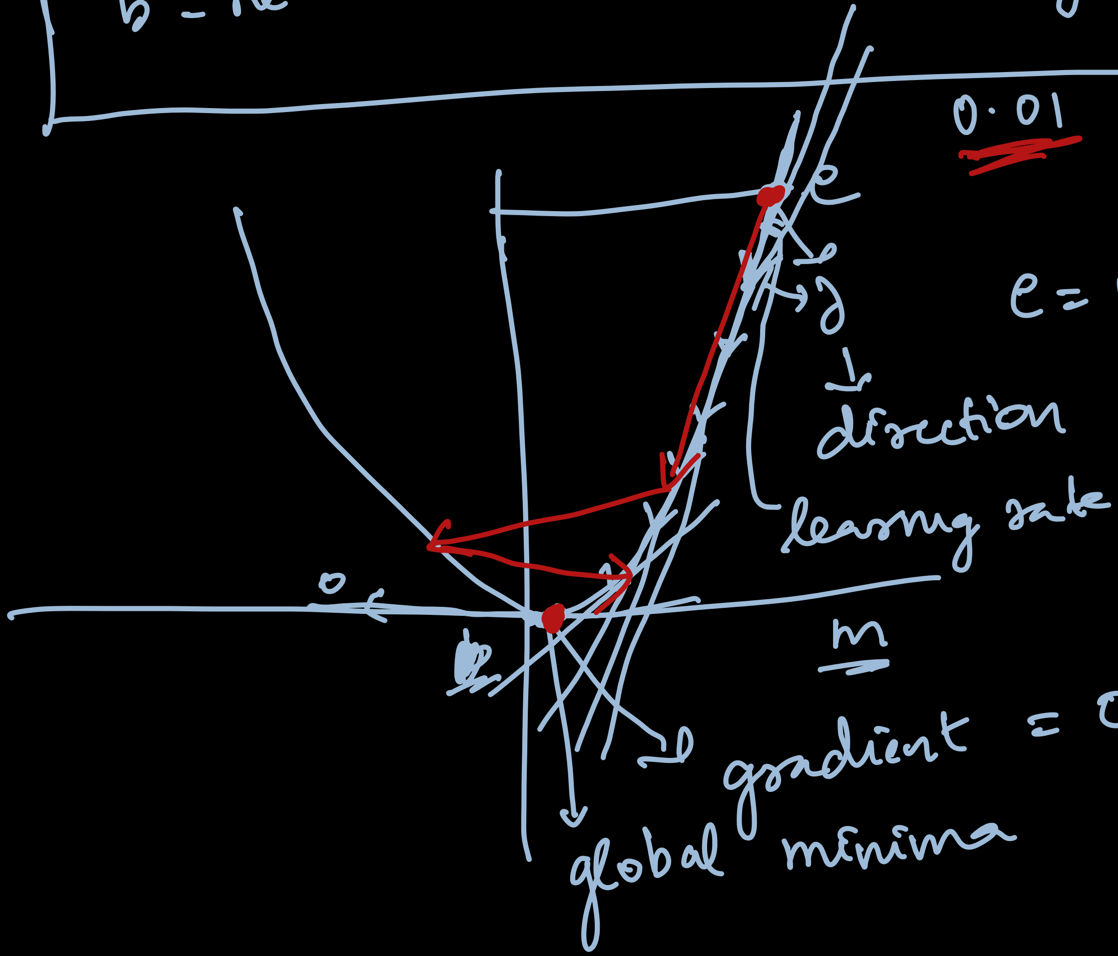
$$\frac{\partial e}{\partial m}, \quad \frac{\partial e}{\partial b}$$

