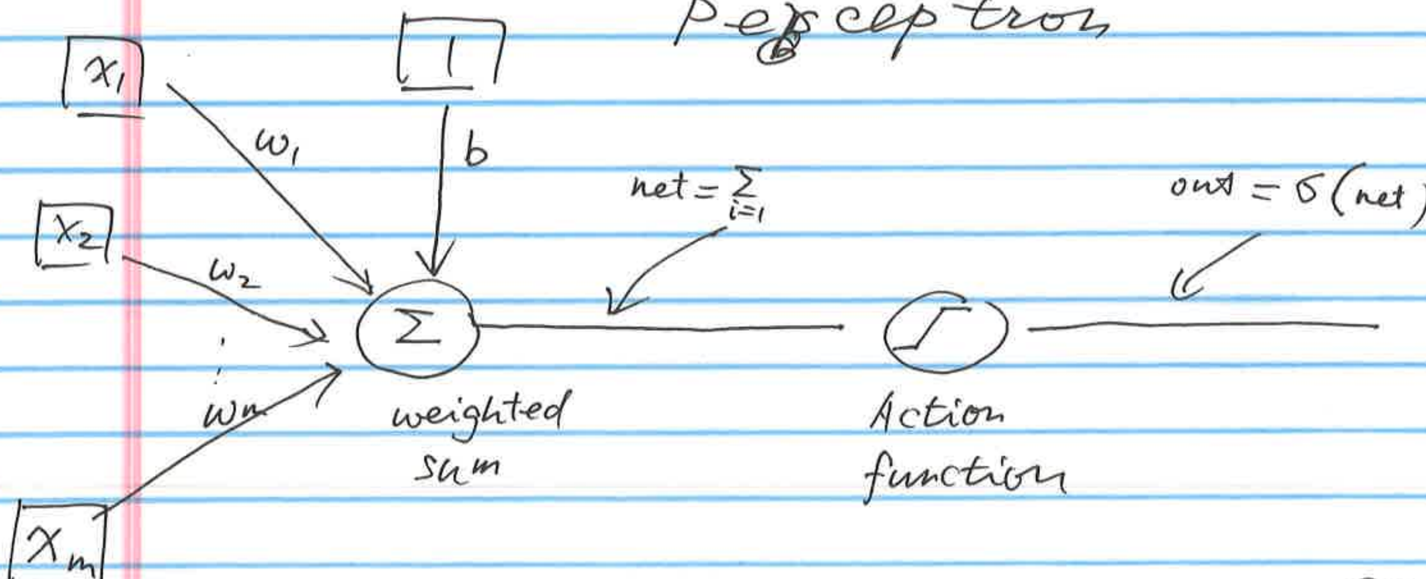
CPSC 4430/5440.

