Lecture 11 Part 2

Back Propagation

Abrar Hasan
Lecturer
Dept. of Software Engineering

1

Back Propagation

Back Propagation

Remember Gradient Descent?

Gradient descent is an optimization algorithm commonly used to train machine learning models and neural networks .

Chain Rule

$$\frac{dy}{dx} = \frac{dy}{da} \quad \frac{da}{}$$

Chain Rule

$y = 5x + 3$

$x = t^2$

Now we have to find $\frac{dy}{dt}$

Chain Rule

What generally we do,

$$y = 5x + 3$$

$$\quad = 5(t^2)+3$$

$$\quad = 5t^2+3$$

$$\frac{dy}{dt} = \frac{d(5t^2+3)}{dt}$$

$$=10t$$

$y = 5x + 3$

$x = t^2$

$$\frac{dy}{dt} = ?$$

Chain Rule

But in chain rule we see

- Y is function of X
- X is a function of T

$y = 5x + 3$

$x = t^2$

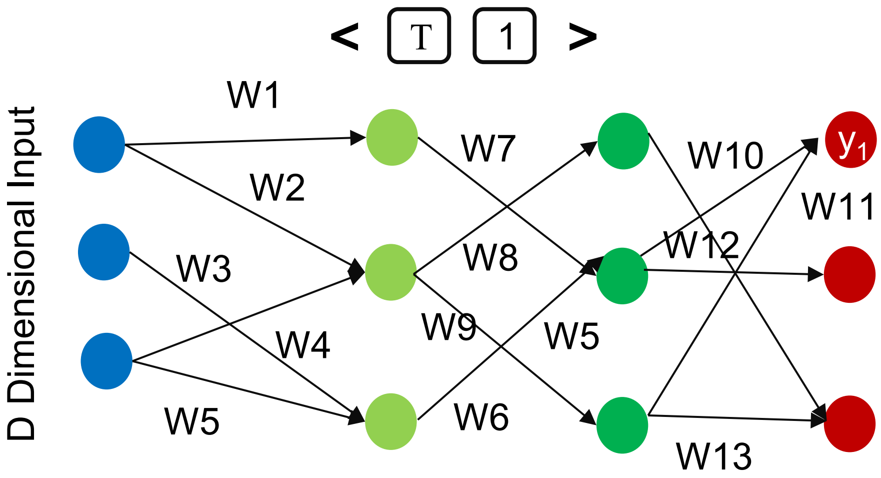
$$\frac{dy}{dt} = ?$$

Instead of doing this $\frac{dy}{dt} = \frac{d(5t^2+3)}{dt}$

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt}$$

$$= \frac{d}{dx}(5x+3) * \frac{d}{dt}(t^2)$$

$$= 5 * 2t = 10t$$



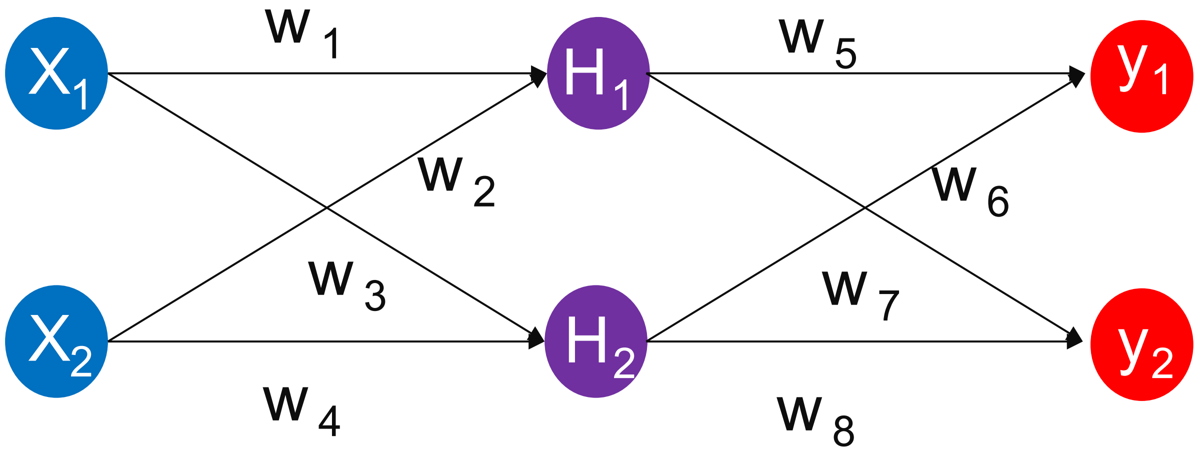
$$\text{Total Error} = \frac{1}{2}(T_1\text{-outy}_1)^2 + \frac{1}{2}(T_2\text{-outy}_2)^2 + \frac{1}{2}(T_3\text{-outy}_3)^2$$

