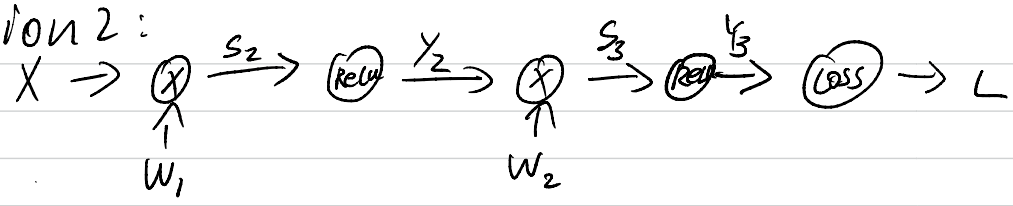
Question 2:



$$\frac{\partial L}{\partial Y_3} = \begin{bmatrix} 5 \\ 10 \end{bmatrix} = \begin{bmatrix} Y_{31} \\ Y_{32} \end{bmatrix}$$

From the forward there is no negative value so I have assum the relu layer is passthrough

$$Y_2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$Y_3 = [2\ 3] \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$$

$$Y_3 = \begin{bmatrix} Y_{21} \cdot W_{211} + Y_{22} \cdot W_{221} \\ Y_{21} \cdot W_{12} + Y_{22} \cdot W_{222} \end{bmatrix}$$

$$Y_3 = [5\ \ 10]$$

$$\frac{\partial L}{\partial W} = \begin{bmatrix} \frac{\partial L}{\partial Y_{31}} \cdot Y_{21} & \frac{\partial L}{\partial Y_{32}} \cdot Y_{21} \\ \frac{\partial L}{\partial Y_{31}} \cdot Y_{22} & \frac{\partial L}{\partial Y_{32}} \cdot Y_{22} \end{bmatrix}$$

$$\frac{\partial L}{\partial W_2} = \begin{bmatrix} 5 \cdot 2 & 10 \cdot 2 \\ 5 \cdot 3 & 10 \cdot 3 \end{bmatrix}$$

$$\frac{\partial L}{\partial W_2} = \begin{bmatrix} 10 & 20 \\ 15 & 30 \end{bmatrix}$$

$$\frac{\partial L}{\partial Y_{31}} \cdot \frac{\partial Y_{31}}{\partial W_{ij}} + \frac{\partial L}{\partial Y_{32}} \cdot \frac{\partial Y_{32}}{\partial W_{ij}}$$

$$\frac{\partial L}{\partial W_{11}} = \frac{\partial L}{\partial Y_{31}} \cdot \frac{\partial Y_{31}}{\partial W_{11}} + \frac{\partial L}{\partial Y_{32}} \cdot \frac{\partial Y_{32}}{\partial W_{11}}$$

$$= \frac{\partial L}{\partial Y_{31}} \cdot Y_{21}$$

$$\frac{\partial L}{\partial W_{12}} = \frac{\partial L}{Y_{31}} \cdot \frac{\partial Y_{31}}{\partial W_{12}} + \frac{\partial L}{\partial Y_{32}} \cdot \frac{\partial Y_{32}}{\partial W_{12}}$$

$$= \frac{\partial L}{\partial Y_{32}} \cdot Y_{21}$$

$$\frac{\partial L}{\partial W_{21}} = \frac{\partial L}{Y_{31}} \cdot \frac{\partial Y_{31}}{\partial W_{21}} = \frac{\partial L}{Y_{31}} \cdot Y_{22}$$

$$\frac{\partial L}{\partial W_{22}} = \frac{\partial L}{Y_{32}} \cdot \frac{\partial Y_{32}}{\partial W_{22}} = \frac{\partial L}{Y_{32}} \cdot Y_{22}$$

$S_2 = X \cdot W_1 \qquad S_3 = S_2 \cdot W_2 \qquad W_1 = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} \qquad W_2 = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$

$\frac{\partial L}{Y_B} = \begin{bmatrix} 5 \\ 10 \end{bmatrix} \qquad S_2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

$S_3 = \begin{bmatrix} S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} w_{211} & w_{212} \\ w_{221} & w_{222} \end{bmatrix} = \begin{bmatrix} S_{21} \cdot w_{211} + S_{22} w_{221} \\ S_{21} \cdot w_{212} + S_{22} w_{222} \end{bmatrix}$

$\frac{\partial L}{\partial Y_{31}} \cdot \frac{\partial Y_{31}}{\partial Y_{2i}} + \frac{\partial L}{\partial Y_{32}} \cdot \frac{\partial Y_{32}}{\partial Y_{2i}}$

