ω^{l}_{jk} = weight from k^{th} neuron in layer $l-1$ to j^{th} neuron in layer l

b^{l}_{j} = bias in neuron j in layer l

z^{l}_{j} = pre activation value of neuron j in layer l

a^{l}_{j} = post activation value of neuron j in layer l

$a^{l}_{j} = \sigma(z^{l}_{j}) = z^{l}_{j}$ – for linear nodes

$a^{l}_{j} = \sigma(z^{l}_{j}, z^{l}_{j*}) =$ for circular node

where z_{j*} is the circular partner for z_{j}

$$\boxed{z^{l}_{j} = \sum_{k} \omega^{l}_{jk} a^{l-1}_{k} + b^{l}_{j}} \qquad (1)$$

Circular Preliminaries

$$r_j^n = \sqrt{(z_j^n)^2 + (z_{j*}^n)^2} = \left[(z_j^n)^2 + (z_{j*}^n)^2\right]^{1/2}$$

$$a_j^n = \frac{(z_j^n)}{\sqrt{(z_j^n)^2 + (z_{j*}^n)^2}} = \frac{z_j^n}{r_j^n} = z_j^n (r_j^n)^{-1}$$

$$a_{j*} = \frac{z_{j*}^2}{\sqrt{(z_j^2)^2 + (z_j^2)^2}} = \frac{z_{j*}^n}{r_j^n} = z_{j*}^n (r_j^n)^{-1}$$

$$\frac{\partial r_j^n}{\partial z_j^n} = \tfrac{1}{2}\left[(z_j^n)^2 + (z_{j*}^n)^2\right]^{-1/2} 2 z_j^n$$

$$= \frac{z_j^n}{r_j^n} = z_j^n (r_j^n)^{-1}$$

$$\frac{\partial a_j^n}{\partial z_j^n} = (r_j^n)^{-1} - z_j^n (r_j^n)^{-2} z_j^n (r_j^n)^{-1}$$

$$\boxed{= (r_j^n)^{-1} - (z_j^n)^2 (r_j^n)^{-3}} = \boxed{(z_{j*}^n)^2 (r_j^n)^{-3}}$$

$$\frac{\partial a_{j*}^n}{\partial z_j^n} = -(z_{j*}^n)(r_j^n)^{-2}(z_j^n)(r_j^n)^{-1}$$

$$= \boxed{-z_j^n z_j^n (r_j^n)^{-3}}$$

For Linear Neurons

$$a_j^n = z_j^n$$

$$\boxed{\frac{\partial a_j^n}{\partial z_j^n} = 1} \qquad (2L)$$

For Circular Neurons

$$a_j^n = z_j^n (r_j^n)^{-1} \quad \text{where} \quad r_j^n = \left[(z_j^n)^2 + (z_{j^*}^n)^2\right]^{1/2} \qquad (r_j^n = r_{j^*}^n)$$

now
$$\frac{\partial r_j^n}{\partial z_j^n} = \tfrac{1}{2}\left[(z_j^n)^2 + (z_{j^*}^n)\right]^{-1/2} \cdot 2 z_j^n$$

$$\frac{\partial r_j^n}{\partial z_j^n} = z_j^n (r_j^n)^{-1}$$

so
$$\frac{\partial a_j^n}{\partial z_j^n} = (r_j^n)^{-1} + (z_j^n)(-1)(r_j^n)^{-2}(z_j^n)(r_j^n)^{-1}$$

$$\boxed{\frac{\partial a_j^n}{\partial z_j^n} = (r_j^n)^{-1} - (z_j^n)^2 (r_j^n)^{-3}} \qquad \left[(r_j^n)^2 - (z_j^n)^2\right](r_j^n)^{-3} \qquad (2c)$$

$$\boxed{= (z_{j^*}^n)^2 (r_j^n)^{-3}} \qquad (2c)$$

$$a_{j^*}^n = z_{j^*}^n (r_j^n)^{-1}$$

$$\frac{\partial a_{j^*}^n}{\partial z_j^n} = (z_{j^*}^n)(-1)(r_j^n)^{-2}(z_j^n)(r_j^n)^{-1}$$

$$\boxed{\frac{\partial a_{j^*}^n}{\partial z_j^n} = -(z_j^n)(z_{j^*}^n)(r_j^n)^{-3}} \qquad (2c^*)$$

$D_j^\ell \stackrel{def}{=} \frac{\partial C}{\partial z_j^\ell}$ for som cost function C

consider outer layer ($\ell=0$), composed of linear neurons $a_j^0 = \sigma(z_j^0) = z_j^0$

$$D_j^0 = \frac{\partial C}{\partial a_k^0}\frac{\partial a_k^0}{\partial z_j^0} = \frac{\partial C}{\partial a_k^0}\delta_{kj} = \frac{\partial C}{\partial a_j^0}$$

Taking $C \stackrel{def}{=} \frac{1}{2}\sum_j (d_j - a_j^0)^2$

$$D_j^0 = \left(\tfrac{1}{2}\right)(2)(d_j - a_j^0)(-1)$$

$$\boxed{= (a_j^0 - d_j)}$$

$$D_j^{\ell-1} \stackrel{def}{=} \frac{\partial C}{\partial z_j^{\ell-1}}$$

$$D_j^{\ell-1} = \sum_k\sum_i \frac{\partial C}{\partial z_k^\ell}\frac{\partial z_k^\ell}{\partial a_i^{\ell-1}}\frac{\partial a_i^{\ell-1}}{\partial z_j^{\ell-1}}$$

$$D_j^{\ell-1} = \sum_k\sum_i D_k^\ell w_{ki}^\ell \frac{\partial a_i^{\ell-1}}{\partial z_j^{\ell-1}}$$ subbing (1)

<u>for linear</u>

$$D_j^{\ell-1} = \sum_k\sum_i D_k^\ell w_{ki}^\ell \delta_{ij}$$ subbing (2L)

$$\boxed{D_j^{\ell-1} = \sum_k D_k^\ell w_{kj}^\ell}$$ (3L)

Circular

$$D_j^{l-1} = \sum_{ki} D_k^l w_{ki}^l \frac{\partial a_i^{l-1}}{\partial z_j^{l-1}}$$

$$= \sum_k D_k^l w_{kj}^l (z_{j*}^{l-1})^2 (r_j^{l-1})^{-3} - \sum_k D_k^l w_{kj*}^l (z_j^{l-1})(z_{j*}^{l-1})(r^{l-1})^{-3}$$

$$= \sum_k D_k^l (r_j^{l-1})^{-3} \left[w_{kj*}^l (z_j^{l-1})^2 - (w_{kj*}^l)(z_j^{l-1})(z_{j*}^{l-1}) \right] \quad (3c^*)$$