$$\text{Outy}_1 = \text{activation}(w_{10}x_1 + w_{11}x_2)$$

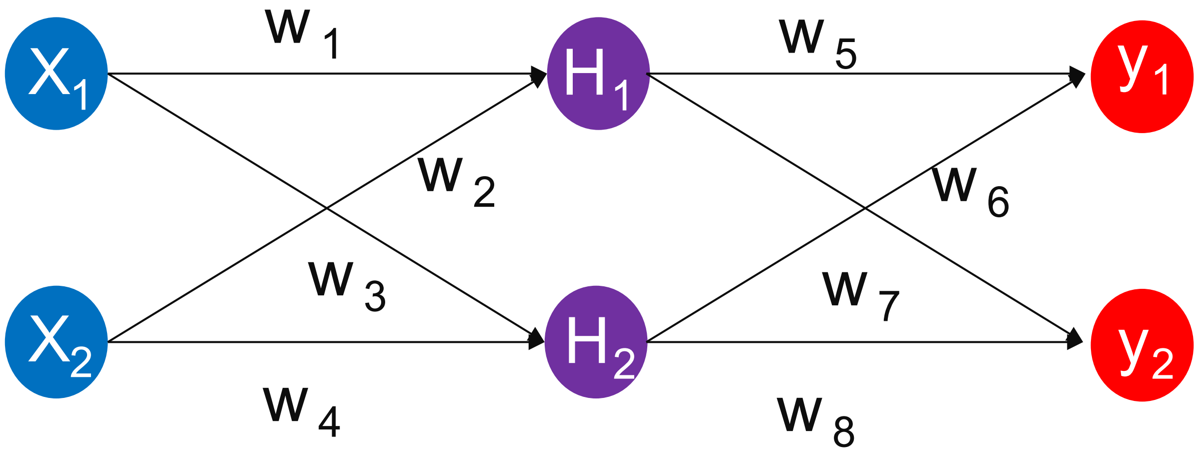
< T 1 >

$$\frac{dError}{dW_{11}} = \frac{dError}{d\ outy_1} * \frac{d\ outy_1}{dy_1} * \frac{d\ y_1}{dw_{11}}$$

