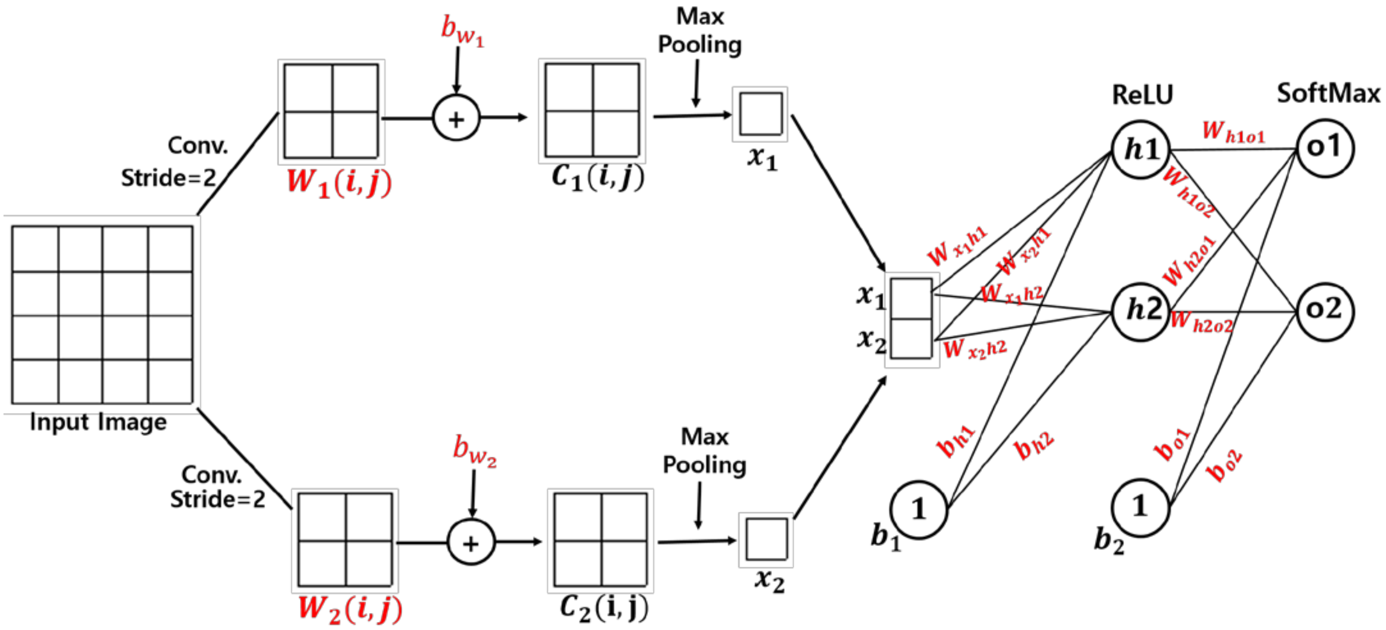
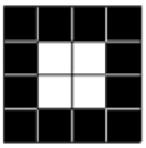
1. Convolutional Neural Network structure



2. Training set

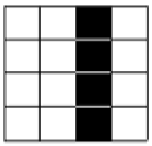
□ = 0

■ = 1



Input Image

$[o1\_out \quad o2\_out] = [1 \quad 0]$



Input Image

$[o1\_out \quad o2\_out] = [0 \quad 1]$

3. Initialize

$$W_1(i,j) = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \qquad W_2(i,j) = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$$

$b_{w_1}=0$, $b_{w_2}=0$

$$\begin{bmatrix} W_{x_1h_1} & W_{x_1h_2} \\ W_{x_2h_1} & W_{x_2h_2} \end{bmatrix} = \begin{bmatrix} 0.5 & -0.5 \\ -0.1 & 0.1 \end{bmatrix} \qquad [b_{h1} \quad b_{h2}] = [0.3 \quad 0.3]$$

$$\begin{bmatrix} W_{h1o1} & W_{h1o2} \\ W_{h2o1} & W_{h2o2} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.3 \\ 0.3 & 0.5 \end{bmatrix} \qquad [b_{o1} \quad b_{o2}] = [0.5 \quad 0.1]$$

