2. $S_2 = Y_2 = \begin{bmatrix} y_{21} & y_{22} \end{bmatrix} = \overset{X}{\begin{bmatrix} 1 & 2 \end{bmatrix}} \overset{W_1}{\begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}} = \begin{bmatrix} 2 & 3 \end{bmatrix}$

$S_3 = Y_3 = \begin{bmatrix} y_{31} & y_{32} \end{bmatrix} = \overset{Y_2}{\begin{bmatrix} 2 & 3 \end{bmatrix}} \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 10 \end{bmatrix}$

$$\frac{\partial L}{\partial w_1} = ? \qquad \frac{\partial L}{\partial w_2} = ?$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial y_{21}} \times \frac{\partial y_{21}}{\partial w_1} = \frac{\partial L}{\partial y_{21}} \times \frac{\partial y_{21}}{\partial s_{21}} \times \frac{\partial s_{21}}{\partial w_1}$$

$$= \frac{\partial L}{\partial y_{21}} \times \cancelto{1}{\phi(s_{21})} \times \frac{\partial [x_1 w_1 + x_2 w_2]}{\partial w_1}$$

$$= \frac{\partial L}{\partial y_{21}} \times \cancelto{1.0}{x_1}$$

$$\frac{\partial L}{\partial y_{21}} = \frac{\partial [\frac{1}{2}(T_1 - y_{31})^2]}{\partial y_{21}} + \frac{\partial [\frac{1}{2}(T_2 - y_{32})^2]}{\partial y_{21}}$$

$$= \frac{\partial [\frac{1}{2}(T_1 - y_{31})^2]}{\partial y_{31}} \times \cancelto{1}{\frac{\partial y_{31}}{\partial s_{31}}} \times \frac{\partial s_{31}}{\partial y_{21}} + \frac{\partial [\frac{1}{2}(T_2 - y_{32})]}{\partial y_{32}} \times \cancelto{1}{\frac{\partial y_{32}}{\partial s_{32}}} \times \frac{\partial s_{32}}{\partial y_{21}}$$

$$= (y_{31} - T_1) \times \cancelto{w_5}{\frac{\partial [w_5 y_{21} + w_6 y_{22}]}{\partial y_{21}}} + (y_{32} - T_2) \times \cancelto{w_7}{\frac{\partial [w_7 y_{21} + w_6 y_{22}]}{\partial y_{21}}}$$

$$= (y_{31} - T_1) w_5 + (y_{32} - T_2) w_7$$

$$\frac{\partial L}{\partial w_1} = [(y_{31} - T_1) w_5 + (y_{32} - T_2) w_7] \times \cancelto{1.0}{x_1}$$

$$(5 - 0)1.0 + (10 - 0)2.0 = \boxed{25}$$

$$w_{1_{new}} =$$

$$\frac{\partial L}{\partial w_2} = x_2 \times \frac{\partial L}{\partial y_{21}}$$ already calculated

$$w_{2new} = [(5-0)1.0 + (10-0)2.0] \times 2 = \boxed{50}$$

$$w_{3new} = \frac{\partial L}{\partial w_3} = \cancel{x_1}^{1} \times \frac{\partial L}{\partial y_{22}}$$

$$\frac{\partial L}{\partial y_{22}} = \frac{\partial\left[\frac{1}{2}(T_1 - y_{31})^2\right]}{\partial y_{22}} + \frac{\partial\left[\frac{1}{2}(T_2 - y_{32})^2\right]}{\partial y_{22}}$$

$$= \frac{\partial\left[\frac{1}{2}(T_1 - y_{31})^2\right]}{\partial y_{31}} \times \frac{\partial y_{31}}{\partial s_{31}} \times \frac{\partial s_{31}}{\partial y_{22}} + \frac{\partial\left[\frac{1}{2}(T_2 - y_{32})^2\right]}{\partial y_{32}} \times \frac{\partial y_{32}}{\partial s_{32}} \times \frac{\partial s_{32}}{\partial y_{22}}$$

$$= (y_{31} - T_1) \times \frac{\partial[y_{21}w_5 + y_{22}w_6]}{\partial y_{22}} + (y_{32} - T_2) \times \frac{\partial[y_{21}w_7 + y_{22}w_8]}{\partial y_{22}}$$

$$= w_6(y_{31} - T_1) + w_8(y_{32} - T_2) = 1(5-0) + 2(10-0) = \boxed{25}$$

$$w_{4new} = \frac{\partial L}{\partial w_4} = x_2 \times \frac{\partial L}{\partial y_{22}} = w2 \times (1(5-0) + 2(10-0)) = \boxed{50}$$

$$w_1 = \begin{bmatrix} w_{1new} & w_{3new} \\ w_{2new} & w_{4new} \end{bmatrix} = \begin{bmatrix} 25 & 25 \\ 50 & 50 \end{bmatrix}$$

$$W_{5new} = \frac{\partial L}{\partial W_5} = y_{21} \times \frac{\partial L}{\partial y_{31}} = y_{21} \times \frac{\partial\left[\frac{1}{2}(T_1 - y_{31})^2 + \frac{1}{2}(T_2 - y_{32})^2\right]}{\partial y_{31}}$$

$$= y_{21} \times (y_{31} - T_1) = 2 \times (5-0) = \boxed{10}$$

$$W_{6new} = \frac{\partial L}{\partial W_6} = y_{22} \times \frac{\partial L}{\partial y_{31}} = y_{22} \times (y_{31} - T_1) = 3 \times (5-0) = \boxed{15}$$

$$W_{7new} = \frac{\partial L}{\partial W_7} = y_{21} \times \frac{\partial L}{\partial y_{32}} = y_{21} \times \frac{\partial\left[\frac{1}{2}(T_1 - y_{31})^2 + \frac{1}{2}(T_2 - y_{32})^2\right]}{\partial y_{32}}$$

$$= y_{21} \times (y_{32} - T_2) = 2 \times (10-0) = \boxed{20}$$

$$W_{8new} = \frac{\partial L}{\partial W_8} = y_{22} \times \frac{\partial L}{\partial y_{32}} = 3 \times (10-0) = \boxed{30}$$

$$W_2 = \begin{bmatrix} W_{5new} & W_{7new} \\ W_{6new} & W_{8new} \end{bmatrix} = \begin{bmatrix} 10 & 20 \\ 15 & 30 \end{bmatrix}$$

The gradient of W_1 and W_2 for a single iteration when calculated manually is the same as the python solution in Jupyter.

