a. In our example $a_3 = \text{Softmax}(z_3)$

$\therefore$ Finding the Softmax derivatives

$$\Rightarrow \frac{\partial a^3_i}{\partial z^3_j} = \frac{\partial \left(e^{z^3_i} / (e^{z^3_1} + e^{z^3_2} + e^{z^3_3})\right)}{\partial z^3_j}$$

- Case 1 : $i == j$

Applying the $\partial(u/v)$ formula

$$\Rightarrow \frac{e^{z^3_i}(e^{z^3_1} + e^{z^3_2} + e^{z^3_3}) - e^{z^3_i} \cdot e^{z^3_i}}{(e^{z^3_1} + e^{z^3_2} + e^{z^3_3})^2}$$

$$\Rightarrow \frac{e^{z^3_i} \; (\sum_j e^{z^3_i} - e^{z^3_i})}{\sum e^{z^3_i} \cdot \sum_i e^{z^3_i}} \quad \to \left(1 - \frac{e^{z^3_i}}{\sum_i e^{z^3_i}}\right)$$

$$\frac{e^{z^3_i}}{\sum e^{z^3_i}}$$

$$\Rightarrow \psi(z^3_i) \cdot (1 - \psi(z^3_i))$$

$$\Rightarrow a^3_i \cdot (1 - a^3_i) \text{ when } i == j$$

- Case 2 : $i \neq j$

Applying the $\partial(u/v)$ formula,

$$\Rightarrow \frac{0 - e^{z^3_i} \cdot e^{z^3_j}}{(e^{z^3_1} + e^{z^3_2} + e^{z^3_3})^2}$$

$$\Rightarrow \frac{-e^{z^3_i}}{(e^{z^3_1}+e^{z^3_2}+e^{z^3_3})} \cdot \frac{e^{z^3_j}}{(e^{z^3_1}+e^{z^3_2}+e^{z^3_3})}$$

$$\Rightarrow -\psi(z^3_i) \cdot \psi(z^3_j)$$

$$\Rightarrow -a^3_i \cdot a^3_j$$

So, $\frac{\partial a_i}{\partial a_j} = a_i(\delta_{ij} - a_j)$

$\therefore$ Applying the chain rule

$$\frac{\partial J}{\partial z_i} = \frac{\partial J}{\partial a_i} \cdot \frac{\partial a_i}{\partial z_i}$$

$$\frac{\partial J}{\partial a_3} = \frac{\partial}{\partial a_3}(y_i \log(a_3))$$

$$= -1/N \cdot (y_i / a_3)$$

Plugging in the values,

$$\frac{\partial J}{\partial z_3} = -1/N \cdot (y_i/a_3) \cdot a_3(\delta_{ij} - a_3)$$

$$= 1/N\,(y_i(a_3 - \delta_{i,j}))$$

As y_i is an one hot matrix, (classification).

$$= 1/N\,(\psi(z_3) - \Delta)$$

(b) $\frac{\partial J}{\partial w^2} = \frac{\partial J}{\partial z^3} \cdot \frac{\partial z^3}{\partial w^2}$

$$\Rightarrow \frac{1}{N}(\psi(z^3) - \Delta) \cdot \frac{\partial(w^2 \cdot a^2 + b^2)}{\partial w^2}$$

$$= 1/N\,(\Psi(Z^3)-\Delta)\cdot a^2$$

For regularized loss:-

$$\frac{\partial \tilde{J}}{\partial Z^3} = \frac{1}{N}(\Psi(Z^3)-\Delta)\cdot\frac{\partial Z^3}{\partial w^2} + \frac{\partial}{\partial w^2}\left(\lambda(\|W^1\|_2^2+\|W^2\|_2^2\right)$$

$$\Rightarrow 1/N(\Psi(Z_3)-\Delta)\cdot\frac{\partial}{\partial W^2}(W^2a^2+b^2) + \frac{\partial}{\partial w^2}\lambda\cdot\left(\|W^1\|^2+\|W^2\|^2\right)$$

$$\Rightarrow 1/N(\Psi(Z_3)-\Delta)\cdot a^2 + 2\cdot\lambda\cdot W^2$$

(c)

(a) $$\frac{\partial J}{\partial w^1} = \frac{\partial J}{\partial Z^3}\cdot\frac{\partial Z^3}{\partial a^2}\cdot\frac{\partial a^2}{\partial Z^2}\cdot\frac{\partial Z^2}{\partial w^1}$$

$$\Rightarrow 1/N\,(\Psi(Z^3-\Delta))\cdot\frac{\partial(W^2a^2+b^2)}{\partial a^2}\cdot\frac{\partial\varphi(Z^2)}{\partial Z^2}\cdot\frac{\partial(w^1\cdot X+b^1)}{\partial w^1}$$

$$\Rightarrow 1/N\,(\Psi(Z^3-\Delta))\cdot W^2\cdot X\cdot\frac{\partial\varphi(Z^2)}{\partial Z^2}$$

$$\Rightarrow \begin{cases} 1/N\,\Psi(Z_3-\Delta)\cdot X\cdot W^2 & \text{if } (Z_2>0) \\ 0 & \text{if } (Z_2<0) \end{cases}$$

∴ For the regularized Loss:

$$\frac{\partial\tilde{J}}{\partial W_1} = \frac{\partial(\text{Cost Fn})}{\partial w^1} + \frac{\partial(\text{Reg Fn})}{\partial w^1}$$

$$\Rightarrow \frac{\partial J}{\partial w_1} + \frac{\partial}{\partial w_1} \cdot \lambda (\|w_1\|^2 + \|w_2\|^2)$$

$$\Rightarrow \frac{1}{N} \Psi(z_3 - \Delta) . X . w^2 + 2 \cdot \lambda \cdot w_1$$

ⓑ $$\frac{\partial J}{\partial b^2} = \frac{\partial J}{\partial z^3} \cdot \frac{\partial z^3}{\partial b^2}$$

$$\Rightarrow 1/N \; (\Psi(z^3 - \Delta)) \cdot \frac{\partial}{\partial b^2}(w^2 a^2 + b^2)$$

$$\Rightarrow 1/N \, (\Psi(z_3) - \Delta) \cdot 1.$$

ⓒ $$\frac{\partial J}{\partial b^1} = \frac{\partial J}{\partial z^3} \cdot \frac{\partial z^3}{\partial a^2} \cdot \frac{\partial a^2}{\partial z^2} \cdot \frac{\partial z^2}{\partial b^1}$$

∴ using the values derived from last eqns

$$\Rightarrow \begin{cases} 1/N \, (\Psi(z_3) - \Delta) \cdot w^2 \cdot 1 & \text{if } z_2 > 0 \\ 0 & \text{if } z_2 < 0 \end{cases}$$