KURISETI RAVI SRI TEJA 2019CS10369

PART-B:

①

a) • Sigmoid Function

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

we have

$$\frac{\partial}{\partial x}(\sigma(x)) = \frac{\partial}{\partial x}\left(\frac{1}{1+e^{-x}}\right) = \frac{-1}{(1+e^{-x})^2} \cdot -e^{-x}$$

$$= \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1}{(1+e^{-x})} \cdot \frac{e^{-x}}{(1+e^{-x})}$$

$$= \sigma(x)(1-\sigma(x)).$$

Hence, $\frac{\partial}{\partial x}(\sigma(x)) = \sigma(x) \cdot (1-\sigma(x))$.

$$\frac{\partial}{\partial w}(\sigma(f(w))) = \frac{\partial}{\partial f}(\sigma(f(w)) \cdot \frac{\partial f}{\partial w}$$

$$\cancel{= f(w)(1-f(w)) \frac{\partial f(w)}{\partial w}} \quad \cancel{f(w) = \sigma(f(w))}$$

$$= \sigma(f(w))(1-\sigma(f(w))) \cdot \frac{\partial f(w)}{\partial w}$$

b) Hyperbolic Tangent.

$$\operatorname{Tanh}(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$$

$$\frac{\partial}{\partial x}(\operatorname{Tanh}(x)) = \frac{\partial}{\partial x}\left(\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}\right)$$

$$= \frac{(e^{x}+e^{-x})\cdot\frac{\partial}{\partial x}(e^{x}-e^{-x}) - (e^{x}-e^{-x})\frac{\partial}{\partial x}(e^{x}+e^{-x})}{(e^{x}+e^{-x})^2}$$

$$= \frac{(e^{x}+e^{-x})\cdot(e^{x}+e^{-x}) - (e^{x}-e^{-x})\cdot(e^{x}-e^{-x})}{(e^{x}+e^{-x})^2}$$

$$= 1-\left(\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}\right)^2$$

$$= 1-\operatorname{Tanh}^2(x).$$

$$\frac{\partial}{\partial w}(\operatorname{Tanh}(f(w))) = \frac{\partial}{\partial f}(\operatorname{Tanh}(f(w))\cdot\frac{\partial f(w)}{\partial w}$$

$$= (1-\operatorname{Tanh}^2(f(w)))\cdot\frac{\partial f(w)}{\partial w}.$$

c) ReLU.

$$\text{ReLU}(x) = \begin{cases} x & x > 0 \\ 0 & \text{else} \end{cases}$$

$$\frac{\partial}{\partial x} \cdot \text{ReLU}(x) = \begin{cases} 1 & x > 0 \\ 0 & \text{else} \end{cases}$$

$$\frac{\partial}{\partial w} \text{ReLU}(f(w)) = \frac{\partial}{\partial f} \text{ReLU}(f(w)) \cdot \frac{\partial f(w)}{\partial w}$$

$$= \begin{cases} \frac{\partial f(w)}{\partial w} & f(w) > 0 \\ 0 & \text{else} \end{cases}$$

2)

a) Cross-Entropy Loss:

If $(p_1, p_2, \ldots, p_n)$ and $(q_1, q_2, \ldots, q_n)$ are two probability distributions, then

$CE(p, q) = -\sum_{i=1}^{n} p_i \log q_i$ where $p = \langle p_1, p_2, \ldots, p_n \rangle$ and $q = \langle q_1, q_2, \ldots, q_n \rangle$

If there are m samples and C-classes, then

$$J = -\frac{1}{m} \sum_{j=1}^{c} \sum_{i=1}^{m} y_{ij} \log P_{ij}$$

where y_{ij} is either 0 or 1 i.e, $y_i = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \end{bmatrix}$ → one-hot encoded vector.

i.e, $\sum_j y_{ij} = 1$

and

P_{ij} is the probability distribution of i^{th} sample over all classes

i.e, $\sum_j P_{ij} = 1$

i.e, if we put $\hat{y}$ as P, we get

$$J(y,\hat{y}) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{c} y_{ij}\log(\hat{y}_{ij})$$

Usually we use a softmax on outputs to compute $\hat{y}$ i.e,

$$\hat{y}_{ij} = \frac{e^{z_{ij}}}{\sum_{k=1}^{c} e^{z_{ik}}}$$ where z_{ij} are the final outputs before softmax layer.