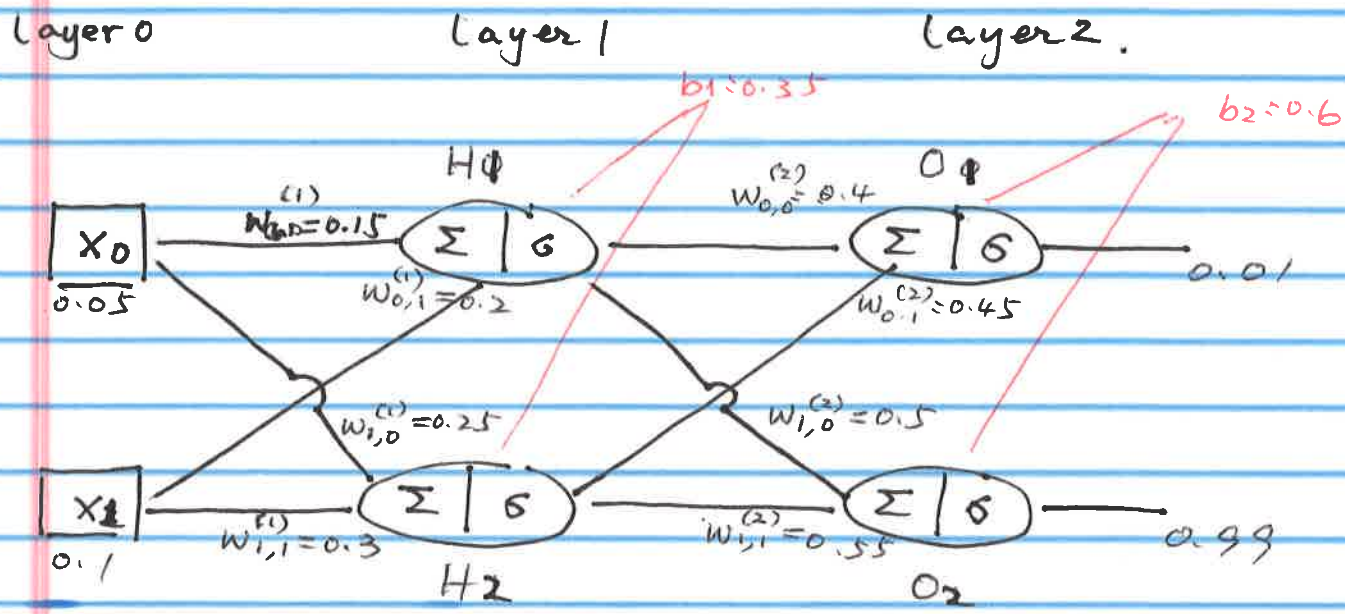
Lesson D02 (a).

Backward Propogation in Scalar.

(Fully Connected Forward Feed)

MLP : Multilayer Perceptron





Layer 0 (input Layer)

$X_0 = 0.05, \qquad X_1 = 0.1.$

Layer 1 (Hidden Layer)

$$W^{(1)} = \begin{bmatrix} w_{0,0}^{(1)} & w_{0,1}^{(1)} \\ w_{1,0}^{(1)} & w_{1,1}^{(1)} \end{bmatrix} = \begin{bmatrix} 0.15 & 0.2 \\ 0.25 & 0.3 \end{bmatrix}.$$

$$b^{(1)} = 0.35$$

Layer 2 (Output Layer)

$$W^{(2)} = \begin{bmatrix} w_{0,0}^{(2)} & w_{0,1}^{(2)} \\ w_{1,0} & w_{1,1}^{(2)} \end{bmatrix} = \begin{bmatrix} 0.4 & 0.45 \\ 0.5 & 0.55 \end{bmatrix}$$

$$b^{(2)} = 0.6.$$

Forward Pass: Compute the output of Network.

Layer 1 (Hidden layer)

$$net_{H1} = w_{0,0}^{(1)} \cdot x_1 + w_{0,1}^{(1)} \cdot x_2 + b^{(1)}$$

$$= 0.15 * 0.05 + 0.2 * 0.1 + 0.35$$

$$= 0.3775$$

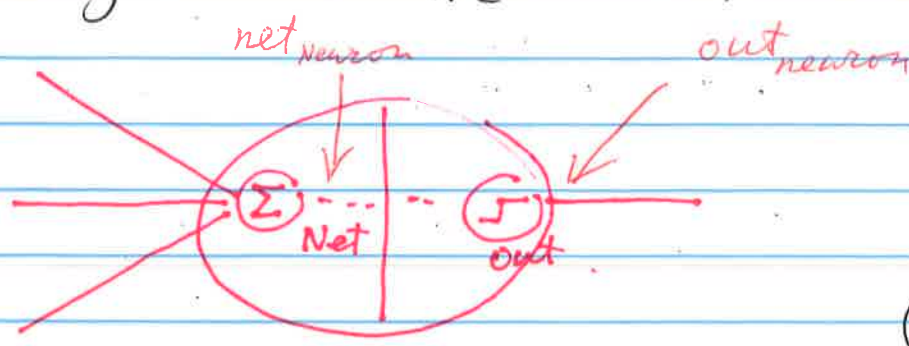
squash it into Sigmoid function

$$out_{H1} = \frac{1}{1+e^{-net_{H1}}} = \frac{1}{1+e^{-0.3775}}$$

$$= 0.59326992$$

in the same way.