4. Forward Propagation

a. convolution(∗ : convolution operation)

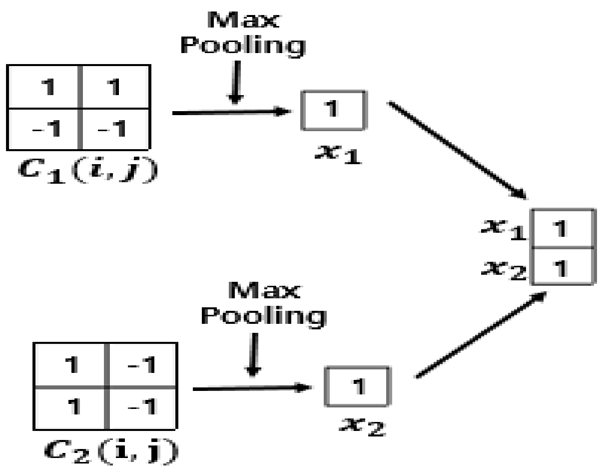
Input Image * $W_1(i,j)\ +\ b_{w_1}\ =\ C_1(i,j)$, **stride=2**

$$\underbrace{\begin{array}{|c|c|c|c|}\hline 1 & 1 & 1 & 1 \\ \hline 1 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 1 \\ \hline 1 & 1 & 1 & 1 \\ \hline \end{array}}_{\text{Input Image}} \ast \underbrace{\begin{array}{|c|c|}\hline 1 & 1 \\ \hline -1 & -1 \\ \hline \end{array}}_{W_1(i,j)} + 0 = \underbrace{\begin{array}{|c|c|}\hline 1 & 1 \\ \hline -1 & -1 \\ \hline \end{array}}_{C_1(i,j)}$$

Input Image * $W_2(i,j)\ +\ b_{W_2}\ =\ C_2(i,j)$, **stride=2**

$$\underbrace{\begin{array}{|c|c|c|c|}\hline 1 & 1 & 1 & 1 \\ \hline 1 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 1 \\ \hline 1 & 1 & 1 & 1 \\ \hline \end{array}}_{\text{Input Image}} \ast \underbrace{\begin{array}{|c|c|}\hline 1 & -1 \\ \hline 1 & -1 \\ \hline \end{array}}_{W_2(i,j)} + 0 = \underbrace{\begin{array}{|c|c|}\hline 1 & -1 \\ \hline 1 & -1 \\ \hline \end{array}}_{C_2(\mathrm{i,j})}$$

b. Max pooling & Vectorization



c. Fully connected

$$[x_1 \quad x_2] \text{ X } \begin{bmatrix} W_{x_1h1} & W_{x_1h2} \\ W_{x_2h1} & W_{x_2h2} \end{bmatrix} + [b_{h1} \quad b_{h2}] = [h1\_in \quad h2\_in]$$

$$\textbf{ReLU}([h1\_in \quad h2\_in]) = [h1\_out \quad h2\_out]$$

$$[1 \quad 1] \text{ X } \begin{bmatrix} 0.5 & -0.5 \\ -0.1 & 0.1 \end{bmatrix} + [0.3 \quad 0.3] = [0.7 \quad -0.1]$$

$$\textbf{ReLU}([0.7 \quad -0.1]) = [0.7 \quad 0]$$

$$[h1\_out \quad h2\_out] \ X \begin{bmatrix} W_{h1o1} & W_{h1o2} \\ W_{h2o1} & W_{h2o2} \end{bmatrix} + [b_{o1} \quad b_{o2}] = [o1\_in \quad o2\_in]$$

$$\textbf{softmax}([o1\_in \quad o2\_in]) = [o1\_out \quad o2\_out]$$

$$[0.7 \quad 0] \ X \begin{bmatrix} 0.5 & 0.3 \\ 0.3 & 0.5 \end{bmatrix} + [0.5 \quad 0.1] = [0.85 \quad 0.31]$$

$$\textbf{softmax}([0.85 \quad 0.31]) = [0.6318 \quad 0.3682]$$

5. Back propagation

a. Total error(cross-entropy)

