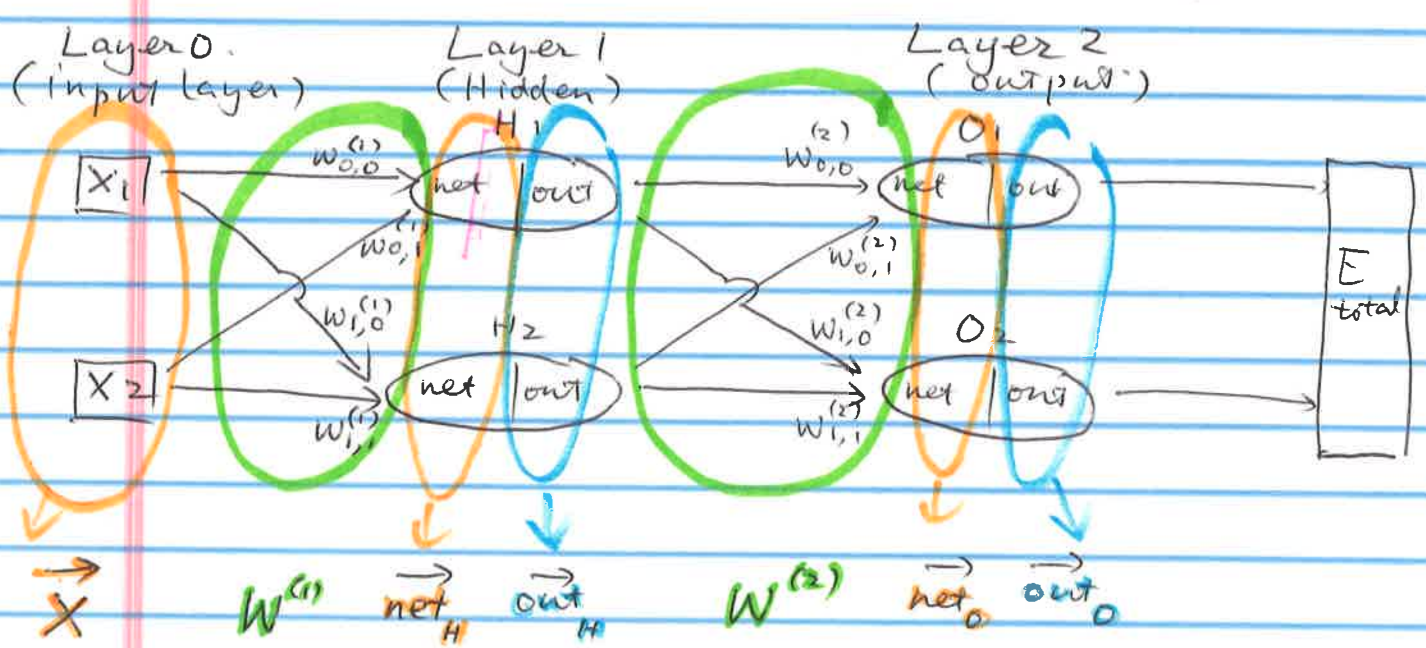
CPSC 4430/5440

Lesson D02 (b)

Matrix-version Backward Propogation



(bias is excluded for simplicity).

Given Observation.

Feature		Label (target)	
x_1	x_2	target 1	target 2
0.05	0.1	0.01	0.99

Forward Pass (Layer 1)

using $\vec{x} = \begin{bmatrix} 0.05 \\ 0.1 \end{bmatrix}$. $W^{(1)} = \begin{bmatrix} 0.15 & 0.20 \\ 0.25 & 0.30 \end{bmatrix}$.

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

$$\vec{net}_H = W^{(1)} * \vec{x} + b^{(1)} * \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

$$= \begin{bmatrix} 0.15 & 0.2 \\ 0.25 & 0.3 \end{bmatrix} \begin{bmatrix} 0.05 \\ 0.1 \end{bmatrix} + 0.35 * \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

$$= \begin{bmatrix} 0.3775 \\ \quad \end{bmatrix}.$$

$$\vec{out}_H = \sigma(\vec{net}_H) = \frac{1}{1 + e^{-net_H}}$$

$$= \begin{bmatrix} 0.59326992 \\ 0.596884378 \end{bmatrix}.$$

Forward Pass (Layer 2)

$$\vec{net}_o = W^{(2)} \times \vec{out}_H + b^{(2)} \times \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

$$= \begin{bmatrix} 0.4 & 0.45 \\ 0.5 & 0.55 \end{bmatrix} \begin{bmatrix} 0.593269992 \\ 0.596884378 \end{bmatrix} + 0.6 \times \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