we can get $out_{H2} = 0.596884378$



Forward Pass (continued)

Layer 2 (Output layer)

$$net_{O1} = w^{(2)}_{0,0} * out_{H1} + w^{(2)}_{0,1} * out_{H2} + b_2$$

$$= 0.4 * 0.593269992 + 0.45 * 0.596884378 + 0.6$$

$$= 1.105905967$$

$$out_{O1} = \frac{1}{1+e^{-net_{O1}}} = \frac{1}{1+e^{-1.105905967}}$$

$$= 0.75136507.$$

In the same way.

$$out_{O2} = 0.772928465$$

$\therefore$ The predicted output is $\begin{bmatrix} 0.75136507 \\ 0.772928465 \end{bmatrix}$

Step II : Compute the Error.

$$E_{total} = \frac{1}{2} \left\| \begin{bmatrix} 0.01 \\ 0.99 \end{bmatrix} - \begin{bmatrix} 0.75136507 \\ 0.772928465 \end{bmatrix} \right\|^2$$

$$= 0.274811083 + 0.02356$$

$$= 0.29837$$

Step III : Backward Pass.

① Compute the gradient with respect to weight.

② update weight to make predicted output closed to the actual output

Let's consider $w^{(2)}_{0,0}$ first.

$$\frac{\partial E_{total}}{\partial w^{(2)}_{0,0}}.$$

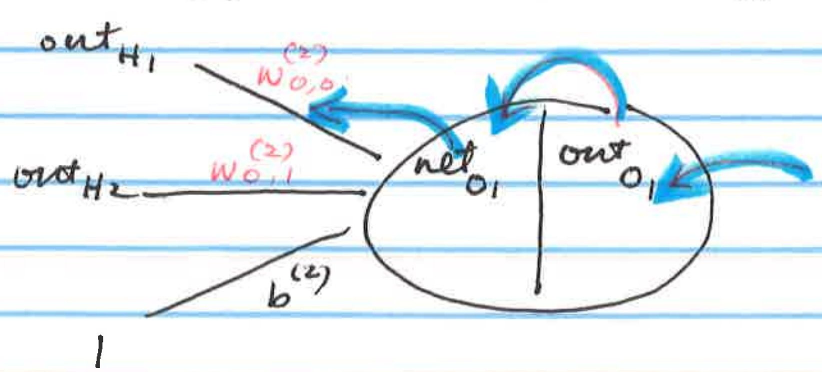
Chain Rule

$$E_{total} = E_{O1} + E_{O2}.$$

$$E_{O1} = \frac{1}{2}(target_{O1} - out_{O1})^2$$

$$out_{O1} = \frac{1}{1+e^{-net_{O1}}}.$$

$$net_{O1} = w^{(2)}_{0,0} \cdot out_{h1} + w^{(2)}_{0,1}\, out_{H2} + b^{(2)}_0$$



$$E_{O1} = \frac{1}{2}(target_{O1} - out_{O1})^2$$

$$E_{total} = E_{O1} + E_{O2}$$

using the Chain Rule.

$$\frac{\partial E_{total}}{\partial w^{(2)}_{0,0}} = \frac{\partial net_{o1}}{\partial w^{(2)}_{0,0}} * \frac{\partial out_{o1}}{\partial net_{o1}} * \frac{\partial E_{total}}{\partial out_{o1}}$$