$\frac{\partial L}{\partial Y_{31}} \cdot \frac{\partial Y_{31}}{\partial Y_{21}} + \frac{\partial L}{\partial Y_{32}} \cdot \frac{\partial Y_{32}}{\partial Y_{21}} = \frac{\partial L}{\partial Y_{31}} \cdot w_{211} + \frac{\partial L}{\partial Y_{32}} \cdot w_{212}$

$\frac{\partial L}{\partial Y_{31}} \cdot \frac{\partial Y_{31}}{\partial Y_{22}} + \frac{\partial L}{\partial Y_{32}} \cdot \frac{\partial Y_{32}}{\partial Y_{22}} = \frac{\partial L}{\partial Y_{31}} \cdot w_{221} + \frac{\partial L}{\partial Y_{32}} \cdot w_{222}$

$\frac{\partial L}{\partial Y_2} = \begin{bmatrix} 5 \cdot 1 + 10 \cdot 2 \\ 5 \cdot 1 + 10 \cdot 2 \end{bmatrix} = \begin{bmatrix} 25 \\ 25 \end{bmatrix}$

$\frac{\partial L}{\partial S_2} =$

$\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial S_2} \cdot \frac{\partial S_2}{\partial W_1} \qquad X = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$S_2 = X^T W_1 = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$

$S_2 = \begin{bmatrix} X_1 & X_2 \end{bmatrix} \begin{bmatrix} w_{111} & w_{112} \\ w_{121} & w_{122} \end{bmatrix}$

$S_2 \begin{bmatrix} X_1 w_{111} + X_2 w_{121} \\ X_1 w_{112} + X_2 w_{122} \end{bmatrix}$

$\frac{\partial L}{\partial S_{21}} \frac{\partial S_{21}}{\partial w_{1ij}} + \frac{\partial L}{\partial S_{22}} \cdot \frac{\partial S_{22}}{\partial w_{1ij}}$

$w_{111} = \frac{\partial L}{\partial S_{21}} \cdot \frac{\partial S_{21}}{\partial w_{111}} = \frac{\partial L}{\partial S_{21}} \cdot X_1$

$w_{112} = \frac{\partial L}{\partial S_{21}} \cdot \frac{\partial S_{21}}{\partial w_{112}}^{0} + \frac{\partial L}{\partial S_{22}} \cdot \frac{\partial S_{22}}{\partial w_{112}} = \frac{\partial L}{\partial S_{22}} \cdot X_1$

$\frac{\partial L}{\partial W_1} = \begin{bmatrix} 25 \cdot 1 & 25 \cdot 1 \\ 25 \cdot 2 & 25 \cdot 2 \end{bmatrix} = \begin{bmatrix} 25 & 25 \\ 50 & 50 \end{bmatrix}$

$$w_{121} = \frac{\partial L}{\partial s_{21}} \cdot \frac{\partial s_{21}}{\partial w_{121}} = \frac{\partial L}{\partial s_{21}} \cdot x_2$$

$$w_{122} = \frac{\partial L}{\partial s_{22}} \cdot \frac{\partial s_{22}}{\partial w_{122}} = \frac{\partial L}{\partial s_{22}} \cdot x_2$$

Problem 3.

$S_2 = X W_1 \quad Y_2 = \text{relu}(S_2) \quad S_3 = Y_2 W_2 \quad Y_3 = \text{relu}(S_3) \quad \text{Loss} = \frac{1}{N}\sum (\hat{y}-y)^2$

$X = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad W_1 = \begin{bmatrix} 2 & 1 \\ 0 & -1 \end{bmatrix} \quad W_2 = \begin{bmatrix} -1 & 2 \\ 1 & 2 \end{bmatrix}$

Foward: $S_2 = X W_1$

$S_2 = [1\ 2]\begin{bmatrix} 2 & 1 \\ 0 & -1 \end{bmatrix} = [2\ -1] = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$

$Y_2 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}[1\ 0] = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$

$S_3 = [2\ 0]\begin{bmatrix} -1 & 2 \\ 1 & 2 \end{bmatrix} = [-2\ \ 4] = \begin{bmatrix} -2 \\ 4 \end{bmatrix} = [$

$Y_3 = \begin{bmatrix} -2 \\ 4 \end{bmatrix}[0\ \ 1] = \begin{bmatrix} S_{31}\cdot r_1 \\ S_{32}\cdot r_2 \end{bmatrix} = \begin{bmatrix} -2\cdot 0 \\ 4\cdot 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 4 \end{bmatrix})$

$\text{Loss} = \frac{1}{2}(0)^2 + \frac{1}{2}(4)^2 = 0 + \frac{16}{2} = 8$

$\frac{\partial L}{\partial Y_3} = \frac{2}{2}(Y_3 - Y)\cdot(1) = (Y_3 - Y) \qquad \frac{\partial L}{\partial Y_3} = \begin{bmatrix} 0 \\ 4 \end{bmatrix}$