< T 1 >



< T 1 >



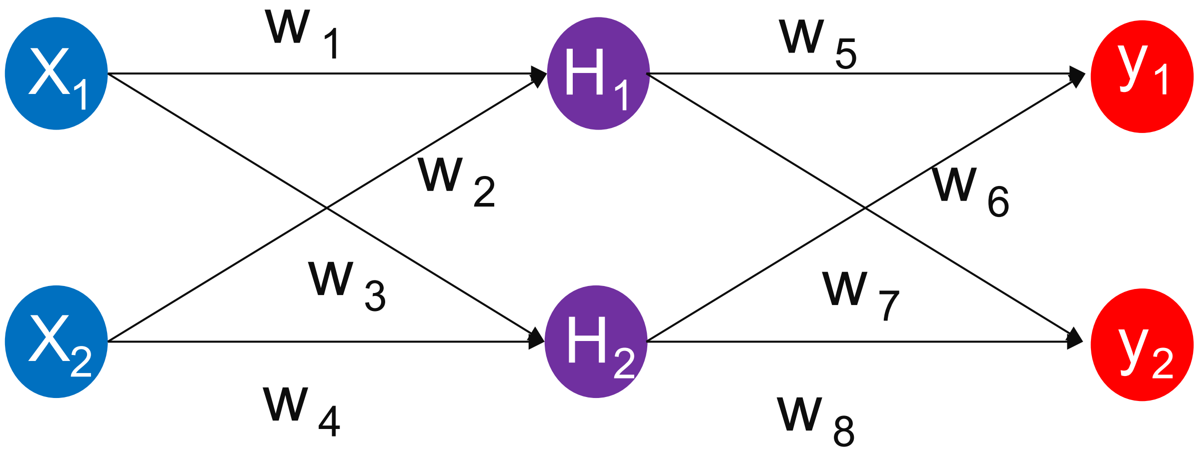
$H_1 = x_1 * w_1 + x_2 * w_2 + b_1$

$H_2 = x_1 * w_3 + x_2 * w_4 + b_2$

$\text{output\_H}_1 = \frac{1}{1+e^{-H_1}}$

$\text{output\_H}_2 = \frac{1}{1+e^{-H_2}}$

< T 1 >



$$y_1 = out\_H_1 * w_5 + out\_H_2 * w_6 + b_3$$

$$y_1 = outH_1 * w_7 + outH_2 * w_8 + b_4$$

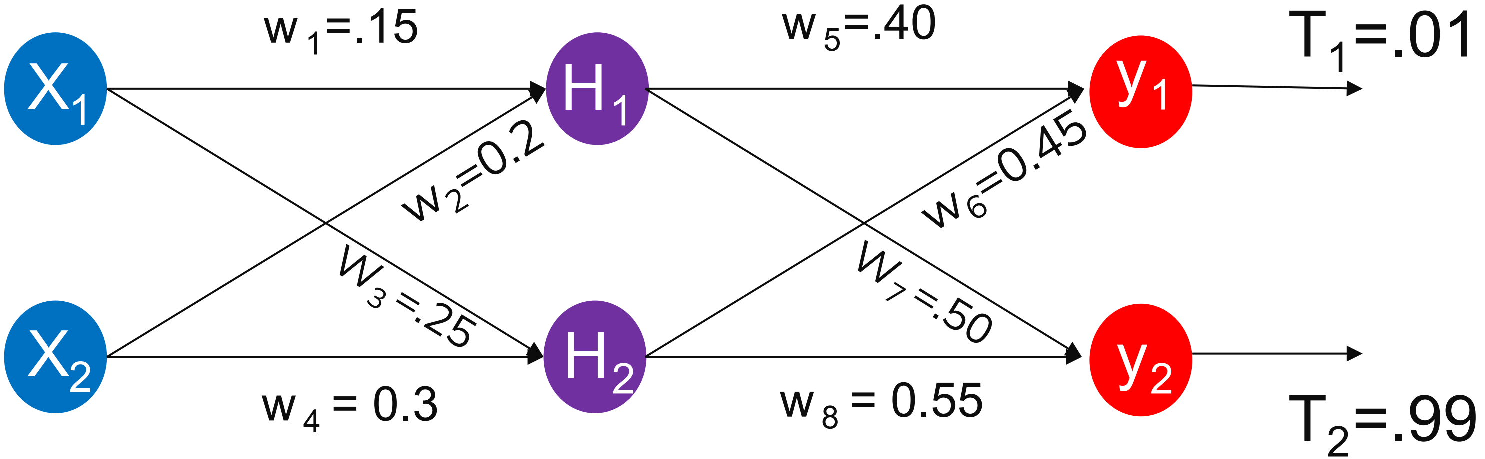
$$output\_y_1 = \frac{1}{1+e^{-y_1}}$$

$$output\_y_1 = \frac{1}{1+e^{-y_1}}$$

What we just learnt is called forward pass

