Question 2: Backpropagation

(a)

Given loss function (in Q1) is,

$$J(\theta, \{x_i, y_i\}_{i=1}^{N}) = \frac{1}{N}\sum_{i=1}^{N} -\log\left[\frac{\exp z_{y_i}^{3}}{\sum_{j=1} \exp z_j^{3}}\right]$$

Let, $\exp(z_{y_i}^{3}) = g_{y_i}$ &

$\sum_j \exp(z_j^{3}) = h_j$

$$a^3 = \frac{g_{y_i}}{h_j} = \psi(z^3) \qquad \text{(softmax output)}$$

So, $$\frac{\partial J}{\partial z^3} = \frac{\partial J}{\partial a^3} \cdot \frac{\partial a^3}{\partial z^3}$$

$$\frac{\partial J}{\partial a^3} = -\frac{1}{N}\frac{\partial}{\partial a^3}\sum_{i=1}^{N} \log a^3 = -\frac{1}{N}\left(\frac{1}{a^3}\right)$$

So, $$\frac{\partial J}{\partial z^3} = -\frac{1}{N}\left(\frac{1}{a^3}\right)\frac{\partial a^3}{\partial z^3}$$

- The derivative $\frac{\partial a^3}{\partial z^3}$ will have two cases,

  when $z_{y_i}^{3} = z^3 = z_j^{3}$ or $y_i = j$ because only then the derivative of $z_{y_i}^{3}$ by z^3 is possible.

- The derivative of z_j^{3} is possible by z^3 in all the cases as z_j^{3} has all the z^3 summed up. So only one we'll remain after the derivative.

- The other case is $z_{y_i}^{3} \neq z^3$ or $y_i \neq j$. Here the derivative of $z_{y_i}^{3}$ by z^3 will be zero.

if $y_i \neq j$, then $\dfrac{\partial g_{y_i}}{\partial z^3} = 0$

$$\frac{\partial J}{\partial z^3} = -\frac{1}{N}\left(\frac{1}{a^3}\right)\left(0 - \frac{g_{y_i}}{h_j^2} \exp\left(z_j^{(3)}\right)\right)$$

$$= \frac{1}{N} \cdot \psi\left(z_j^3\right) = \frac{1}{N}\psi\left(z_j^3\right)$$

if $y_i = j$, then $\dfrac{\partial g_{y_i}}{\partial z^3} = g_{y_i}$ as only the part of y_i differentiated by z^3 are contained.

$$\frac{\partial J}{\partial z^3} = -\frac{1}{N}\left(\frac{1}{a^3}\right)\left(\frac{g_{y_i}}{h_j} - \frac{g_{y_i}}{h_j^2} \exp\left(z_j^3\right)\right)$$

$$= -\frac{1}{N}\left(1 - \frac{\exp\left(z_j^3\right)}{h_j}\right)$$

$$= \frac{1}{N}\left(\frac{\exp\left(z_j^3\right)}{\sum \exp\left(z_j^3\right)} - 1\right) = \frac{1}{N}\left(\psi\left(z_j^3\right) - 1\right)$$

$$= \frac{1}{N}\left(\psi\left(z^3\right) - 1\right)$$

So, $$\frac{\partial J}{\partial z^3} = \frac{1}{N}\left(\psi\left(z^3\right) - \Delta\right)$$

where $\Delta = \begin{cases} 1 & \text{when } y_i = j \\ 0 & \text{otherwise} \end{cases}$

(b)

Deriving loss function with w^2

$$\frac{\partial J}{\partial w^2}\left(\{x_i, y_i\}_{i=1}^{N}\right)$$

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

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

$$= \frac{1}{N}\left(\psi(z^3) - \Delta\right) \cdot a^2$$

Considering the partial derivatives in case of regularization

$$\frac{\partial \tilde{J}}{\partial w^2} = \underbrace{\frac{1}{N}\left(\psi(z^3) - \Delta\right) \cdot a^2}_{\text{from previous case}} + \frac{\partial}{\partial w^2} \lambda \left(\|w^{(1)}\|_2^2 + \|w^{(2)}\|_2^2\right)$$

$$= \frac{1}{N}\left(\psi(z^3) - \Delta\right) \cdot a^{2'} + 2\lambda w^{(2)}$$

(c) The partial derivatives of loss with respect to model parameters $(w^{(1)}, b^{(1)}, b^{(2)})$

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

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

$$= \frac{1}{N}\left(\psi(z^3) - \Delta\right) \cdot (w^2) \cdot \phi(z^2)$$

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

$$= \frac{1}{N}\left[\psi(z^3) - \Delta\right) \frac{\partial}{\partial b^2}\left(a^2 w^2 + b^2\right)$$

$$= \frac{1}{N}\left(\psi(z^3) - \Delta\right)$$

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

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

$$\frac{\partial}{\partial w^1}\left(a^1 w^1 + b^1\right) + \frac{\partial}{\partial w^1}\left(\lambda\left(\|w^1\|^2 + \|w^2\|^2\right)\right)$$

$$= \frac{1}{N}\left(\psi(z^3) - \Delta\right) w^2 \phi'(z^2) \cdot a^1 + 2\lambda w^1$$