Now,

$$\frac{\partial J}{\partial W} = \sum\sum \frac{\partial J}{\partial z_{ij}} \frac{\partial z_{ij}}{\partial W}$$

$$\frac{\partial J}{\partial z_{ij}} = -\frac{1}{m}\sum_{u=1}^{m}\sum_{v=1}^{c} \frac{\partial}{\partial z_{ij}}(y_{uv}\log(\hat{y}_{uv}))$$

$$= -\frac{1}{m}\sum_{u=1}^{m}\sum_{v=1}^{c}\left[\frac{\partial y_{uv}}{\partial z_{ij}}\log(\hat{y}_{uv}) + y_{uv}\frac{\partial}{\partial z_{ij}}(\log\hat{y}_{uv})\right]$$

$$= -\frac{1}{m}\sum_{u=1}^{m}\sum_{v=1}^{c}\frac{y_{uv}}{\hat{y}_{uv}}\frac{\partial\hat{y}_{uv}}{\partial z_{ij}}$$ [$\because y_{uv}$ is constant for u,v]

$$\frac{\partial\hat{y}_{uv}}{\partial z_{ij}} = \frac{\partial}{\partial z_{ij}}\left(\frac{e^{z_{uv}}}{\sum_{k=1}^{c} e^{z_{uk}}}\right)$$

$$= \frac{\sum_{k=1}^{c} e^{z_{uk}}\cdot e^{z_{uv}}\cdot\frac{\partial z_{uv}}{\partial z_{ij}} - e^{z_{uv}}\sum_{k=1}^{c} e^{z_{uk}}\frac{\partial z_{uk}}{\partial z_{ij}}}{\left(\sum_{k=1}^{c} e^{z_{uk}}\right)^2}$$

$$= \hat{y}_{uv}\left(\frac{\partial z_{uv}}{\partial z_{ij}} - \frac{1}{\sum_{k=1}^{c} e^{z_{uk}}}\cdot\sum_{k=1}^{c} e^{z_{uk}}\frac{\partial z_{uk}}{\partial z_{ij}}\right)$$

$$= \begin{cases} 0 & u\neq i \\ \hat{y}_{uv}(1-\hat{y}_{uj}) & u=i,\ v=j \\ -\hat{y}_{uv}\hat{y}_{uj} & u=i\ \ v\neq j \end{cases}$$

$$\Rightarrow \frac{\partial J}{\partial z_{ij}} = -\frac{1}{m} \sum_{u=1}^{m} \sum_{v=1}^{c} \frac{y_{uv}}{\hat{y}_{uv}} \cdot \frac{\partial \hat{y}_{uv}}{\partial z_{ij}}$$

$$= -\frac{1}{m} \sum_{v=1}^{c} \frac{y_{iv}}{\hat{y}_{iv}} \frac{\partial \hat{y}_{iv}}{\partial z_{ij}} \quad [\because \text{zero if } u \neq i]$$

$$= -\frac{1}{m} \left[\sum_{v \neq j} \frac{y_{iv}}{\hat{y}_{iv}} - \hat{y}_{iv} \hat{y}_{ij} + \frac{y_{ij}}{\hat{y}_{ij}} \cdot \hat{y}_{ij} (1 - \hat{y}_{ij}) \right]$$

$$= \frac{1}{m} \left[\sum_{v \neq j} y_{iv} \hat{y}_{ij} - y_{ij} (1 - \hat{y}_{ij}) \right]$$

$$= \cancel{\frac{\hat{y}_{ij}}{m} \left[\sum_{v \neq j} y_{iv} - y_{ij} + y_{ij} \hat{y}_{ij} \right]}$$

$$= \frac{1}{m} \left[\sum_{v} y_{iv} \hat{y}_{ij} - y_{ij} \right]$$

$$= \frac{1}{m} \left[\hat{y}_{ij} - y_{ij} \right] \quad [\because \sum_{v} y_{iv} = 1]$$

$$\Rightarrow \frac{\partial J}{\partial W} = \sum_{j} \sum_{i} \frac{1}{m} \cdot \left[\hat{y}_{ij} - y_{ij} \right] \cdot \frac{\partial z_{ij}}{\partial W}$$