Now Lets see a mathematical example

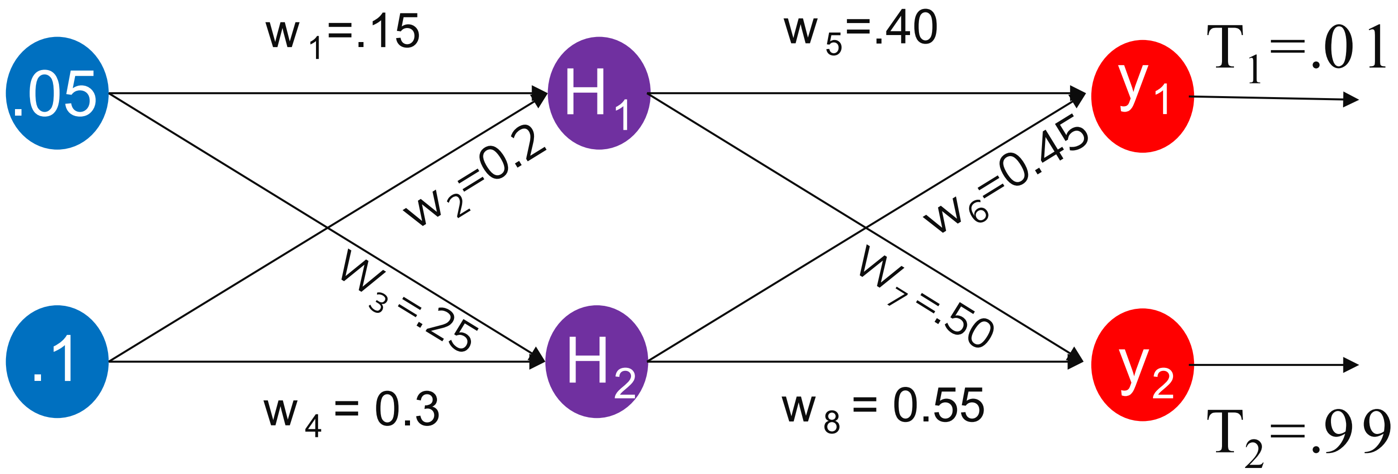
< T 1 >



$X_1=.05$
$X_2=.01$

$b_1 = b_2 = .35$
$b_3 = b_4 = .60$

< T 1 >



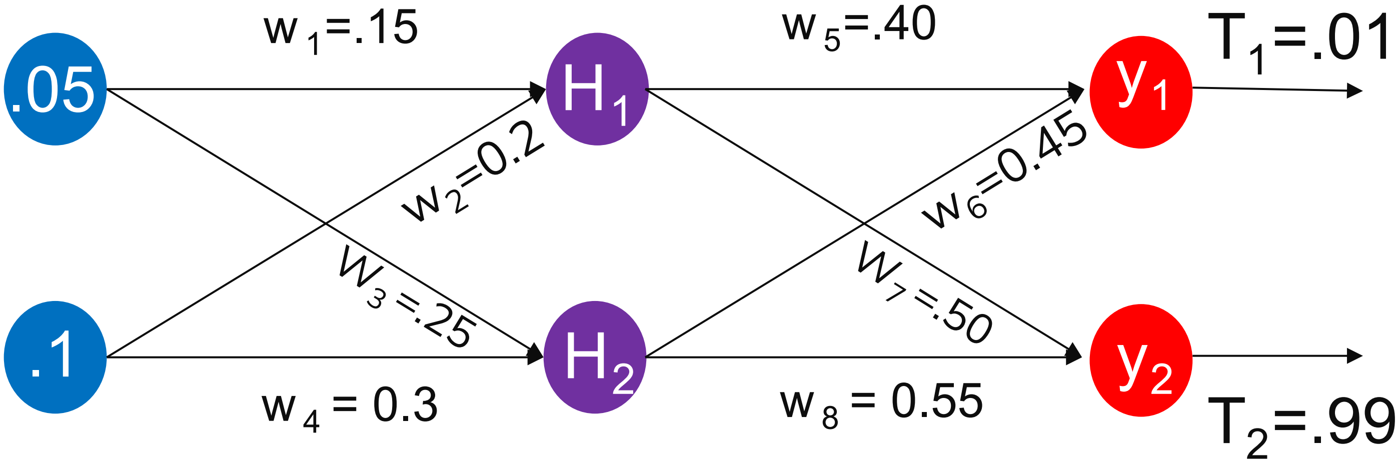
$$H_1 = x_1 * w_1 + x_2 * w_2 + b_1$$

$$H_1 = .05 * .15 + .1 * .2 + .35$$

$$= .3775$$

$$output\_H_1 = \frac{1}{1+e^{-H_1}}$$

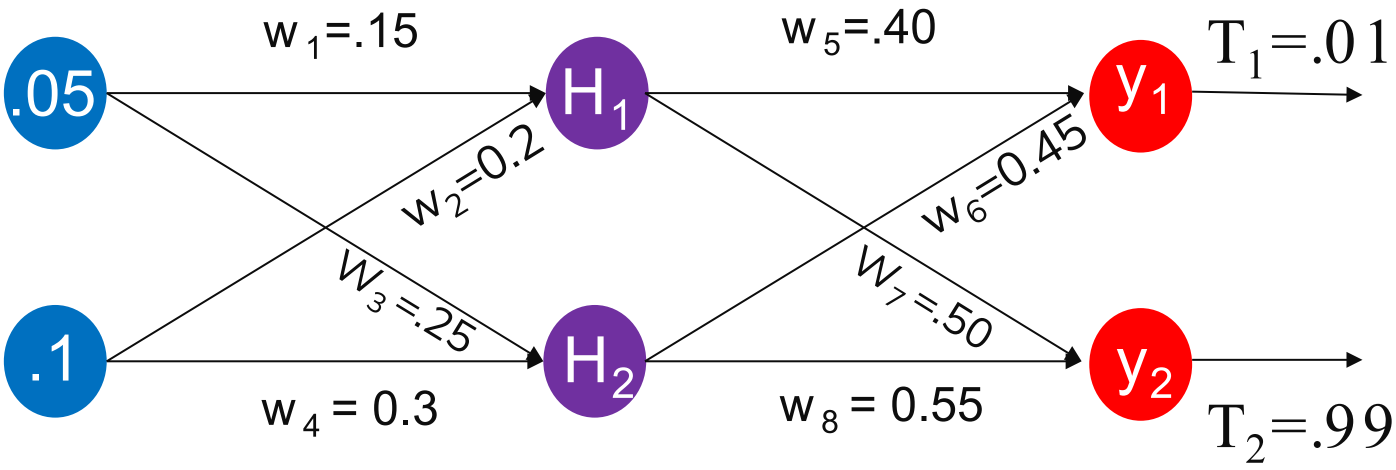
$$= .5932$$



H_2= .3925

output_H $_2$= .59688

< T 1 >



$$y_1 = out\_H_1 * w_5 + out\_H_2 * w_6 + b_3$$