① $E_{total} = \frac{1}{2}(target_{o1} - out_{o1})^2 + \frac{1}{2}(target_{o2} - out_{o2})^2.$

$$\Rightarrow \frac{\partial E_{total}}{\partial out_{o1}} = 2 * \frac{1}{2}(target_{o1} - out_{o1}) * (-1) + 0$$

$$= -(target_{o1} - out_{o1})$$

$$= -(0.01 - 0.75136507)$$

$$= 0.74136507.$$

$$\boxed{\begin{aligned} f(x) &= \frac{1}{1+e^{-x}} = \frac{e^x}{1+e^x}. \\ \frac{d}{dx} f(x) &= \frac{e^x(1+e^x) - e^x \cdot e^x}{(1+e^x)^2} = \frac{e^x}{(1+e^x)^2} \\ &= f(x) \cdot (1 - f(x)) \end{aligned}}$$

$$out_{o_1} = \frac{1}{1+e^{-net_{o_1}}}$$

$$\Rightarrow \frac{\partial\, out_{o_1}}{\partial\, net_{o_1}} = out_{o_1}(1 - out_{o_1})$$

$$= 0.75136507\,(1 - 0.75136507)$$

$$= 0.186815602.$$

$$\frac{\partial E_{total}}{\partial w_{0,0}^{(2)}} = \frac{\partial E_{total}}{\partial out_{O_1}} \cdot \frac{\partial out_{O_1}}{\partial net_{O_1}} \cdot \frac{\partial net_{O_1}}{\partial w_{0,0}^{(2)}}$$

$$net_{O_1} = w_{0,0}^{(2)} * out_{H_1} + w_{0,1}^{(2)} \cdot out_{H_2} + b_2 * 1$$

$$\Rightarrow \frac{\partial net_{O_1}}{\partial w_{0,0}^{(2)}} = 1 * out_{H_1} + 0 + 0.$$

$$= out_{H_1} = 0.59326992.$$

As a result.

$$\frac{\partial E_{total}}{\partial w_{0,0}^{(2)}} = 0.74136507 * 0.1868 * 0.59327$$

$$= 0.08216704.$$

In the SAME way we can compute

$$\frac{\partial E_{total}}{\partial w_{0,1}^{(2)}}, \quad \frac{\partial E_{total}}{\partial w_{1,0}^{(2)}}, \quad \frac{\partial E_{total}}{\partial w_{1,1}^{(2)}}$$

To decrease the Error, we can update the current weight for Layer 2 (output layer)

Given Learning Rate η.

$$W_{0,0}^{(2)*} = W_{0,0}^{(2)} - \eta * \frac{\partial E_{total}}{\partial W_{0,0}^{(2)}}$$

$$= 0.4 - 0.5 * 0.082167041$$

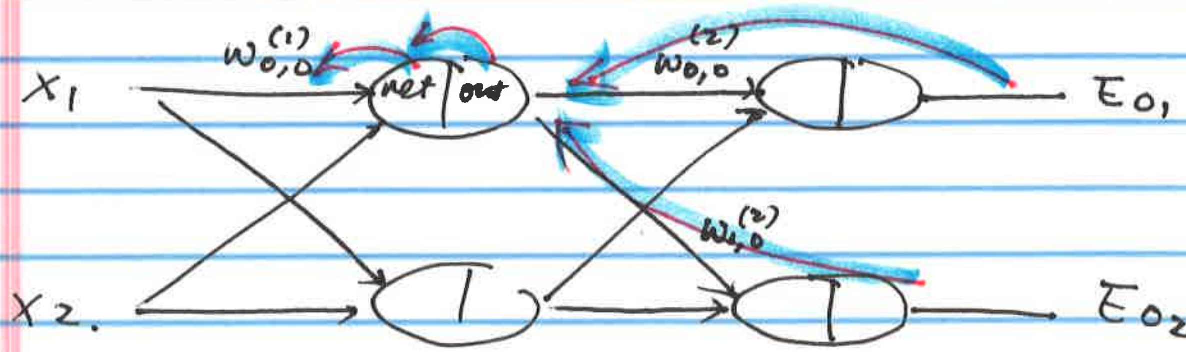
$$= 0.35891648$$

$$W_{0,1}^{(2)*} = 0.408666186.$$

$$W_{1,0}^{(2)*} = 0.511301270.$$

$$W_{1,1}^{(2)*} = 0.561370121$$

Backward Pass for the Hidden Layer (Layer 1)