$$crossentopy = -\sum_{i=1}^{2}\left(y_i \cdot log(oi\_out)\right) + ((1-y_i)\cdot log(1-oi\_out))$$

$$E_{total} = -\left(1\cdot log(0.6318) + 1\cdot log(1-0.3682)\right) = 0.9184$$

y_i : *true value* , oi_out: *predicted value*

5-1) $\Delta W_{h_i o_j}$

a. Total error and oi_out

$$\frac{\partial E_{total}}{\partial o1\_out} = \frac{\partial(-1\cdot(y_1\cdot log(o1\_out)+(1-y_1)\cdot log(1-o1\_out)))}{\partial o1\_out} = -1\cdot(y_1\frac{1}{o1\_out}+(1-y_1)\frac{-1}{1-o1\_out})$$

$$\begin{bmatrix}\frac{\partial E_{total}}{\partial o1\_out}\\ \frac{\partial E_{total}}{\partial o2\_out}\end{bmatrix} = \begin{bmatrix}-1\cdot(y_1\frac{1}{o1\_out}+(1-y_1)\frac{-1}{1-o1\_out})\\ -1\cdot(y_2\frac{1}{o2\_out}+(1-y_2)\frac{-1}{1-o2\_out})\end{bmatrix} = \begin{bmatrix}-1.5828\\ 1.5828\end{bmatrix}$$

b. oi_out and oi_in

$$Softmax = \frac{e^{x_a}}{\sum_{a=1}^{n} e^{x_a}} = \frac{e^{x_1}}{e^{x_1}+e^{x_2}},\ \frac{\partial(Softmax)}{\partial x_1} = \frac{e^{x_1}e^{x_2}}{(e^{x_1}+e^{x_2})^2}$$

$$\begin{bmatrix}\frac{\partial o1\_out}{\partial o1\_in}\\ \frac{\partial o2\_out}{\partial o2\_in}\end{bmatrix} = \begin{bmatrix}\frac{e^{o1\_in}e^{o2\_in}}{(e^{o1\_in}+e^{o2\_in})^2}\\ \frac{e^{o1\_in}e^{o2\_in}}{(e^{o1\_in}+e^{o2\_in})^2}\end{bmatrix} = \begin{bmatrix}0.2326\\ 0.2326\end{bmatrix}$$

c. oi_in and $W_{h_i o_j}$

$$\frac{\partial o1\_in}{\partial W_{h1o1}} = \frac{\partial(h_1\_out \cdot W_{h1o1} + h_2\_out \cdot W_{h2o1} + b_2 \cdot b_{o1})}{\partial W_{h1o1}} = h_1\_out$$

By symmetry, $$\begin{bmatrix} \frac{\partial o1\_in}{\partial W_{h1o1}} \\ \frac{\partial o1\_in}{\partial W_{h2o1}} \\ \frac{\partial o1\_in}{\partial b_{o1}} \end{bmatrix} = \begin{bmatrix} \frac{\partial o2\_in}{\partial W_{h1o2}} \\ \frac{\partial o2\_in}{\partial W_{h2o2}} \\ \frac{\partial o2\_in}{\partial b_{o2}} \end{bmatrix} = \begin{bmatrix} h1\_out \\ h2\_out \\ b_2 \end{bmatrix} = \begin{bmatrix} 0.7 \\ 0 \\ 1 \end{bmatrix}$$

d. Gradient $\Delta W_{h_i o_j}$

By chain rule, $$\frac{\partial E_{total}}{\partial W_{h1o1}} = \frac{\partial E_{total}}{\partial o1\_out} \cdot \frac{\partial o1\_out}{\partial o1\_in} \cdot \frac{\partial o1\_in}{\partial W_{h1o1}}$$