3. $S_2 = \overset{x}{\begin{bmatrix} 1 & 2 \end{bmatrix}} \overset{W_1}{\begin{bmatrix} 2 & -1 \\ 0 & -1 \end{bmatrix}} = \begin{bmatrix} 2 & -1 \end{bmatrix}$

$Y_2 = \begin{bmatrix} 2 & 0 \end{bmatrix}$

$S_3 = \overset{Y_2}{\begin{bmatrix} 2 & 0 \end{bmatrix}} \overset{W_2}{\begin{bmatrix} -1 & 2 \\ 1 & 2 \end{bmatrix}} = \begin{bmatrix} -2 & 4 \end{bmatrix}$

$Y_3 = \begin{bmatrix} 0 & 4 \end{bmatrix}$

$$W_{1new} = \frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial y_{21}} \times \cancel{\frac{\partial y_{21}}{\partial S_{21}}}^{1} \times \frac{\partial S_{21}}{\partial w_1} = \frac{\partial L}{\partial y_{21}} \times \frac{\partial [x_1 w_1 + x_2 w_2]}{\partial w_1}$$

$$= \frac{\partial L}{\partial y_{21}} \times x_1 = [(y_{31} - T_1) w_5 + (y_{32} - T_2) w_7] \times x_1$$

$$= [(0-0)-1 + (4-0)2] \times 1 = \boxed{8}$$

$$W_{2new} = \frac{\partial L}{\partial y_{21}} \times x_2 = [(0-0)-1 + (4-0)2] \times 2 = \boxed{16}$$

$$W_{3new} = \frac{\partial L}{\partial W_3} = \frac{\partial L}{\partial y_{22}} \times \frac{\partial y_{22}}{\partial S_{22}} \times \frac{\partial S_{22}}{\partial w_3} = \frac{\partial L}{\partial y_{22}} \times \cancel{\phi'(S_{22})}^{0} \times \frac{\partial S_{22}}{\partial w_3} = 0$$

$$W_{4new} = \frac{\partial L}{\partial W_4} = \frac{\partial L}{\partial y_{22}} \times \frac{\partial y_{22}}{\partial S_{22}} \times \frac{\partial S_{22}}{\partial W_4} = \frac{\partial L}{\partial y_{22}} \times \cancel{\phi'(S_{22})}^{0} \times \frac{\partial S_{22}}{\partial w_4} = 0$$

$$W_1 = \begin{bmatrix} W_{1new} & W_{3new} \\ W_{2new} & W_{4new} \end{bmatrix} = \begin{bmatrix} 8 & 0 \\ 16 & 0 \end{bmatrix}$$

$$w_{5new} = \frac{\partial L}{\partial w_5} = \frac{\partial L}{\partial y_{31}} \times \cancel{\frac{\partial y_{31}}{\partial S_{31}}}^{0} \times \frac{\partial S_{31}}{\partial w_5} = \boxed{0}$$

$$w_{6new} = \frac{\partial L}{\partial w_6} = \frac{\partial L}{\partial y_{31}} \times \cancel{\frac{\partial y_{21}}{\partial S_{31}}}^{0} \times \frac{\partial S_{31}}{\partial w_6} = \boxed{0}$$

$$w_{7new} = \frac{\partial L}{\partial w_7} = \frac{\partial L}{\partial y_{32}} \times \cancel{\frac{\partial y_{32}}{\partial S_{32}}}^{1} \times \frac{\partial S_{32}}{\partial w_7} = y_{21} \times \frac{\partial L}{\partial y_{32}}$$

$$= y_{21} \times \frac{\partial\left[\frac{1}{2}(T_1-y_{31})^2 + \frac{1}{2}(T_2-y_{32})^2\right]}{\partial y_{32}}$$

$$= y_{21} \times (y_{32}-T_2) = 2\times(4-0) = \boxed{8}$$

$$w_{8new} = \frac{\partial L}{\partial w_8} = \frac{\partial L}{\partial y_{32}} \times \cancel{\frac{\partial y_{32}}{\partial S_{32}}}^{1} \times \frac{\partial S_{32}}{\partial w_8} = y_{22} \times \frac{\partial L}{\partial y_{32}}$$

$$= y_{22} \times (y_{32}-T_2) = 0\times(4-0) = \boxed{0}$$

$$W_2 = \begin{bmatrix} w_{5new} & w_{7new} \\ w_{6new} & w_{8new} \end{bmatrix} = \begin{bmatrix} 0 & 8 \\ 0 & 0 \end{bmatrix}$$

The gradient of W_1 and W_2 for a single iteration when calculated manually is the same as the python solution in Jupyter.

For a single image:

4. input layer : 28×28
first hidden layer : 512
second hidden layer : 512
output layer : 10

Number of parameters : $(28 \times 28 \times 512) + (512 \times 512) + (512 \times 10) + (512 + 512 + 10) = 669706$

Number of operations $= (2 \times (28 \times 28 \times 512) + 512) + (2 \times (512 \times 512) + 512) + (2 \times (512 \times 10) + 10) = 1338378$