Back propagation notes

$A = g(Z)$; $A, Z \in \mathbb{R}^{(n_A, m)}$

$Z = W A\_prev$; $A_{prev} \in \mathbb{R}^{(n_x, m)}$; $W \in \mathbb{R}^{(n_A, n_x)}$

$$\frac{\partial Z_{kl}}{\partial W_{ij}} = \left| \begin{array}{ll} 0 & \text{if } k \neq i \\ A_{prev\,lj} & \text{if } k = i \end{array} \right| = A_{prev\,lj} \quad ; \quad \frac{\partial A_{kl}}{\partial Z_{ij}} = g'(Z_{kl})\, \delta_{k,i}\, \delta_{l,j}$$

$$\frac{\partial A_{kl}}{\partial W_{ij}} = \sum_{e=1}^{n_A} \sum_{f=1}^{m} \frac{\partial A_{kl}}{\partial Z_{ef}} \cdot \frac{\partial Z_{ef}}{\partial W_{ij}} = \sum_{e=1}^{n_A} \sum_{f=1}^{m} g'(Z_{kl}) \cdot \delta_{k,e} \cdot \delta_{l,f} \cdot A_{prev\,fj}\, \delta_{e,i}$$

$$= g'(Z_{kl})\, A_{prev\,l,j}\, \delta_{k,i}$$

$$\frac{\partial L}{\partial W_{ij}} = \sum_{k=1}^{n_A} \sum_{l=1}^{M} \frac{\partial L}{\partial A_{kl}} \frac{\partial A_{kl}}{\partial W_{ij}} = \sum_{k=1}^{n_A} \sum_{e=1}^{M} \frac{\partial L}{\partial A_{kl}}\, g'(Z_{kl}) \cdot A_{prev\,l,j}\, \delta_{k,i} =$$

$$= \sum_{l=1}^{m} \frac{\partial L}{\partial A_{il}} \cdot g'(Z_{il}) \cdot A_{l,j} = \left(\frac{\partial L}{\partial A_{i1}} \cdot g'(Z_{i1}) \cdot A_{1j} + \ldots + \frac{\partial L}{\partial A_{i n_A}} g'(Z_{i n_A}) A_{n_A j} \right)$$

* Vectorized implementation $\longrightarrow$

$$dW = (dAL * g'(Z)) \cdot A_{prev}^T$$

element wise multiplication.

dot product

$$\frac{\partial Z_{ke}}{\partial A_{prev_{ij}}} = W_{ki}\,\delta_{e,j} \quad ; \quad \frac{\partial A_{ke}}{\partial A_{prev_{ij}}} = \sum_{e=1}^{n_A}\sum_{f=1}^{m} \frac{\partial A_{ke}}{\partial Z_{ef}} \cdot \frac{\partial Z_{ef}}{\partial A_{prev_{ij}}} = g'(Z_{ke}) \cdot W_{k,i} \cdot \delta_{e,j}$$

$$\frac{\partial \mathcal{L}}{\partial A_{prev_{ij}}} = \sum_{e=1}^{n_A}\sum_{f=1}^{m} \frac{\partial \mathcal{L}}{\partial A_{ef}} \cdot \frac{\partial A_{ef}}{\partial A_{prev_{ij}}} = \sum_{e=1}^{n_A}\sum_{f=1}^{m} \frac{\partial \mathcal{L}}{\partial A_{ef}} \cdot g(Z_{ef})\, W_{e,i}\, \delta_{f,j}$$

$$= \sum_{e=1}^{n_A} \frac{\partial \mathcal{L}}{\partial A_{ej}}\, g'(Z_{ej})\, W_{e,i}$$

* Vectorized implementation. $\longrightarrow$

$$dA\_prev = W^T \cdot (dAL * g'(Z))$$