$\downarrow$ gradient of m　　　$\downarrow$ gradient of b

$\downarrow$

$$\boxed{m\_new = old\_m - learning\ rate * \frac{\partial e}{\partial m}}$$

$$b\_new = old\_b - learning\ rate * \frac{\partial e}{\partial b}$$



$$e = \hat{y} - y$$

For w_3

$$E = \frac{(\hat{y} - y)^2}{2}$$

$$\frac{\partial E}{\partial \hat{y}} = \frac{\cancel{2}(\hat{y} - y) \times 1 - 0}{\cancel{2}}$$

$$= \hat{y} - y$$

$$= 1.9 - 8$$

$$= -6.1$$

$$x^n \rightarrow n x^{n-1}$$

$$y = (x - z)^2$$

$$\frac{dy}{dx} = 1 - 0 \qquad 2(x-z)^{2-1} \qquad 2(x-z)^1$$

$$x^1 \qquad 1x^{1-1} \qquad 1x^0 \Rightarrow 1$$

$$\frac{\partial E}{\partial w_3} = ?$$

$$e = \frac{(\hat{y} - y)^2}{2}$$

$$\frac{\partial e}{\partial w_3} = \boxed{\frac{\partial e}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_3}} \rightarrow \text{chain rule}$$

$$\hat{y} = h_1 \cdot w_3$$

$$\frac{\partial \hat{y}}{\partial w_3} = h_1$$

$$\frac{\partial \hat{y}}{\partial w_3} = 3.8$$

$$\frac{\partial e}{\partial w_3} = \frac{\partial e}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_3}$$

$$= -6.1 * 3.8$$

$$\frac{\partial e}{\partial w_3} = -23.18$$

gradient of w_3

$$\begin{array}{r} 4 \\ 38 \\ \times\ 61 \\ \hline 38 \\ 228 \\ \hline 2318 \end{array}$$

Weight update

$$w_3\_new = w_3 - \text{learning rate} * \frac{\partial e}{\partial w_3}$$

$$= 0.5 - 0.01 * (-23.18)$$

$$= 0.5 + .2318$$

$$w_3\text{-new} = 0.7318$$

0.01
0.001
0.0001

For w_1

$$e \frac{(\hat{y}-y)^2}{2}$$

$$\frac{\partial e}{\partial w_1} = \frac{\partial e}{\partial h_1} \cdot \frac{\partial h_1}{\partial w_1}$$

$$= \frac{\partial e}{\cancel{\partial \hat{y}}} \cdot \frac{\cancel{\partial \hat{y}}}{\cancel{\partial h_1}} \cdot \frac{\cancel{\partial h_1}}{\partial w_1}$$

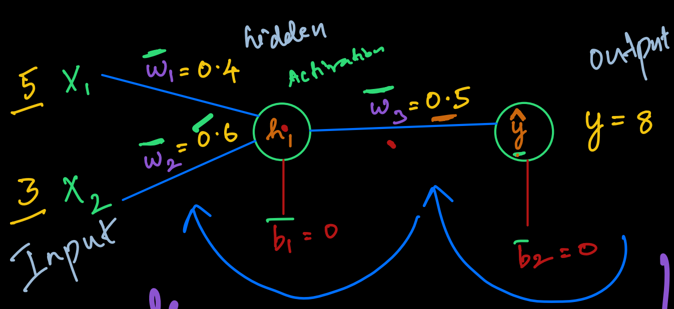
$$\frac{\partial e}{\partial \hat{y}}$$

$$\hat{y} = h_1 . w_3$$

$$\frac{\partial \hat{y}}{\partial h_1} = w_3$$

$$h_1 = x_1 w_1 + x_2 w_2$$

$$\frac{\partial h_1}{\partial w_1} = x_1$$



without chain rule

$$\frac{\partial e}{\partial w_1}$$

$$e = (\hat{y} - y)^2$$

$$e = (h_1 \cdot w_3 - y)^2$$

$$\frac{1}{\cancel{x}} * \frac{\cancel{x}}{\cancel{x}} * \frac{\cancel{x}}{6} = \frac{1}{6}$$

$$\frac{\partial e}{\partial w_1} = \frac{\partial e}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial h_1} \cdot \frac{\partial h_1}{\partial w_1}$$

$$e = ((x_1 w_1 + x_2 w_2) \cdot w_3 - y)^2$$

$$= ((x_1 w_1 w_3 + x_2 w_2 w_3) - y)^2$$

$$e = ((x_1 w_1 w_3 + x_2 w_2 w_3)^2 + y^2 - 2y(x_1 w_1 w_3 + x_2 w_2 w_3))$$

$$\frac{\partial e}{\partial w_1} =$$

why chain rule ?