KURISETI RAVI SRI TEJA 2019CS10369.

b) Let us assume SVM to be binary classifier with output labels +1, -1.

SVM Loss is given by,

$$L(y, \hat{y}) = \max(0, 1 - y\hat{y})$$

Now for m samples we have

$$L(y, \hat{y}) = \sum_{i=1}^{m} \max(0, 1 - y_i\hat{y}_i) = \sum_{i=1}^{n} L(y_i, \hat{y}_i)$$

$$\frac{\partial L(y, \hat{y})}{\partial w} = \sum_{i=1}^{n} \frac{\partial L(y_i, \hat{y}_i)}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial w}.$$

Now,

$$\frac{\partial L(y_i, \hat{y}_i)}{\partial \hat{y}_i} = \frac{\partial}{\partial \hat{y}_i}(\max(0, 1 - y_i\hat{y}_i))$$

$$= \begin{cases} -y_i & y_i\hat{y}_i < 1 \\ 0 & \text{else.} \end{cases}$$

Therefore,

$$\frac{\partial L(y, \hat{y})}{\partial w} = \sum_{i=1}^{n} \underbrace{\frac{\partial L(y_i, \hat{y}_i)}{\partial \hat{y}_i}}_{\text{Defined above.}} \cdot \frac{\partial \hat{y}_i}{\partial w}$$

c) L1-Loss :-

We have for a sample of n-objects,

$$J(y,\hat{y}) = \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

$$\frac{\partial J}{\partial w} = \sum_{k=1}^{n} \frac{\partial J}{\partial \hat{y}_k} \cdot \frac{\partial \hat{y}_k}{\partial w}$$

$$\frac{\partial J(y,\hat{y})}{\partial \hat{y}_k} = \sum_{i=1}^{n} \frac{\partial}{\partial \hat{y}_k} |y_i - \hat{y}_i|$$

$$= \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{y_i - \hat{y}_i} \cdot - \frac{\partial \hat{y}_i}{\partial \hat{y}_k}$$

$$= \frac{|\hat{y}_k - y_k|}{\hat{y}_k - y_k}.$$

$$\frac{\partial J}{\partial w} = \sum_{k=1}^{n} \frac{|\hat{y}_k - y_k|}{\hat{y}_k - y_k} \cdot \frac{\partial \hat{y}_k}{\partial w}.$$

d) Huber-loss :-

$$L_\delta(y_i, \hat{y}_i) = \begin{cases} \frac{1}{2}(y_i - \hat{y}_i)^2 & |y_i - \hat{y}_i| \leq \delta \\ \delta(|y_i - \hat{y}_i| - \frac{\delta}{2}) & \text{else} \end{cases}$$

Let there be m samples.

We have

$$L = \sum_{j=1}^{n} L_\delta(y_j, \hat{y}_j)$$

$$\frac{\partial L}{\partial w} = \sum_{j=1}^{n} \frac{\partial}{\partial w} L_\delta(y_j, \hat{y}_j)$$

$$= \sum_{j=1}^{n} \frac{\partial L_\delta(y_j, \hat{y}_j)}{\partial \hat{y}_j} \cdot \frac{\partial \hat{y}_j}{\partial w}.$$

where,

$$\frac{\partial L_\delta(y_j, \hat{y}_j)}{\partial \hat{y}_j} = \begin{cases} \frac{1}{2} \cdot \frac{\partial (y_j - \hat{y}_j)^2}{\partial \hat{y}_j} & |y_j - \hat{y}_j| \leq \delta \\ \delta \cdot \frac{\partial}{\partial \hat{y}_j}(|y_j - \hat{y}_j| - \delta/2) & \text{else where} \end{cases}$$