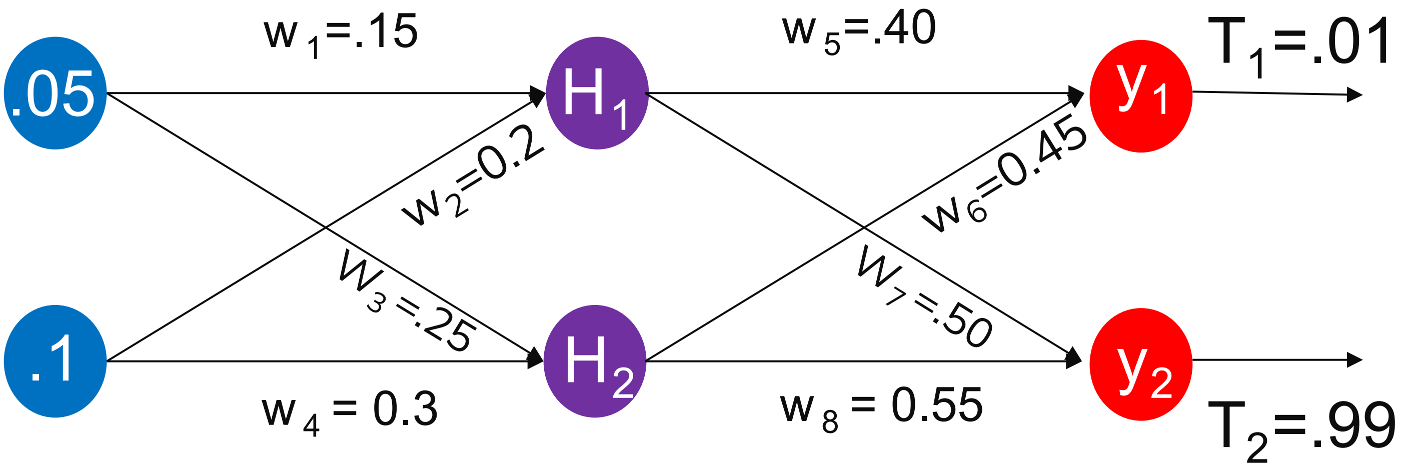
$$= .5932 * .4 + .59688 * .45 + .6$$

$$= 1.1056$$

$$output\_y_1 = \frac{1}{1+e^{-y_1}}$$

$$= .7513$$

< T 1 >



y2 = 1.22

output_y $_1$ = .77

Total Error = $\frac{1}{2}(T_1\text{-outy}_1)^2 + \frac{1}{2}(T_2\text{-outy}_2)^2$

$$\frac{dError}{dW_5} = \frac{dError}{d\,outy_1} * \frac{d\,outy_1}{dy_1} * \frac{d\,y_1}{dw_5}$$

$$\frac{dError}{d\,outy_1} = \frac{d}{d\,outy_1}\left(\frac{1}{2}(T_1\text{-outy}_1)^2 + \frac{1}{2}(T_2\text{-outy}_2)^2\right)$$