$$e = \frac{(\hat{y} - y)^2}{2}$$

$$\frac{\partial e}{\partial w_1} = ?$$

$$\boxed{\frac{1}{\cancel{x}} \times \frac{\cancel{x}}{\cancel{y}} \times \frac{\cancel{y}}{z} = \frac{1}{z}}$$

$$\frac{\partial e}{\partial w_1} = \frac{\partial e}{\cancel{\partial \hat{y}}} \cdot \frac{\cancel{\partial \hat{y}}}{\cancel{\partial h_1}} \cdot \frac{\cancel{\partial h_1}}{\partial w_1}$$

chain rule

$$\frac{\partial e}{\partial \hat{y}} = -6.1$$

$$\hat{y} = h_1 \cdot w_3$$

$$\frac{\partial \hat{y}}{\partial h_1} = w_3 = 0.5$$

$$h_1 = x_1 w_1 + x_2 w_2$$

$$\frac{\partial h_1}{\partial w_1} = x_1 = 5$$

$$\frac{\partial e}{\partial w_1} = -6.1 * 0.5 * 5 = -15.25$$

Weight update

$$w_{1}\_new = w_1 - \text{learning rate} * \frac{\partial e}{\partial w_1}$$

$$= 0.4 - (0.01 * (-15.25))$$

$$w_{1}\_new = 0.5525$$

For w_2

$$\frac{\partial e}{\partial w_2} \Rightarrow \frac{\partial e}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial h_1} \cdot \frac{\partial h_1}{\partial w_2}$$

$$h_1 = x_1 w_1 + x_2 w_2$$

$$\frac{\partial h_1}{\partial w_2} = x_2 = 3$$

$$\frac{\partial e}{\partial w_2} = -6.1 * 0.5 * 3$$

$$= -9.15$$

$$w_2\text{-new} = w_2 - \text{learning rate} * \frac{\partial e}{\partial w_2}$$

$$= 0.6 - (0.01 * -9.15)$$

$$w_2\text{-new} = 0.6915$$

$$w_1\text{-new} = 0.5525$$

$$w_2\text{-new} = 0.6915$$

$$w_3\text{-new} = 0.7318$$

$$\hat{y} = h_1 * w_3\text{-new}$$

$$h_1 = x_1 w_1^{new} + x_2 w_2^{new}$$

$$= 5 \times 0.5525 + 3 * 0.6915$$

$$= 4.837$$

$$\hat{y} = 4.837 * 0.7318$$

$$= 3.5397$$

$$\text{error\_new} = \frac{(\hat{y} - y)^2}{2}$$

$$= \frac{(3.5397 - 8)^2}{2}$$

$$\Rightarrow 9.947$$

From 15.8 in the first iteration the error reduced to 9.947

Hyperparameters

Batch, Iteration, Epoch

1000 sample

Batch size = 10

10 samples will be sent

→ update the weights on every 10 samples

Scalar, Vector, Matrix, Tensor

↓

Tensorflow → framework

Pytorch → Meta

CNTK →

↓

google

↓

NN

Iteration:

1000 Inputdata

Batch of 10

10 → weight will update

100 → Iteration to complete the whole sample

Epoch:

One Complete pass through your entire training dataset

sample = 1000 , Batch = 10

epoch = 100

iteration = 10 / 1000 = 100 × 100 = 10000

epoch

$$\begin{array}{ccc} 1 \times 3 & 3 \times 1 & 1 \times 1 \end{array}$$

$$\rightarrow \begin{bmatrix} x_{11} & x_{21} & x_{31} \\ x_{21} & x_{22} & x_{32} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} \quad \end{bmatrix}$$

3 × 3