By symmetry,

$$\Delta W_{hioj} = \begin{bmatrix} \frac{\partial E_{total}}{\partial W_{h1o1}} & \frac{\partial E_{total}}{\partial W_{h1o2}} \\ \frac{\partial E_{total}}{\partial W_{h2o1}} & \frac{\partial E_{total}}{\partial W_{h2o2}} \\ \frac{\partial E_{total}}{\partial b_{o1}} & \frac{\partial E_{total}}{\partial b_{o2}} \end{bmatrix} = \begin{bmatrix} \frac{\partial E_{total}}{\partial o1\_out} \cdot \frac{\partial o1\_out}{\partial o1\_in} \cdot \frac{\partial o1\_in}{\partial W_{h1o1}} & \frac{\partial E_{total}}{\partial o2\_out} \cdot \frac{\partial o2\_out}{\partial o2\_in} \cdot \frac{\partial o2\_in}{\partial W_{h1o2}} \\ \frac{\partial E_{total}}{\partial o1\_out} \cdot \frac{\partial o1\_out}{\partial o1\_in} \cdot \frac{\partial o1\_in}{\partial W_{h2o1}} & \frac{\partial E_{total}}{\partial o2\_out} \cdot \frac{\partial o2\_out}{\partial o2\_in} \cdot \frac{\partial o2\_in}{\partial W_{h2o2}} \\ \frac{\partial E_{total}}{\partial o1\_out} \cdot \frac{\partial o1\_out}{\partial o1\_in} \cdot \frac{\partial o1\_in}{\partial b_{o1}} & \frac{\partial E_{total}}{\partial o2\_out} \cdot \frac{\partial o2\_out}{\partial o2\_in} \cdot \frac{\partial o2\_in}{\partial b_{o2}} \end{bmatrix}$$

$$= \begin{bmatrix} \Delta W_{h1o1} & \Delta W_{h1o2} \\ \Delta W_{h2o1} & \Delta W_{h2o2} \\ \Delta b_{o1} & \Delta b_{o2} \end{bmatrix} = \begin{bmatrix} -1.5828 \cdot 0.2326 \cdot 0.7 & 1.5828 \cdot 0.2326 \cdot 0.7 \\ -1.5828 \cdot 0.2326 \cdot 0 & 1.5828 \cdot 0.2326 \cdot 0 \\ -1.5828 \cdot 0.2326 \cdot 1 & 1.5828 \cdot 0.2326 \cdot 1 \end{bmatrix} = \begin{bmatrix} -0.2577 & 0.2577 \\ 0 & 0 \\ -0.3682 & 0.3682 \end{bmatrix}$$

5-2) $\Delta W_{x_i h_j}$

a. hi_out and hi_in

$$ReLU = \max(0, x), \quad \frac{\partial(ReLU)}{\partial x} = \begin{bmatrix} (x>0);1 \\ (x<0);0 \end{bmatrix}, \quad \text{So} \begin{bmatrix} \frac{\partial h1\_out}{\partial h1\_in} \\ \frac{\partial h2\_out}{\partial h2\_in} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

b. hi_in and $W_{x_i h_j}$

$$\frac{\partial h1\_in}{\partial W_{x1h1}} = \frac{\partial (x1 \cdot W_{x1h1} + x2 \cdot W_{x2h1} + b_1 \cdot b_{h1})}{\partial W_{x1h1}} = x1$$

By symmetry, $$\begin{bmatrix} \frac{\partial h1\_in}{\partial W_{x1h1}} \\ \frac{\partial h1\_in}{\partial W_{x2h1}} \\ \frac{\partial h1\_in}{\partial b_{h1}} \end{bmatrix} = \begin{bmatrix} \frac{\partial h2\_in}{\partial W_{x1h2}} \\ \frac{\partial h2\_in}{\partial W_{x2h2}} \\ \frac{\partial h2\_in}{\partial b_{h2}} \end{bmatrix} = \begin{bmatrix} x1 \\ x2 \\ b_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

c. Gradient $\Delta W_{x_i h_j}$

By chain rule, $$\frac{\partial E_{total}}{\partial W_{x1h1}} = \frac{\partial E_{total}}{\partial h1\_out} \cdot \frac{\partial h1\_out}{\partial h1\_in} \cdot \frac{\partial h1\_in}{\partial W_{x1h1}}$$

