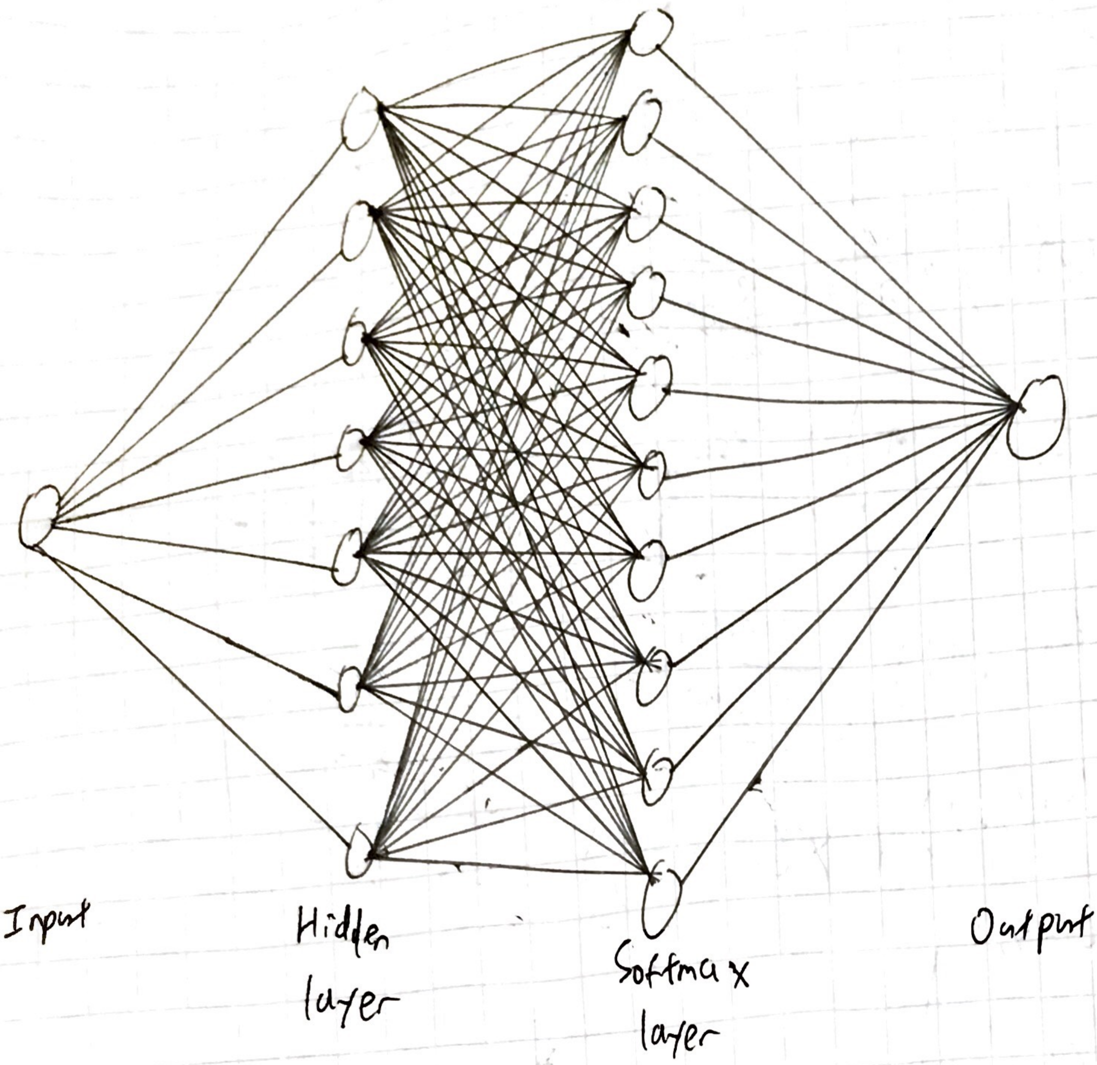

1. For $m = 1$,

$$E = -t \cdot \log(y)$$

2. $$\frac{\partial E}{\partial z^{(2)}} = \frac{\partial E}{\partial y^{(2)}} \cdot \frac{\partial y^{(2)}}{\partial z^{(2)}}$$

$$= \frac{\partial E}{\partial y^{(2)}} \cdot \frac{\partial}{\partial z_i^{(2)}} \left(\frac{e^{z_i^{(2)}}}{e^{z_1^{(2)}} + e^{z_2^{(2)}}} \right) \left(\frac{e^{z_i^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}} \right)$$

$$= -\frac{t}{y^{(2)}} \cdot \left(\frac{e^{z_i^{(2)}} \left(\sum_{k=1}^{10} e^{z_k} - e^{z_i^{(2)}} \right)}{\sum_{k=1}^{10}} \right)$$