$$= \begin{bmatrix} 1.105905967 \\ \underline{\hspace{3cm}} \end{bmatrix}$$

$$\vec{out}_o = \sigma(\vec{net}_o) = \frac{1}{1 + e^{-\vec{net}_o}}$$

$$= \begin{bmatrix} 0.75136507 \\ 0.772928465 \end{bmatrix}.$$

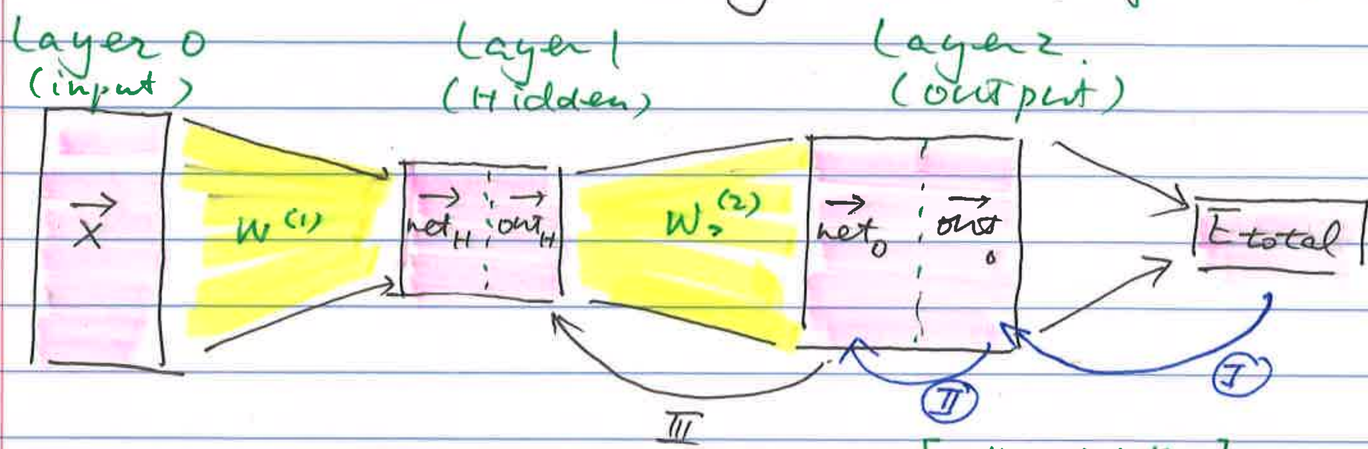
Compute The Total Error

$$E_{total} = \frac{1}{2} \| \vec{out_o} - \vec{target} \|^2$$

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

$$= 0.274811083.$$

Backward Propogation (Layer 2).



$$W^{(1)} = \begin{bmatrix} 0.15 & 0.2 \\ 0.25 & 0.3 \end{bmatrix} . \qquad W^{(2)} = \begin{bmatrix} 0.4 & 0.45 \\ 0.5 & 0.55 \end{bmatrix}$$

I $$E_{total} = \frac{1}{2} \| \vec{out}_o - \vec{target} \|^2$$

$$= \frac{1}{2} (\vec{out}_o - \vec{target})^T (\vec{out}_o - \vec{target})$$

$$\Rightarrow \frac{\partial E_{total}}{\partial \vec{out}_o} = \vec{out}_o - \vec{target} .$$

II $$\vec{out}_o = \frac{1}{1 + e^{-\vec{net}_o}} \quad (\text{sigmoid}) .$$

$$\Rightarrow \frac{\partial \vec{out}_o}{\partial \vec{net}_o} = \sum_{i=1}^{2} - E_{ii} \, \vec{out}_o \cdot \vec{e}_i$$

III

$$\overrightarrow{out}_H \qquad \overrightarrow{net}_O$$

$$\overrightarrow{net}_o = W^{(2)} \cdot \overrightarrow{out}_H + \overrightarrow{b}^{(2)}$$

$$\Longrightarrow \frac{\partial \overrightarrow{net}_o}{\partial W^{(2)}} = \begin{bmatrix} \begin{bmatrix} \overrightarrow{out}_H^{\,T} \\ 0 \end{bmatrix} \\ \begin{bmatrix} 0 \\ \overrightarrow{out}_H^{\,T} \end{bmatrix} \end{bmatrix}$$

Using Chain Rule.

$$\frac{\partial E_{total}}{\partial W^{(2)}} = \frac{\partial \overrightarrow{net}_o}{\partial W^{(2)}} \cdot \frac{\partial \overrightarrow{out}_o}{\partial \overrightarrow{net}_o} \cdot \frac{\partial E_{total}}{\partial \overrightarrow{out}_o}.$$