By symmetry,

$$\Delta W_{xihj} = \begin{bmatrix} \frac{\partial E_{total}}{\partial W_{x1h1}} & \frac{\partial E_{total}}{\partial W_{x1h2}} \\ \frac{\partial E_{total}}{\partial W_{x2h1}} & \frac{\partial E_{total}}{\partial W_{x2h2}} \\ \frac{\partial E_{total}}{\partial b_{h1}} & \frac{\partial E_{total}}{\partial b_{h2}} \end{bmatrix} = \begin{bmatrix} \frac{\partial E_{total}}{\partial h1\_out} \cdot \frac{\partial h1\_out}{\partial h1\_in} \cdot \frac{\partial h1\_in}{\partial W_{x1h1}} & \frac{\partial E_{total}}{\partial h2\_out} \cdot \frac{\partial h2\_out}{\partial h2\_in} \cdot \frac{\partial h2\_in}{\partial W_{x1h2}} \\ \frac{\partial E_{total}}{\partial h1\_out} \cdot \frac{\partial oh\_out}{\partial h1\_in} \cdot \frac{\partial h1\_in}{\partial W_{x2h1}} & \frac{\partial E_{total}}{\partial h2\_out} \cdot \frac{\partial h2\_out}{\partial h2\_in} \cdot \frac{\partial h2\_in}{\partial W_{x2h2}} \\ \frac{\partial E_{total}}{\partial h1\_out} \cdot \frac{\partial h1\_out}{\partial h1\_in} \cdot \frac{\partial h1\_in}{\partial b_{h1}} & \frac{\partial E_{total}}{\partial h2\_out} \cdot \frac{\partial h2\_out}{\partial h2\_in} \cdot \frac{\partial h2\_in}{\partial b_{h2}} \end{bmatrix}$$

, Where

$$\begin{bmatrix} \frac{\partial E_{total}}{\partial h1\_out} \\ \frac{\partial E_{total}}{\partial h2\_out} \end{bmatrix} = \begin{bmatrix} \frac{\partial E_{total}}{\partial o1\_out}\cdot\frac{\partial o1\_out}{\partial o1\_in}\cdot\frac{\partial o1\_in}{\partial h1\_out} + \frac{\partial E_{total}}{\partial o2\_out}\cdot\frac{\partial o2\_out}{\partial o2\_in}\cdot\frac{\partial o2\_in}{\partial h1\_out} \\ \frac{\partial E_{total}}{\partial o1\_out}\cdot\frac{\partial o1\_out}{\partial o1\_in}\cdot\frac{\partial o1\_in}{\partial h2\_out} + \frac{\partial E_{total}}{\partial o2\_out}\cdot\frac{\partial o2\_out}{\partial o2\_in}\cdot\frac{\partial o2\_in}{\partial h2\_out} \end{bmatrix}$$

$$= \begin{bmatrix} -1.5828\cdot 0.2326\cdot W_{h1o1} + 1.5828\cdot 0.2326\cdot W_{h1o2} \\ -1.5828\cdot 0.2326\cdot W_{h2o1} + 1.5828\cdot 0.2326\cdot W_{h2o2} \end{bmatrix} = \begin{bmatrix} -0.0737 \\ 0.0737 \end{bmatrix}$$

So, $\Delta W_{xihj} = \begin{bmatrix} \Delta W_{x1h1} & \Delta W_{x1h2} \\ \Delta W_{x2h1} & \Delta W_{x2h2} \\ \Delta b_{h1} & \Delta b_{h2} \end{bmatrix} = \begin{bmatrix} -0.0737\cdot 1\cdot 1 & 0.0737\cdot 0\cdot 1 \\ -0.0737\cdot 1\cdot 1 & 0.0737\cdot 0\cdot 1 \\ -0.0737\cdot 1\cdot 1 & 0.0737\cdot 0\cdot 1 \end{bmatrix} = \begin{bmatrix} -0.0737 & 0 \\ -0.0737 & 0 \\ -0.0737 & 0 \end{bmatrix}$