2. $$\frac{\partial E}{\partial z^{(2)}} = \frac{\partial E}{\partial y^{(2)}} \cdot \frac{\partial y^{(2)}}{\partial z^{(2)}}$$

$$= \sum_{i=1}^{10} \frac{\partial E}{\partial y_k^{(2)}} \cdot \frac{\partial}{\partial z_i^{(2)}} \left(\frac{e^{z_i^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}} \right)$$

2. Consider softmax function:

$$s(z_i^{(2)}) = \frac{e^{z_i^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}}$$

$$\frac{\partial s(z_i^{(2)})}{\partial z_i} = \frac{e^{z_i^{(2)}}\left(\sum_{k=1}^{10} e^{z_i}\right) - e^{z_i^{(2)}}\left(e^{z_i^{(2)}}\right)}{\left(\sum_{k=1}^{10} e^{z_k^{(2)}}\right)^2}$$

$$= \frac{e^{z_i^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}} * \left(1 - \frac{e^{z_i^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}}\right)$$

$$= y_i(1-y_i) \quad -(1)$$

$$\frac{\partial s(z_i^{(2)})}{\partial z_j}, j \neq i = \frac{-e^{z_i^{(2)}} e^{z_j^{(2)}}}{\left(\sum_{k=1}^{10} e^{z_k^{(2)}}\right)^2}$$

$$= y_i y_j \quad -(2)$$

$$\frac{\partial E}{\partial z^{(2)}} = \frac{\partial E}{\partial y^{(2)}} \frac{\partial y^{(2)}}{\partial z^{(2)}}$$

$$= \frac{\partial E}{\partial y} \frac{\partial y}{\partial z^{(2)}}$$

$$= \frac{\partial}{\partial y}\left(-t \cdot \log(y)\right) \frac{\partial}{\partial z^{(2)}} \left(\frac{e^{z_1^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}}, \frac{e^{z_2^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}}, \ldots, \frac{e^{z_{10}^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}}\right)$$

$$= \left(-\frac{t_1}{y_1}, -\frac{t_2}{y_2}, \ldots, -\frac{t_{10}}{y_{10}}\right) \begin{pmatrix} \frac{\partial}{\partial z_1}\left(\frac{e^{z_1^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}}\right), & \frac{\partial}{\partial z_1}\left(\frac{e^{z_2^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}}\right), & \ldots, & \frac{\partial}{\partial z_1}\left(\frac{e^{z_{10}^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}}\right) \\ \frac{\partial}{\partial z_2}\left(\frac{e^{z_1^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}}\right), & \frac{\partial}{\partial z_2}\left(\frac{e^{z_2^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}}\right), & \ldots, & \frac{\partial}{\partial z_2}\left(\frac{e^{z_{10}^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}}\right) \\ \vdots & \vdots & & \vdots \\ \frac{\partial}{\partial z_{10}}\left(\frac{e^{z_1^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}}\right), & \frac{\partial}{\partial z_{10}}\left(\frac{e^{z_2^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}}\right), & \ldots, & \frac{\partial}{\partial z_{10}}\left(\frac{e^{z_{10}^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}}\right) \end{pmatrix}$$

$$= \left(\sum_{i=1}^{10} \frac{t_i}{y_i} \frac{\partial}{\partial z_i^{(1)}} \left(\frac{e^{z_1^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}} \right), \sum_{i=1}^{10} \frac{t_i}{y_i} \frac{\partial}{\partial z_i^{(2)}} \left(\frac{e^{z_2^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}} \right), \ldots, \sum_{i=1}^{10} \frac{t_i}{y_i} \frac{\partial}{\partial z_i^{(2)}} \left(\frac{e^{z_{10}^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}} \right) \right)$$

$\therefore$ j^{th} element of $\frac{\partial E}{\partial z^{(2)}}$, $\left(\frac{\partial E}{\partial z^{(2)}}\right)_j = \sum_{i=1}^{10} \frac{t_i}{y_i} \frac{\partial}{\partial z_i^{(2)}} \left(\frac{e^{z_j^{(2)}}}{\sum_{k=1}^{10} e^{z_k^{(2)}}} \right)$