$$= \begin{cases} \hat{y}_j - y_j & |y_j - \hat{y}_j| \leq \delta \\ \delta \cdot \frac{|\hat{y}_j - y_j|}{\hat{y}_j - y_j} & \text{else where.} \end{cases}$$

$$\Rightarrow \frac{\partial L}{\partial w} = \sum_{j=1}^{n} \frac{\partial L_\delta(y_j, \hat{y}_j)}{\partial \hat{y}_j} \cdot \frac{\partial \hat{y}_j}{\partial w}$$

KURISETI RAVI SRI TEJA 2019CS10369

e) L2-LOSS.

We have for a n-sample output $\hat{y}_1, \hat{y}_2 \dots \hat{y}_n$ with labelled output $y_1, y_2, \dots, y_n$.

$$L = \frac{1}{2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Let $\hat{y}_i = f(x_i, w_j)$

$$\frac{\partial L}{\partial w_j} = \frac{1}{2} \sum_{i=1}^{n} \frac{\partial}{\partial w_j} (y_i - \hat{y}_i)^2$$

$$= \frac{1}{2} \cdot 2 \cdot \sum_{i=1}^{n} (y_i - \hat{y}_i) \cdot \frac{-\partial \hat{y}_i}{\partial w_j}$$

$$= \sum_{i=1}^{n} (\hat{y}_i - y_i) \cdot \frac{\partial f(x_i, w_j)}{\partial w_j}$$

~~If $\hat{y} = f(x, w)$, we get~~

$$\cancel{\frac{\partial L}{\partial w} = \sum_{i=1}^{n} (\hat{y}_i - y_i) \frac{\partial f(x, w)}{\partial w}}.$$

f) Cosine Similarity:

Let $y = \langle y_1, y_2, \ldots, y_n \rangle$
$\hat{y} = \langle \hat{y}_1, \hat{y}_2, \ldots \hat{y}_n \rangle$

$$J(y, \hat{y}) = \cos\theta = \frac{\sum_{i=1}^{n} y_i \hat{y}_i}{\sqrt{\sum_{i=1}^{n} y_i^2}\sqrt{\sum_{i=1}^{n} \hat{y}_i^2}}$$

Let $\hat{y} = f(x, w)$

$$\frac{\partial J}{\partial w} = \sum_{i=1}^{n} \frac{\partial J}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial w}$$

$$\frac{\partial J}{\partial \hat{y}_i} = \frac{\partial}{\partial \hat{y}_i}\left(\frac{\sum_{j=1}^{n} y_j \hat{y}_j}{\sqrt{\sum_{j=1}^{n} y_j^2}\sqrt{\sum_{j=1}^{n} \hat{y}_j^2}}\right)$$

$$= \frac{1}{\sqrt{\sum_{j=1}^{n} y_j^2}} \frac{\partial}{\partial \hat{y}_i}\left(\frac{\sum_{j=1}^{n} y_j \hat{y}_j}{\sqrt{\sum_{j=1}^{n} \hat{y}_j^2}}\right)$$

$$= \frac{1}{\sqrt{\sum_{j=1}^{n} y_j^2}} \frac{\sqrt{\sum_{j=1}^{n} \hat{y}_j^2} \cdot \sum_{j=1}^{n} y_j \frac{\partial \hat{y}_j}{\partial \hat{y}_i} - \sum_{j=1}^{n} y_j \hat{y}_j \cdot \frac{\partial}{\partial \hat{y}_i}\left(\sqrt{\sum_{j=1}^{n} \hat{y}_j^2}\right)}{\sum_{j=1}^{n} \hat{y}_j^2}$$