Matrix Calculus. (Appendix)

① $\vec{y} = \dfrac{1}{\vec{x}}$.

e.g. $\vec{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{x_1} \\ \frac{1}{x_2} \end{bmatrix} ; \vec{x}$.

$$\frac{\partial \vec{y}}{\partial \vec{x}} = \begin{bmatrix} \partial(\frac{1}{x_1})/\partial x_1 & \partial(\frac{1}{x_1})/\partial x_2 \\ \partial(\frac{1}{x_2})/\partial x_1 & \partial(\frac{1}{x_2})/\partial x_2 \end{bmatrix}.$$

$$= \begin{bmatrix} -\frac{1}{x_1^2} & 0 \\ 0 & -\frac{1}{x_2^2} \end{bmatrix}$$

$$= -\sum_{i=1}^{2} \left(E_{ii} \begin{bmatrix} \frac{1}{x_1^2} \\ \frac{1}{x_2^2} \end{bmatrix} e_i \right)$$

$$= -\sum_{i=1}^{2} \left(E_{ii} \vec{y}^{\,2} e_i \right)$$

A

(2) $\vec{y} = e^{-\vec{x}}$.

e.g. $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} e^{-x_1} \\ e^{-x_2} \end{bmatrix}$.

$$\frac{\partial \vec{y}}{\partial \vec{x}} = \begin{bmatrix} \partial e^{-x_1}/\partial x_1 & \partial e^{-x_1}/\partial x_2 \\ \partial e^{-x_2}/\partial x_1 & \partial e^{-x_2}/\partial x_2 \end{bmatrix}.$$

$$= \begin{bmatrix} -e^{-x_1} & 0 \\ 0 & -e^{-x_2} \end{bmatrix}.$$

$$= \sum_{i=1}^{2} -(E_{ii}\vec{y} \cdot e_i)$$

(E_i is a metrix has 1 in [i,i]
e_i is a vector has 1 in i)

(B)

③ $\vec{y} = A * \vec{x} + \vec{b}$.

Assume $\vec{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$. $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$.

$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. $\vec{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$.

Then $$\frac{\partial \vec{y}}{\partial A} = \begin{bmatrix} \frac{\partial y_1}{\partial A} \\ \frac{\partial y_2}{\partial A} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \frac{\partial y_1}{\partial a_{11}} & \frac{\partial y_1}{\partial a_{12}} \\ \frac{\partial y_2}{\partial a_{21}} & \frac{\partial y_2}{\partial a_{22}} \end{bmatrix} \\ \begin{bmatrix} \frac{\partial y_2}{\partial a_{11}} & \frac{\partial y_2}{\partial a_{12}} \\ \frac{\partial y_2}{\partial a_{21}} & \frac{\partial y_2}{\partial a_{22}} \end{bmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} \begin{bmatrix} \frac{\partial(a_{11}x_1 + a_{12}x_2 + b_1)}{\partial a_{11}} & \frac{\partial(a_{11}x_1 + a_{12}x_2 + b_1)}{\partial a_{12}} \\ \frac{\partial(a_{11}x_1 + a_{12}x_2 + b_1)}{\partial a_{21}} & \frac{\partial(a_{11}x_1 + a_{12}x_2 + b_1)}{\partial a_{22}} \end{bmatrix} \\ \begin{bmatrix} \frac{\partial(a_{21}x_1 + a_{22}x_2 + b_2)}{\partial a_{11}} & \frac{\partial(a_{21}x_1 + a_{22}x_2 + b_2)}{\partial a_{12}} \\ \frac{\partial(a_{21}x_1 + a_{22}x_2 + b_2)}{\partial a_{21}} & \frac{\partial(a_{21}x_1 + a_{22}x_2 + b_2)}{\partial a_{22}} \end{bmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} \begin{bmatrix} x_1 & x_2 \\ 0 & 0 \end{bmatrix} \\ \begin{bmatrix} 0 & 0 \\ x_1 & x_2 \end{bmatrix} \end{bmatrix}.$$ Through generalization.

$$\frac{\partial \vec{y}}{\partial A} = \begin{bmatrix} \begin{bmatrix} \vec{x}^T \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ \begin{bmatrix} 0 \\ \vec{x}^T \\ \vdots \\ 0 \end{bmatrix} \\ \vdots \\ \begin{bmatrix} 0 \\ \vdots \\ \vec{x}^T \end{bmatrix} \end{bmatrix}.$$

c