Using results (1) and (2) above,

$$\left(\frac{\partial E}{\partial z^{(2)}}\right)_j = \sum_{i=1, i\neq j}^{10} \left[\frac{t_i}{y_i} (-y_i y_j) \right] - \frac{t_j}{y_j} (y_j (1-y_j))$$

$$= \sum_{i=1, i\neq j}^{10} (t_i y_j) + y_j t_j - t_j$$

$$= \sum_{i=1}^{10} (t_i y_j) - t_j$$

$$= y_j \sum_{i=1}^{10} t_i - t_j$$

$$= y_j - t_j$$

$$\therefore \frac{\partial E}{\partial z^{(2)}} = (y_1 - t_1, y_2 - t_2, \ldots, y_{10} - t_{10})$$

$$= (y-t)^T \text{ (shown)}$$

3. $$\frac{\partial E}{\partial y^{(1)}} = \frac{\partial E}{\partial z^{(2)}} \frac{\partial z^{(2)}}{\partial y^{(1)}}$$

$$= (y-t)^T W^{(2)} \text{ (shown)}$$

4. $$\frac{\partial E}{\partial w_{uv}^{(2)}} = \frac{\partial E}{\partial z^{(2)}} \frac{\partial z^{(2)}}{\partial w_{uv}^{(2)}}$$

$$= \frac{\partial E}{\partial z_u^{(2)}} y_v^{(1)}$$

$$= \left(\frac{\partial E}{\partial z^{(2)}}\right)_u y_v^{(1)}$$

$$= (y_u^{(2)} - t_u) y_v^{(1)} \text{ (shown)}$$

5. $\sigma(z) = \frac{1}{1+e^{-z}}$

$$\cancel{\frac{\partial \sigma(z)}{\partial z} = \frac{-(-1)(1+e^{-z})(e^{-z})}{(1+e^{-z})^2} = \frac{e^{-z}(1+e^{-z})}{(1+e^{-z})^2}}$$

$$\frac{\partial \sigma(z)}{\partial z} = \frac{-(-1)(e^{-z})}{(1+e^{-z})^2}$$

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

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

$$= \sigma(z)\,(1-\sigma(z))$$

$$\frac{\partial y^{(1)}}{\partial z^{(1)}} = \cancel{\partial\sigma(z^{(1)})}\ \frac{\partial}{\partial z^{(1)}}\left(\sigma(z^{(1)})\right)$$

$$= \begin{bmatrix} \frac{\partial\sigma(z_1^{(1)})}{\partial z_1^{(1)}}, & \frac{\partial\sigma(z_1^{(1)})}{\partial z_2^{(1)}}, & \cdots & \frac{\partial\sigma(z_1^{(1)})}{\partial z_{10}^{(1)}} \\ \frac{\partial\sigma(z_2^{(1)})}{\partial z_1^{(1)}}, & \frac{\partial\sigma(z_2^{(1)})}{\partial z_2^{(1)}}, & \cdots & \frac{\partial\sigma(z_2^{(1)})}{\partial z_{10}^{(1)}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial\sigma(z_{10}^{(1)})}{\partial z_1^{(1)}}, & \frac{\partial\sigma(z_{10}^{(1)})}{\partial z_2^{(1)}}, & \cdots, & \frac{\partial\sigma(z_{10}^{(1)})}{\partial z_{10}} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma(z_1^{(1)})(1-\sigma(z_1^{(1)})), & 0 & 0 \\ 0 & , \sigma(z_2^{(1)})(1-\sigma(z_2^{(1)})), & \cdots, 0 \\ \vdots & & \\ 0 & , 0, & \cdots, \sigma(z_{10}^{(1)})(1-\sigma(z_{10}^{(1)})) \end{bmatrix}$$

$$= \operatorname{diag}\left(y^{(1)} .* (1-y^{(1)})\right) \quad \text{(shown)}$$

6. $$\frac{\partial E}{\partial z^{(1)}} = \frac{\partial E}{\partial y^{(1)}} \frac{\partial y^{(1)}}{\partial z^{(1)}}$$

$$= (y^{(2)} - t)^T W^{(2)} \operatorname{diag}\left(y^{(1)} .* (1-y^{(1)})\right) \quad \text{(shown)}$$

7. $$\frac{\partial E}{\partial W_{uv}^{(1)}} = \frac{\partial E}{\partial z^{(1)}} \frac{\partial z^{(1)}}{\partial W_{uv}^{(1)}}$$

$$= \frac{\partial E}{\partial z_u^{(1)}} \frac{\partial z_u^{(1)}}{\partial W_{uv}^{(1)}}$$

$$= \frac{\partial E}{\partial z_u} \cancel{\frac{\partial (W_{uv}^{(1)} x_v)}{\partial W_{uv}^{(1)}}} \frac{\partial \left(\sum_i W_{ui} x_i\right)}{\partial W_{uv}}$$

$$= \frac{\partial E}{\partial z_u} x_v \quad \text{(shown)}$$