$$= \frac{1}{\sqrt{\sum_{j=1}^{n} y_j^2}} \cdot \frac{\sqrt{\sum_{j=1}^{n} \hat{y}_j^2}\; y_i - \left(\sum_{j=1}^{n} y_j \hat{y}_j\right)\left(\frac{1}{2\sqrt{\sum_{j=1}^{n} \hat{y}_j^2}} \cdot \frac{\partial}{\partial \hat{y}_i} \sum_{j=1}^{n} \hat{y}_j^2\right)}{\sum_{j=1}^{n} \hat{y}_j^2}$$

$$= \frac{1}{\sqrt{\sum_{j=1}^{n} y_j^2}} \quad \frac{\left(\sum_{j=1}^{n} \hat{y}_j^2\right) y_i - \left(\sum_{j=1}^{n} y_j \hat{y}_j\right) \cdot \sum_{j=1}^{n} \hat{y}_j \frac{\partial \hat{y}_j}{\partial \hat{y}_i}}{\left(\sum_{j=1}^{n} \hat{y}_j^2\right)^{3/2}}$$

$$= \frac{1}{\sqrt{\sum_{j=1}^{n} y_j^2}} \cdot \frac{\left(\sum_{j=1}^{n} \hat{y}_j^2\right) y_i - \left(\sum_{j=1}^{n} y_j \hat{y}_j\right) \cdot \hat{y}_i}{\left(\sum_{j=1}^{n} \hat{y}_j^2\right)^{3/2}}$$

$$= \frac{1}{\sqrt{\sum_{j=1}^{n} y_j^2}} \quad \frac{\sum_{j=1}^{n} \hat{y}_j \left(y_i \hat{y}_j - y_j \hat{y}_i\right)}{\left(\sum_{j=1}^{n} \hat{y}_j^2\right)^{3/2}}$$

(OR)

$$= \frac{y_i}{\sqrt{\sum_{j=1}^{n} y_j^2 \sum_{j=1}^{n} \hat{y}_j^2}} - \frac{\hat{y}_i \sum_{j=1}^{n} y_j \hat{y}_j}{\sqrt{\sum_{j=1}^{n} y_j^2} \cdot \left(\sum_{j=1}^{n} \hat{y}_j^2\right)^{3/2}}$$

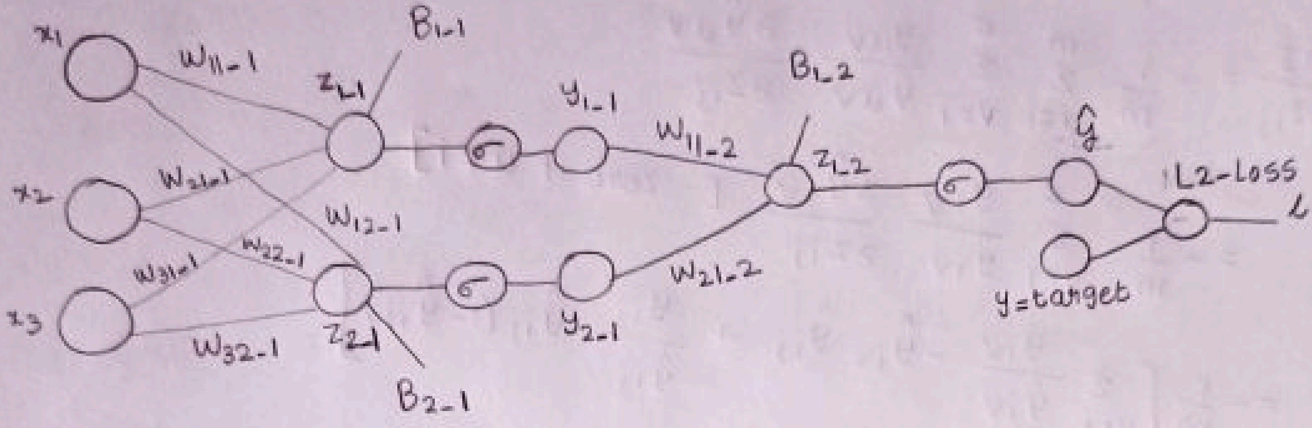
$$= \frac{y_i}{|y||\hat{y}|} - \frac{\hat{y}_i J(y, \hat{y})}{|\hat{y}|^2} \qquad |y| = \sqrt{\sum_{j=1}^{n} y_j^2} \qquad |\hat{y}| = \sqrt{\sum_{j=1}^{n} y_j^2}$$