5-3) Δx_i

By chain rule

$$\frac{\partial E_{total}}{\partial x1} = \begin{aligned} & \frac{\partial E_{total}}{\partial o1\_out}\cdot\frac{\partial o1\_out}{\partial o1\_in}\cdot\frac{\partial o1\_in}{\partial h1\_out}\cdot\frac{\partial h1\_out}{\partial h1\_in}\cdot\frac{\partial h1\_in}{\partial x1} \\ & +\frac{\partial E_{total}}{\partial o1\_out}\cdot\frac{\partial o1\_out}{\partial o1\_in}\cdot\frac{\partial o1\_in}{\partial h2\_out}\cdot\frac{\partial h2\_out}{\partial h2\_in}\cdot\frac{\partial h2\_in}{\partial x1} \\ & +\frac{\partial E_{total}}{\partial o2\_out}\cdot\frac{\partial o2\_out}{\partial o2\_in}\cdot\frac{\partial o2\_in}{\partial h1\_out}\cdot\frac{\partial h1\_out}{\partial h1\_in}\cdot\frac{\partial h1\_in}{\partial x1} \\ & +\frac{\partial E_{total}}{\partial o2\_out}\cdot\frac{\partial o2\_out}{\partial o2\_in}\cdot\frac{\partial o2\_in}{\partial h2\_out}\cdot\frac{\partial h2\_out}{\partial h2\_in}\cdot\frac{\partial h2\_in}{\partial x1} \end{aligned}$$

$$= \frac{\partial h1\_out}{\partial h1\_in}\cdot\frac{\partial h1\_in}{\partial x1}\left(\frac{\partial E_{total}}{\partial o1\_out}\cdot\frac{\partial o1\_out}{\partial o1\_in}\cdot\frac{\partial o1\_in}{\partial h1\_out} + \frac{\partial E_{total}}{\partial o2\_out}\cdot\frac{\partial o2\_out}{\partial o2\_in}\cdot\frac{\partial o2\_in}{\partial h1\_out} \right)$$

$$+ \frac{\partial h2\_out}{\partial h2\_in}\cdot\frac{\partial h2\_in}{\partial x1}\left(\frac{\partial E_{total}}{\partial o1\_out}\cdot\frac{\partial o1\_out}{\partial o1\_in}\cdot\frac{\partial o1\_in}{\partial h2\_out} + \frac{\partial E_{total}}{\partial o2\_out}\cdot\frac{\partial o2\_out}{\partial o2\_in}\cdot\frac{\partial o2\_in}{\partial h2\_out} \right)$$

$$= \frac{\partial E_{total}}{\partial h1\_out} \cdot \frac{\partial h1\_out}{\partial h1\_in} \cdot \frac{\partial h1\_in}{\partial x1} + \frac{\partial E_{total}}{\partial h2\_out} \cdot \frac{\partial h2\_out}{\partial h2\_in} \cdot \frac{\partial h2\_in}{\partial x1}$$

By symmetry,

$$\Delta x_i = \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = \begin{bmatrix} \frac{\partial E_{total}}{\partial h1\_out} \cdot \frac{\partial h1\_out}{\partial h1\_in} \cdot \frac{\partial h1\_in}{\partial x1} + \frac{\partial E_{total}}{\partial h2\_out} \cdot \frac{\partial h2\_out}{\partial h2\_in} \cdot \frac{\partial h2\_in}{\partial x1} \\ \frac{\partial E_{total}}{\partial h1\_out} \cdot \frac{\partial h1\_out}{\partial h1\_in} \cdot \frac{\partial h1\_in}{\partial x2} + \frac{\partial E_{total}}{\partial h2\_out} \cdot \frac{\partial h2\_out}{\partial h2\_in} \cdot \frac{\partial h2\_in}{\partial x2} \end{bmatrix},$$

where $\frac{\partial h1\_in}{\partial x1} = W_{x1h1}$

$$= \begin{bmatrix} -0.0737 \cdot 0.2326 \cdot 0.5 + 0.0737 \cdot 0.2326 \cdot (-0.5) \\ -0.0737 \cdot 0.2326 \cdot (-0.1) + 0.0737 \cdot 0.2326 \cdot (0.1) \end{bmatrix} = \begin{bmatrix} -0.0171 \\ 0.0034 \end{bmatrix}$$

5-4) $\Delta C_i(k,l)$

Let x1's value came from C1(1,1), and x2's value came from C2(1,1).