$$=2 * \frac{1}{2} * (T_1\text{-outy}_1) * (0-1) + 0$$

$$= -\,(T_1\text{-outy}_1)$$

< T 1 >

Total Error = $\frac{1}{2}(T_1\text{-}outy_1)^2 + \frac{1}{2}(T_2\text{-}outy_2)^2$

$$\frac{dError}{dW_5} = \frac{dError}{d\,outy_1} * \frac{d\,outy_1}{dy_1} * \frac{d\,y_1}{dw_5}$$

$$\frac{d\,outy_1}{dy_1} = \text{Derivative of Sigmoid}$$

$$= outy_1 * (1 - outy_1)$$

Derivative of Sigmoid

$$\frac{d\sigma(x)}{dy_1} = \sigma(x) * (1 - \sigma(x))$$

< T 1 >

$$\text{Total Error} = \frac{1}{2}(T_1 - outy_1)^2 + \frac{1}{2}(T_2 - outy_2)^2$$

$$\frac{dError}{dW_5} = \frac{dError}{d\,outy_1} * \frac{d\,outy_1}{dy_1} * \frac{d\,y_1}{dw_5}$$

$$\frac{d\,y_1}{dw_5} = \frac{d}{dw_5}(\,out\_H_1 * w_5 + out\_H_2 * w_6 + b_3)$$

$$= out\_H_1 + 0 + 0$$

$$= out\_H_1$$

< T 1 >

Total Error = $\frac{1}{2}(T_1\text{-}outy_1)^2 + \frac{1}{2}(T_2\text{-}outy_2)^2$

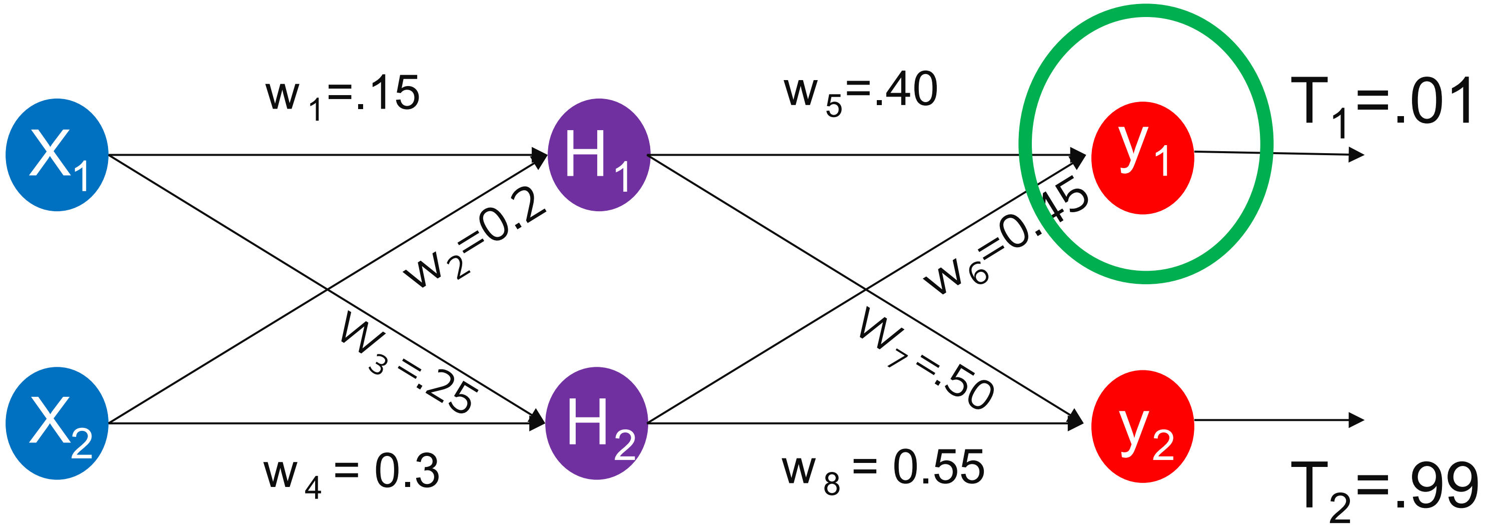
$$\frac{dError}{dW_5} = \frac{dError}{d\,outy_1} * \frac{d\,outy_1}{dy_1} * \frac{d\,y_1}{dw_5}$$