$$= \frac{y_i |\hat{y}| - \hat{y}_i J(y, \hat{y})}{|y||\hat{y}|^2}$$

$$\Rightarrow \frac{\partial J}{\partial w} = \sum_{i=1}^{n} \frac{y_i |\hat{y}| - \hat{y}_i J(y, \hat{y})}{|y||\hat{y}|^2} \cdot \frac{\partial \hat{y}_i}{\partial w}$$

KURISETI RAVI SRI TEJA 2019CS10369

3)



Parameters:-

$x_1 = 0.02$

$x_2 = 0.03$

$x_3 = -2.06$

$\text{Target} = 0.61 = y$

$w_{11-1} = 0.36$

$w_{21-1} = 0.74$

$w_{31-1} = -0.28$

$B_{1-1} = 0.7$

$w_{12-1} = 0.29$

$w_{22-1} = -0.27$

$w_{32-1} = -0.84$

$B_{2-1} = 0.17$

$w_{11-2} = 0.4$

$w_{21-2} = 1.16$

$B_{1-2} = -0.41$

Step-1:-

Forward Computations:-

$$z_{1-1} = x_1 w_{11-1} + x_2 w_{21-1} + x_3 w_{31-1} + B_{1-1}$$

$$= \cancel{1.3062}\ (0.02)(0.36) + (0.03)(0.74) + (-2.06)(-0.28) + 0.7$$

$$= 1.3062$$

$$z_{2-1} = x_1 w_{12-1} + x_2 w_{22-1} + x_3 w_{32-1} + B_{2-1}$$

$$= (0.02)(0.29) + (0.03)(-0.27) + (-2.06)(-0.84) + 0.17$$

$$= 1.8981$$

$$y_{1-1} = \sigma(z_{1-1}) = \sigma(1.3062) = 0.7869$$

$$y_{2-1} = \sigma(z_{2-1}) = \sigma(1.8981) = 0.8697$$

$$z_{1-2} = y_{1-1} w_{11-2} + y_{2-1} w_{21-2} + B_{1-2}$$

$$= (0.7869)(0.4) + (0.8697)(1.16) + (-0.41)$$

$$= 0.9136$$

$$\hat{y} = \sigma(z_{1-2})$$

$$= \sigma(0.9136)$$

$$= 0.7137$$

$$L = \frac{1}{2}(y-\hat{y})^2 = \frac{1}{2}(0.61 - 0.7137)^2 = 0.01075 \times 0.5 = 0.0054$$

Step-2 :-

Back-Propagation :

$$\frac{\partial L}{\partial \hat{y}} = \frac{\partial}{\partial \hat{y}} \frac{1}{2}(y-\hat{y})^2 = \hat{y} - y = 0.1037$$

$$\frac{\partial \hat{y}}{\partial z_{1-2}} = \cancel{z_{1-2}(1-z_{1-2}) = (0.9136)(1-0.9136) = 0.0789} \quad \hat{y}(1-\hat{y}) = 0.2043$$

$$\cancel{\frac{\partial z_{1-2}}{\partial y_{1-1}}}$$

$$\frac{\partial z_{1-2}}{\partial w_{11-2}} = y_{1-1} = 0.7869$$

$$\frac{\partial z_{1-2}}{\partial w_{21-2}} = y_{2-1} = 0.8697$$

$$\frac{\partial z_{1-2}}{\partial B_{1-2}} = 1$$

$$\frac{\partial z_{1-2}}{\partial y_{1-1}} = w_{11-2} = 0.4$$

$$\frac{\partial z_{1-2}}{\partial y_{2-1}} = w_{21-2} = 1.16$$

$$\frac{\partial y_{1-1}}{\partial z_{1-1}} = \cancel{z_{1-1}k}\ y_{1-1}(1-y_{1-1}) = 0.7869(1-0.7869) = 0.1677$$