$$\left\{\begin{aligned} \frac{\partial E_{total}}{\partial w_{0,0}^{(1)}} &= \frac{\partial E_{total}}{\partial out_{H_1}} * \frac{\partial out_{H_1}}{\partial net_{H_1}} * \frac{\partial net_{H_1}}{\partial w_{0,0}^{(1)}} \\ \frac{\partial E_{total}}{\partial out_{H_1}} &= \frac{\partial E_{O_1}}{\partial out_{H_1}} + \frac{\partial E_{O_2}}{\partial out_{H_1}} \end{aligned}\right.$$

$$\left\{\begin{aligned} E_{O_1} &= \frac{1}{2}(target_{O_1} - out_{O_1})^2 . \\ out_{O_1} &= \frac{1}{1+e^{-net_{O_1}}} \\ net_{O_1} &= w_{0,0}^{(2)} \cdot out_{H_1} + w_{0,1}^{(2)} \cdot out_{H_2} + b_{0}^{(2)} * 1 \end{aligned}\right.$$

$$\Rightarrow \frac{\partial net_{O_1}}{\partial out_{H_1}} = w_{0,0}^{(2)}$$

$$\frac{\partial E_{O_1}}{\partial out_{H_1}} = \frac{\partial E_{O_1}}{\partial out_{O_1}} * \frac{\partial out_{O_1}}{\partial net_{O_1}} * \frac{\partial net_{O_1}}{\partial out_{H_1}}$$

$$= \frac{\partial E_{O_1}}{\partial net_{O_1}} * \frac{\partial net_{O_1}}{\partial out_{H_1}}$$

$$= 0.13849856_2 * 0.4$$

$$= 0.055399425.$$

in the same way,

$$\frac{\partial E_{O_2}}{\partial out_{H_1}} = -0.019049119.$$

$$\therefore \frac{\partial E_{total}}{\partial out_{H_1}} = \frac{\partial E_{O_1}}{\partial out_{H_1}} + \frac{\partial E_{O_2}}{\partial out_{H_1}}$$

$$= 0.055399425 - 0.019049119$$

$$= 0.036350363$$

In summery.

$$\frac{\partial E_{total}}{\partial w_{0,0}^{(1)}} = \frac{\partial E_{total}}{\partial out_{H_1}} * \frac{\partial out_{H_1}}{\partial net_{H_1}} * \frac{\partial net_{h_1}}{\partial w_{0,0}^{(1)}}$$

$$out_{h_1} = \frac{1}{1+e^{-net_{h_1}}} \Rightarrow \frac{\partial out_{h_1}}{\partial net_{h_1}} = out_{h_1}(1-out_{h_1})$$

$$= 0.241300709.$$

$$net_{h_1} = w_{0,0}^{(1)} x_1 + w_{0,1}^{(1)} x_2 + b^{(1)} * 1.$$

$$\Rightarrow \frac{\partial net_{h_1}}{\partial w_{0,0}^{(1)}} = x_1 = 0.05$$

$$\therefore \frac{\partial E_{total}}{\partial w_{0,0}^{(1)}} = 0.036350306 * 0.241300709 * 0.05$$

$$= 0.000438568$$

Next we can update $w_{0,0}^{(1)}$

$$w_{0,0}^{(1)*} = w_{0,0}^{(1)} - \eta \cdot \frac{\partial E_{total}}{\partial w_{0,0}^{(1)}}$$

$$= 0.15 - 0.5 * 0.000438568$$

$$= 0.149780716$$

in the same way, we can update

$$w_{0,1}^{(1)*} = 0.19956143$$

$$w_{1,0}^{(1)*} = 0.24975114$$

$$w_{1,1}^{(1)*} = 0.29950229$$

So far, all the weights of network have been updated.

given observation ((0.05, 1), (0.01, 0.99))

In the beginning,

$$E_{total} = 0.298371109$$

After one epoch of training.

$$E_{total} = 0.291027924$$

After 10000 epoch. $E_{total} = 0.000035085$.