$$= -(T_1 - outy_1) * outy_1 * (1 - outy_1) * out\_H_1$$

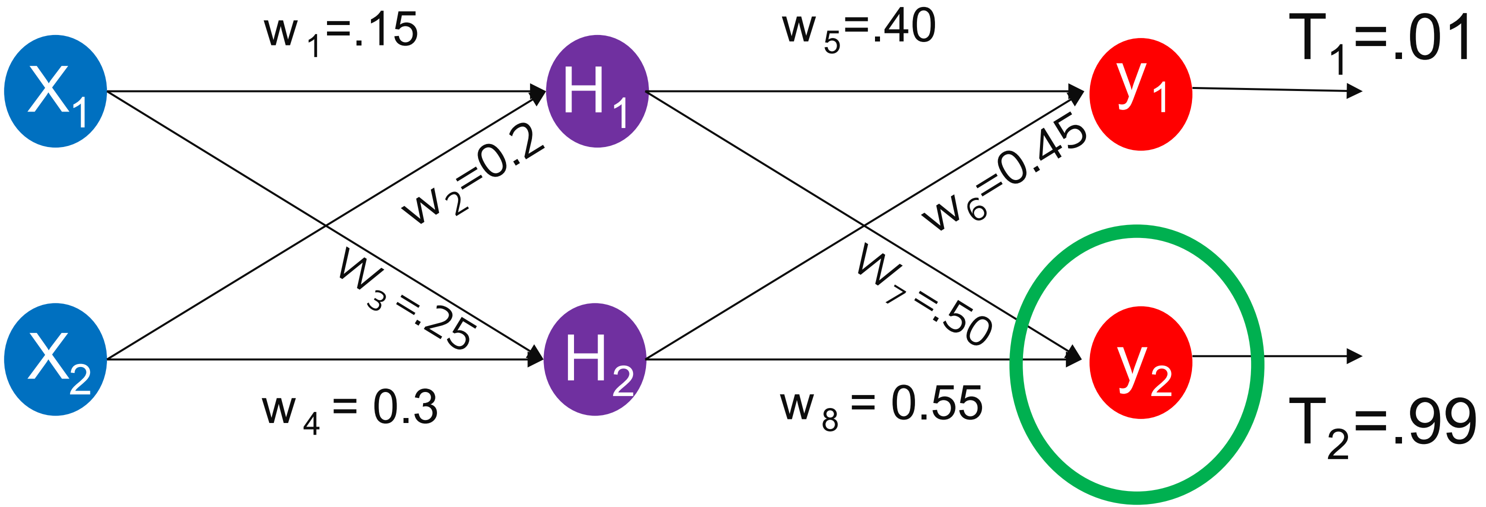
< T 1 >

$$W_5 = W_5 - \eta * \frac{dError}{dW_5}$$

Do the same for all weight

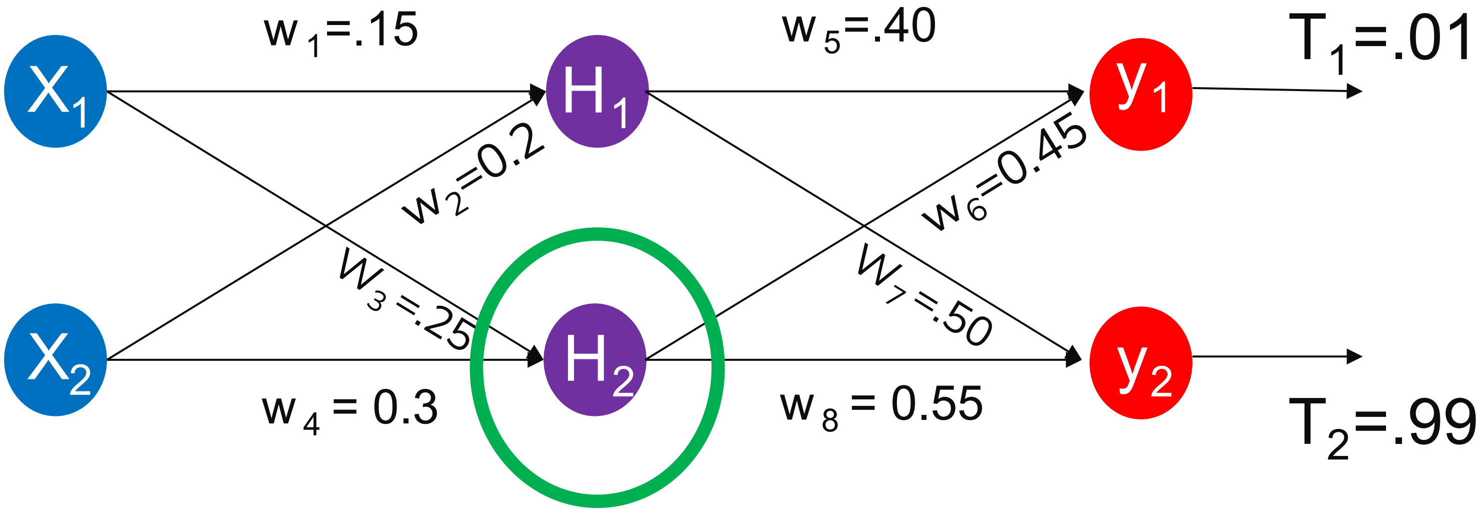

w1=.15
w5=.40
T1=.01
X1
H1
y1
w2=0.2
w6=0.45
W3=.25
W7=.50
X2
H2
y2
w4 = 0.3
w8 = 0.55
T2=.99

< T 1 >

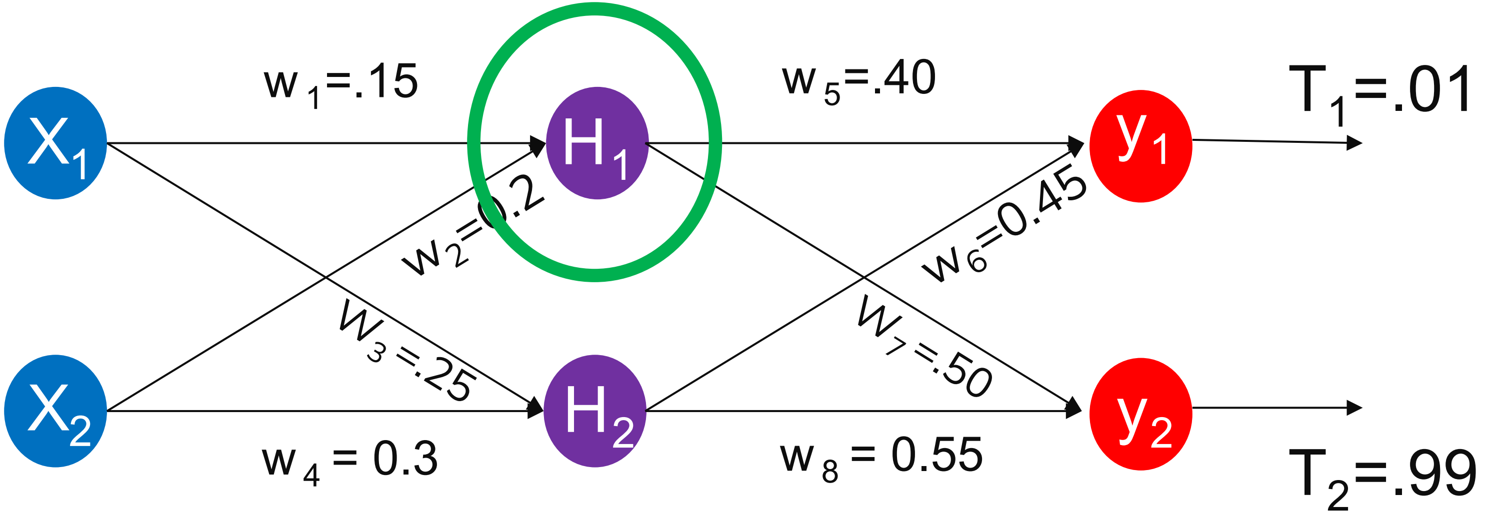


w_1=.15
w_5=.40
T_1=.01
X_1
H_1
y_1
w_2=0.2
w_6=0.45
W_3 =.25
W_7 =.50
X_2
H_2
y_2
w_4 = 0.3
w_8 = 0.55
T_2=.99

< T 1 >


X_1
X_2
H_1
H_2
y_1
y_2
w_1=.15
w_2=0.2
W_3 =.25
w_4 = 0.3
w_5=.40
w_6=0.45
W_7 =.50
w_8 = 0.55
T_1=.01
T_2=.99

<
T
1
>



w 1=.15
w 5=.40
T1=.01
X1
H1
y1
w2=0.2
w6=0.45
W3 =.25
W7 =.50
X2
H2
y2
w 4 = 0.3
w 8 = 0.55
T2=.99

Thank You