$$\frac{\partial y_{2-1}}{\partial z_{2-1}} = y_{2-1}(1-y_{2-1}) = 0.8697(1-0.8697) = 0.1133$$

$$\frac{\partial z_{1-1}}{\partial x_1} = w_{11-1} = 0.36 \qquad \frac{\partial z_{1-1}}{\partial w_{11-1}} = x_1 = 0.02$$

$$\frac{\partial z_{1-1}}{\partial x_2} = w_{21-1} = 0.74 \qquad \frac{\partial z_{1-1}}{\partial w_{21-1}} = x_2 = 0.03$$

$$\frac{\partial z_{1-1}}{\partial x_3} = w_{31-1} = -0.28 \qquad \frac{\partial z_{1-1}}{\partial w_{31-1}} = x_3 = -2.06$$

$$\frac{\partial z_{1-1}}{\partial B_{1-1}} = 1$$

$$\frac{\partial z_{2-1}}{\partial w_{12-1}} = x_1 = 0.02$$

$$\frac{\partial z_{2-1}}{\partial x_1} = w_{12-1} = 0.29 \qquad \frac{\partial z_{2-1}}{\partial w_{22-1}} = x_2 = 0.03$$

$$\frac{\partial z_{2-1}}{\partial x_2} = w_{22-1} = -0.27 \qquad \frac{\partial z_{2-1}}{\partial w_{32-1}} = x_3 = -2.06$$

$$\frac{\partial z_{2-1}}{\partial x_3} = w_{32-1} = -0.84$$

$$\frac{\partial z_{2-1}}{\partial B_{2-1}} = 1$$

$$\frac{\partial l}{\partial \hat{y}} = 0.1037$$

$$\frac{\partial l}{\partial z_{1-2}} = \frac{\partial l}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_{1-2}} = 0.1037 \times 0.2043 = 0.0212$$

$$\frac{\partial l}{\partial B_{1-2}} = \frac{\partial l}{\partial z_{1-2}} \cdot \frac{\partial z_{1-2}}{\partial B_{1-2}} = 0.0212 \times 1 = 0.0212$$

$$\frac{\partial l}{\partial w_{11-2}} = \frac{\partial l}{\partial z_{1-2}} \cdot \frac{\partial z_{1-2}}{\partial w_{11-2}} = 0.0212 \times 0.7869 = 0.0167$$

$$\frac{\partial l}{\partial w_{21-2}} = \frac{\partial l}{\partial z_{1-2}} \cdot \frac{\partial z_{1-2}}{\partial w_{21-2}} = 0.0212 \times 0.8697 = 0.0184$$

$$\frac{\partial l}{\partial y_{1-1}} = \frac{\partial l}{\partial z_{1-2}} \cdot \frac{\partial z_{1-2}}{\partial y_{1-1}} = 0.0212 \times 0.4 = 0.0085$$

$$\frac{\partial l}{\partial y_{2-1}} = \frac{\partial l}{\partial z_{1-2}} \cdot \frac{\partial z_{1-2}}{\partial y_{2-1}} = 0.0212 \times 1.16 = 0.0246$$

$$\frac{\partial l}{\partial z_{1-1}} = \frac{\partial l}{\partial y_{1-1}} \cdot \frac{\partial y_{1-1}}{\partial z_{1-1}} = 0.0085 \times 0.1677 = 0.0014$$

$$\frac{\partial l}{\partial z_{2-1}} = \frac{\partial l}{\partial y_{2-1}} \cdot \frac{\partial y_{2-1}}{\partial z_{2-1}} = 0.0246 \times 0.1133 = 0.0028$$

$$\frac{\partial l}{\partial B_{1-1}} = \frac{\partial l}{\partial z_{1-1}} \cdot \frac{\partial z_{1-1}}{\partial B_{1-1}} = 0.0014 \times 1 = 0.0014$$