$\frac{\partial L}{\partial S_3} = \frac{\partial L}{\partial Y_3}\cdot\frac{\partial Y_3}{\partial S_3} = \frac{\partial L}{\partial Y_{31}}\cdot\frac{\partial Y_{31}}{\partial S_{3i}} + \frac{\partial L}{\partial Y_{32}}\cdot\frac{\partial Y_{32}}{\partial S_i}$

$\text{for } S_{31} = \frac{\partial L}{\partial Y_{31}}\cdot\frac{\partial Y_{31}}{\partial S_{31}} + \frac{\partial L}{\partial Y_{32}}\cdot\frac{\partial Y_{32}}{\partial S_{31}} = \frac{\partial L}{\partial Y_{31}}\cdot 0 + \frac{\partial}{\partial Y_{32}}\cdot 0 = 0$

$\text{for } S_{32} = \frac{\partial L}{\partial Y_{31}}\cdot\frac{\partial Y_{31}}{\partial S_{32}} + \frac{\partial L}{\partial Y_{32}}\cdot\frac{\partial Y_{32}}{\partial S_{32}} = \frac{\partial L}{\partial Y_{32}}\cdot 1 = \frac{\partial L}{\partial Y_{32}} = 4$

$\frac{\partial L}{\partial S_3} = \begin{bmatrix} 0 \\ 4 \end{bmatrix}$

$\frac{\partial L}{\partial W_2} = \frac{\partial L}{\partial S_3}\cdot\frac{\partial S_3}{\partial W_2} \qquad S_3 = Y_2 W_2$

$S_3 = [Y_{21}\ \ Y_{22}]\begin{bmatrix} W_{211} & W_{212} \\ W_{221} & W_{222} \end{bmatrix}$

$S_3 = \begin{bmatrix} Y_{21}\cdot W_{211} + Y_{22}\cdot W_{221} \\ Y_{21}\cdot W_{212} + Y_{22}\cdot W_{222} \end{bmatrix}$

$$\frac{\partial L}{\partial W_2} = \frac{\partial L}{\partial S_{31}} \cdot \frac{\partial S_{31}}{\partial W_{2ij}} + \frac{\partial L}{\partial S_{32}} \cdot \frac{\partial S_{32}}{\partial W_{2ij}} = \begin{bmatrix} 0 & 4\cdot 2 \\ 0 & 4\cdot 0 \end{bmatrix} = \begin{bmatrix} 0 & 8 \\ 0 & 0 \end{bmatrix}$$

$$W_{211} = \frac{\partial L}{\partial S_{31}} \cdot \frac{\partial S_{31}}{\partial W_{211}} = \frac{\partial L}{\partial S_{31}} \cdot Y_{21}$$

$$W_{212} = \frac{\partial L}{\partial S_{32}} \cdot \frac{\partial S_{32}}{\partial W_{212}} = \frac{\partial L}{\partial S_{32}} \cdot Y_{21}$$

$$W_{221} = \frac{\partial L}{\partial S_{31}} \cdot \frac{\partial S_{31}}{\partial W_{221}} = \frac{\partial L}{\partial S_{31}} \cdot Y_{22}$$

$$W_{222} = \frac{\partial L}{\partial S_{32}} \frac{\partial S_{32}}{\partial W_{222}} = \frac{\partial L}{\partial S_{32}} \cdot Y_{22}$$

$$\frac{\partial L}{\partial S_3} = \begin{bmatrix} 0 \\ 4 \end{bmatrix}$$

$$W_2 = \begin{bmatrix} -1 & 2 \\ 1 & 2 \end{bmatrix}$$

$$\frac{\partial L}{\partial Y_2} = \frac{\partial L}{\partial S_3} \cdot \frac{\partial S_3}{\partial Y_2} = \frac{\partial L}{\partial S_{31}} \cdot \frac{\partial S_{31}}{\partial Y_{21}} + \frac{\partial L}{\partial S_{32}} \cdot \frac{\partial S_{32}}{\partial Y_{21}}, \quad \frac{\partial L}{\partial S_{31}} \cdot \frac{\partial S_{31}}{\partial Y_{22}} + \frac{\partial L}{\partial S_{32}} \cdot \frac{\partial S_{32}}{\partial Y_{22}}$$

$$Y_{21} = \frac{\partial L}{\partial S_{31}} \cdot W_{211} + \frac{\partial L}{\partial S_{32}} \cdot W_{212} \qquad Y_{22} = \frac{\partial L}{\partial S_{31}} \cdot W_{221} + \frac{\partial L}{\partial S_{32}} \cdot W_{222}$$