Then, $\Delta C_1(i,j) = \begin{array}{|c|c|} \hline \Delta x1 & 1 \\ \hline -1 & -1 \\ \hline \end{array}$ $\Delta C_2(i,j) = \begin{array}{|c|c|} \hline \Delta x2 & -1 \\ \hline 1 & -1 \\ \hline \end{array}$

5-5) $\Delta W_i(k,l)$

Let Input image be $I(i,j)$, then

$$\Delta C_1(1,1)=I(1,1)\cdot W_1(1,1)+I(1,2)\cdot W_1(1,2)+I(2,1)\cdot W_1(2,1)+I(2,2)\cdot W_1(2,2)+b_{w_1}=\Delta x1$$

So, $$\frac{\partial E_{total}}{\partial W_1(1,1)}=\frac{\partial E_{total}}{\partial x1}\cdot\frac{\partial x1}{\partial W_1(1,1)}=\Delta x1\cdot I(1,1),\ \frac{\partial E_{total}}{\partial b_{w_1}}=\Delta x1\cdot 1$$

By symmetry,

$$\Delta W_1(i,j)=\begin{bmatrix}\Delta W_1(1,1)\\ \Delta W_1(2,1)\\ \Delta W_1(1,2)\\ \Delta W_1(2,2)\end{bmatrix}=\begin{bmatrix}\Delta x1\cdot I(1,1)\\ \Delta x1\cdot I(2,1)\\ \Delta x1\cdot I(1,2)\\ \Delta x1\cdot I(2,2)\end{bmatrix}=\begin{bmatrix}-0.0171\\ -0.0171\\ -0.0171\\ 0\end{bmatrix},\ \Delta W_2(i,j)=\begin{bmatrix}\Delta W_2(1,1)\\ \Delta W_2(2,1)\\ \Delta W_2(1,2)\\ \Delta W_2(2,2)\end{bmatrix}=\begin{bmatrix}\Delta x2\cdot I(1,1)\\ \Delta x2\cdot I(2,1)\\ \Delta x2\cdot I(1,2)\\ \Delta x2\cdot I(2,2)\end{bmatrix}=\begin{bmatrix}0.0034\\ 0.0034\\ 0.0034\\ 0\end{bmatrix}$$

$$\Delta b_{w_1}=\Delta x1=-0.0171,\ \Delta b_{w_2}=\Delta x2=0.0034$$

5-5) Update the weights(learning rate = 0.1)

$$W_1(i,j)=W_1(i,j)-0.1\cdot\Delta W_1(i,j),\ b_{w_1}=b_{W_1}-0.1\cdot\Delta b_{W_1}$$

$$W_2(i,j)=W_2(i,j)-0.1\cdot\Delta W_2(i,j),\ b_{w_2}=b_{W_2}-0.1\cdot\Delta b_{W_2}$$

$$W_{xihj}=W_{xihj}-0.1\cdot\Delta W_{xihj},\ \ b_{h_i}=b_{h_i}-0.1\cdot\Delta b_{h_i}$$

$$W_{hioj}=W_{hioj}-0.1\cdot\Delta W_{hioj},\ \ b_{o_i}=b_{o_i}-0.1\cdot\Delta b_{o_i}$$

6) 2nd forward propagation

softmax([o1_in   o2_in])

= **softmax(**$[\mathbf{0.918358} \quad \mathbf{0.262182}]) = [\mathbf{0.6584} \quad \mathbf{0.3416}]$

$$\boldsymbol{crossentopy} = -\sum_{i=1}^{2}\left(\boldsymbol{y_i} \cdot \boldsymbol{log}(\boldsymbol{oi\_out})\right) + ((\mathbf{1} - \boldsymbol{y_i}) \cdot \boldsymbol{log}(\mathbf{1} - \boldsymbol{oi\_out}))$$

$$\boldsymbol{E_{total}} = -\left(\mathbf{1} \cdot \boldsymbol{log}(\mathbf{0.6584}) + \mathbf{1} \cdot \boldsymbol{log}(\mathbf{1} - \mathbf{0.3416})\right) = \mathbf{0.8359}$$

$\boldsymbol{y_i}$: $\boldsymbol{true\ value}$, $\boldsymbol{oi\_out}$: $\boldsymbol{predicted\ \ value}$

Total error is reduced by 0.0825, and will converge to 0 if it is repeated several times.