$$\frac{\partial l}{\partial B_{2-1}} = \frac{\partial l}{\partial z_{2-1}} \cdot \frac{\partial z_{2-1}}{\partial B_{2-1}} = 0.0028 \times 1 = 0.0028.$$

$$\frac{\partial l}{\partial w_{11-1}} = \frac{\partial l}{\partial z_{1-1}} \cdot \frac{\partial z_{1-1}}{\partial w_{11-1}} = 0.0014 \times 0.02 = 2.8 \times 10^{-5}$$

$$\frac{\partial l}{\partial w_{21-1}} = \frac{\partial l}{\partial z_{1-1}} \cdot \frac{\partial z_{1-1}}{\partial w_{21-1}} = 0.0014 \times 0.03 = 4.2 \times 10^{-5}$$

$$\frac{\partial l}{\partial w_{31-1}} = \frac{\partial l}{\partial z_{1-1}} \cdot \frac{\partial z_{1-1}}{\partial w_{31-1}} = 0.0014 \times -2.06 = -0.00285$$

$\frac{\partial l}{\partial w_{22-1}}$

$$\frac{\partial l}{\partial w_{12-1}} = \frac{\partial l}{\partial z_{2-1}} \cdot \frac{\partial z_{2-1}}{\partial w_{12-1}} = 0.0028 \times 0.02 = 5.6 \times 10^{-5}$$

$$\frac{\partial l}{\partial w_{22-1}} = \frac{\partial l}{\partial z_{2-1}} \cdot \frac{\partial z_{2-1}}{\partial w_{22-1}} = 0.0028 \times 0.03 = 8.4 \times 10^{-5}$$

$$\frac{\partial l}{\partial w_{32-1}} = \frac{\partial l}{\partial z_{3-1}} \cdot \frac{\partial z_{3-1}}{\partial w_{32-1}} = 0.0028 \times -2.06 = -0.00577$$

Let, $W_1 = \begin{bmatrix} w_{11-1} & w_{12-1} \\ w_{21-1} & w_{22-1} \\ w_{31-1} & w_{32-1} \end{bmatrix}$ $\quad B_1 = \begin{bmatrix} B_{1-1} \\ B_{2-1} \end{bmatrix}$

$W_2 = \begin{bmatrix} w_{11-2} \\ w_{21-2} \end{bmatrix}$ $\quad B_2 = \begin{bmatrix} B_{1-2} \end{bmatrix}$

Then,

$$\frac{\partial l}{\partial W_1} = \begin{bmatrix} \frac{\partial l}{\partial w_{11-1}} & \frac{\partial l}{\partial w_{12-1}} \\ \frac{\partial l}{\partial w_{21-1}} & \frac{\partial l}{\partial w_{22-1}} \\ \frac{\partial l}{\partial w_{31-1}} & \frac{\partial l}{\partial w_{32-1}} \end{bmatrix} = \begin{bmatrix} 2.8 \times 10^{-5} & 5.6 \times 10^{-5} \\ 4.2 \times 10^{-5} & 8.4 \times 10^{-5} \\ -0.00285 & -0.00577 \end{bmatrix}$$

$$\frac{\partial l}{\partial W_2} = \begin{bmatrix} \frac{\partial l}{\partial w_{11-2}} \\ \frac{\partial l}{\partial w_{21-2}} \end{bmatrix} = \begin{bmatrix} 0.0167 \\ 0.0184 \end{bmatrix}$$

$$\frac{\partial l}{\partial B_1} = \begin{bmatrix} \frac{\partial l}{\partial B_{1-1}} \\ \frac{\partial l}{\partial B_{2-1}} \end{bmatrix} = \begin{bmatrix} 0.0014 \\ 0.0028 \end{bmatrix} \qquad \frac{\partial l}{\partial B_2} = \begin{bmatrix} \frac{\partial l}{\partial B_{1-2}} \end{bmatrix} = \begin{bmatrix} 0.0212 \end{bmatrix}$$

Loss $= l = 0.0054$