$$\frac{\partial L}{\partial Y_2} = \begin{bmatrix} 0 + 4\cdot 2 \\ 0 + 4\cdot 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \end{bmatrix}$$

$$\begin{bmatrix} S_{21} \\ S_{22} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ r_1 & r_2 \end{bmatrix} = \begin{bmatrix} S_{21} r_1 \\ S_{22} r_2 \end{bmatrix}$$

$$\frac{\partial L}{\partial S_2} = \frac{\partial L}{\partial Y_2} \frac{\partial Y_2}{\partial S_2} = \frac{\partial L}{\partial Y_{21}} \cdot \frac{\partial Y_{21}}{\partial S_{2i}} + \frac{\partial L}{\partial Y_{22}} \cdot \frac{\partial Y_{22}}{\partial S_{2i}} = \begin{bmatrix} 8 \\ 0 \end{bmatrix} = \frac{\partial L}{\partial S_2}$$

for $S_{21} = \frac{\partial L}{\partial Y_{21}} \cdot r_1 + \frac{\partial L}{\partial Y_{22}} \cdot 0 = 8$

$S_{22} = \frac{\partial L}{\partial Y_{21}} \cdot 0 + \frac{\partial L}{\partial Y_{22}} \cdot r_2 = 0$

$S_2 = X W_1 \quad Y_2 = \text{relu}(S_2) \quad S_3 = Y_2 W_2 \quad Y_3 = \text{relu}(S_3)$

$X = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad W_1 = \begin{bmatrix} 2 & 1 \\ 0 & -1 \end{bmatrix} \quad W_2 = \begin{bmatrix} -1 & 2 \\ 1 & 2 \end{bmatrix} \quad \frac{\partial L}{\partial S_2} = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$

$$\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial S_2} \cdot \frac{\partial S_2}{\partial W_1} = \frac{\partial L}{\partial S_{21}} \cdot \frac{\partial S_{21}}{\partial W_{ij}} + \frac{\partial L}{\partial S_{22}} \cdot \frac{\partial S_{22}}{\partial W_{ij}}$$

$S_2 = X W_1$

$$S_2 = X^T W_1 = \begin{matrix} X_1 & X_2 \end{matrix} \begin{bmatrix} 2 & 1 \\ 0 & -1 \end{bmatrix} \begin{matrix} W_{111} & W_{112} \\ W_{121} & W_{122} \end{matrix}$$

$$S_2 = \begin{bmatrix} X_1 W_{111} + X_2 W_{121} \\ X_1 W_{112} + X_2 W_{122} \end{bmatrix}$$

$$W_{111} = \frac{\partial L}{\partial S_{21}} \cdot X_1$$

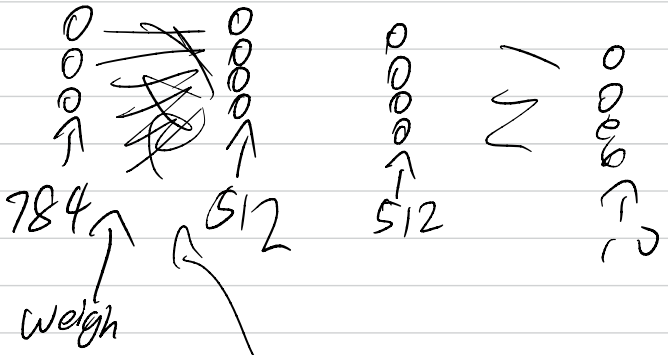
$$W_{121} = \frac{\partial L}{\partial S_{21}} \cdot X_2$$

$$W_{112} = \frac{\partial L}{\partial S_{22}} \cdot X_1$$

$$W_{122} = \frac{\partial L}{\partial S_{22}} \cdot X_2$$

$$\frac{\partial L}{\partial W_1} = \begin{bmatrix} 8 \cdot 1 & 0 \\ 8 \cdot 2 & 0 \end{bmatrix} = \begin{bmatrix} 8 & 0 \\ 16 & 0 \end{bmatrix}$$

Question 4:



$(784 \cdot 512) + 512 = 401\,920$

$(512 \cdot 512) + 512 = 262,656$ $+ = 669,706$ total parameter

$(512 \cdot 10) + 10 = 5130$

1 image = 784 input $\quad y = xw + b$

784 node (512 mutiply) + 512 node (784 addition) + 512 relu

512 node (512 mutiply) + 512 node (512 addition) + 512 relu

512 node (10 mutiply) + 10 node (512 addition) + 10 relu

$= 1,338,378$ total operation in 1 forward pass