CODING
<LANE>

NOTES

Our purpose, is to find $dZ_3 = \partial cost / dw_3$, for multi-class classification

To find $\frac{\partial L}{\partial W_3}$, we need to find $\frac{\partial a_i}{\partial z_j}$.

$$Loss = -\sum_{i=k}^{n}[y_k * log(a_k)]$$

But $\frac{\partial a_i}{\partial z_j}$ will be different for 2 cases.

$$a_i = \frac{e^{z_i}}{\sum_{i=k}^{n} e^{z_k}}$$

Case I : $i \neq j$

$$\frac{\partial a_i}{\partial z_j} = \frac{\partial}{\partial z_j}\left[\frac{e^{z_i}}{e^{z_j} + \sum'_{k\neq j} e^{z_k}}\right]$$

$$= \frac{\left(e^{z_j} + \sum'_{k\neq j} e^{z_k}\right)\frac{\partial}{\partial z_j}(e^{z_i}) - e^{z_i}\frac{\partial}{\partial z_j}\left[e^{z_j} + \sum'_{k\neq j} e^{z_k}\right]}{\left(e^{z_j} + \sum'_{k\neq j} e^{z_k}\right)^2}$$

$$\left[\because \frac{\partial}{\partial x}\left(\frac{u}{v}\right) = \frac{v\frac{\partial u}{\partial x} - u\frac{\partial v}{\partial x}}{v^2}\right]$$

$$\frac{\partial a_i}{\partial z_j} = \frac{0 - e^{z_i}.e^{z_j}}{\left(\sum_{k=1}^{n} e^{z_k}\right)} = \frac{-e^{z_i}}{\left(\sum_{k=1}^{n} e^{z_k}\right)} \cdot \frac{e^{z_j}}{\left(\sum' e^{z_k}\right)} = -a_i . a_j$$

Case II : $i=j$

$$\frac{\partial a_i}{\partial z_i} = \frac{\partial}{\partial z_i}\left(\frac{e^{z_i}}{e^{z_i}+\sum_{k\neq i} e^{z_k}}\right) = \frac{\left(e^{z_i}+\sum_{k\neq i} e^{z_k}\right)\frac{\partial}{\partial z_i}\left(e^{z_i}\right) - \left(e^{z_i}\right)\frac{\partial}{\partial z_i}\left(e^{z_i}+\sum_{k\neq i} e^{z_k}\right)}{\left(e^{z_i}+\sum_{k\neq i} e^{z_k}\right)^2}$$

$$\frac{\partial a_i}{\partial z_i} = \frac{\left(e^{z_i}+\sum_{k\neq i} e^{z_k}\right)\cdot\left(e^{z_i}\right) - \left(e^{z_i}\right)\cdot\left(e^{z_i}+0\right)}{\left(e^{z_i}+\sum_{k\neq i} e^{z_k}\right)^2}$$

$$\frac{\partial a_i}{\partial z_i} = \frac{e^{z_i}}{\left(\sum_{k=1}^{n} e^{z_k}\right)}\cdot\frac{\sum_{k\neq i} e^{z_k}}{\left(\sum_{k=1}^{n} e^{z_k}\right)} = a_i(1-a_i)$$

$$\left[\because\ a_i = \frac{e^{z_i}}{\sum_{k=1}^{n} e^{z_k}}\ \&\ (1-a_i) = \frac{\sum_{k\neq i} e^{z_k}}{\sum_{k=1}^{n} e^{z_k}}\right]$$

Thus, $\dfrac{\partial a_i}{\partial z_j} = \begin{cases} -a_i a_j, & \text{for } i \neq j \\ a_i(1-a_i), & \text{for } i = j \end{cases}$

Now, $\dfrac{\partial L}{\partial z_i} = \dfrac{\partial}{\partial z_i}\left[-y_i \cdot \log(a_i) - \sum_{k \neq i} y_k \cdot \log(a_k)\right]$

$$= \frac{-y_i}{a_i} \cdot \frac{\partial a_i}{\partial z_i} - \sum_{k \neq i} \frac{y_k}{a_k} \cdot \frac{\partial a_k}{\partial z_i}$$

$$= \frac{-y_i}{a_i} \cdot a_i \cdot (1-a_i) + \sum_{k \neq i} \frac{y_k}{a_k} \cdot a_k \cdot a_i$$

$$= -y_i + y_i a_i + a_i \sum_{k \neq i} y_k$$

$$= -y_i + a_i\left[y_i + \sum_{k \neq i} y_k\right]$$

$$= -y_i + a_i\left[\sum_{k=1}^{n} y_k\right] = -y_i + a_i(1)$$

$\left[\text{As, } \sum_{k=1}^{n} y_k = 1\right]$

$$= \underline{\underline{a_i - y_i}}$$

Thus, $\frac{dL}{dZ_3} = a_3 - y$

And thus, $dZ_3 = \frac{\partial \text{cost}}{\partial W_3} = \underline{\underline{A_3 - Y}}$