$$\frac{1}{2}\sum_{k=1}^{\#outputs}\left(y_k - a_k^{(l)}\right)^2 = RSS$$

$$\boxed{W_{kj}^{(l-1)} := W_{kj}^{(l-1)} - n \cdot \frac{\partial RSS}{\partial W_{kj}^{(l-1)}}} \quad (1)$$

$$\frac{\partial RSS}{\partial W_{kj}} = \frac{\partial RSS}{\partial a_k^{(l)}} \cdot \frac{\partial a_k^{(l)}(z_k^{(l)})}{\partial (z_k^{(l)})} \cdot \frac{\partial z_k^{(l)}}{\partial W_{kj}} \Rightarrow$$

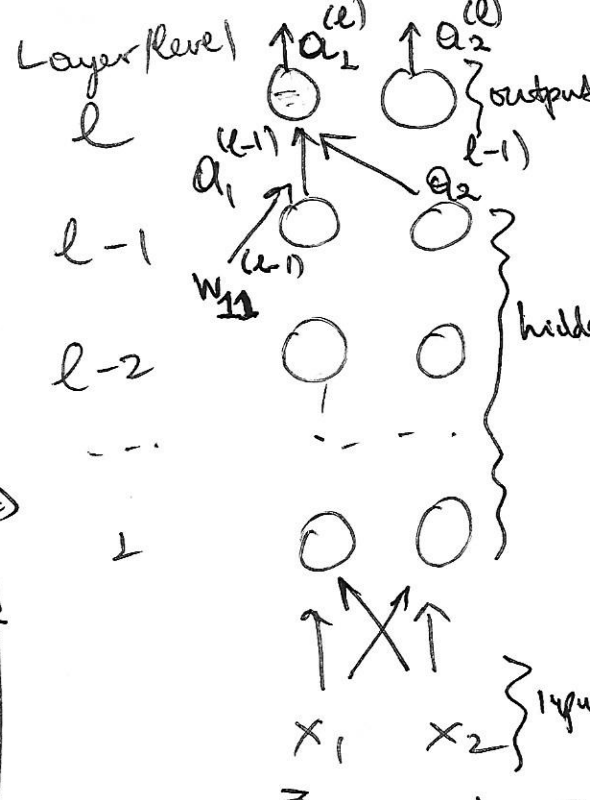
$$\Rightarrow \frac{\partial RSS}{\partial W_{kj}} = \underbrace{-\left(y_k - a_k^{(l)}\right)\left(1 - a_k^{(l)}\right) a_k^{(l)}}_{\delta_k^{(l)}} \, a_j^{(l-1)}$$

$$\Rightarrow \boxed{\frac{\partial RSS}{\partial W_{kj}} = -\delta_k^{(l)} \cdot a_j^{(l-1)}} \quad (2)$$

$$(1) \wedge (2) \Rightarrow W_{kj}^{(l-1)} := W_{kj}^{(l-1)} + n\,\delta_k^{(l)} a_j^{(l-1)}$$

learning rate

All these calculations for the final layer (w's)



$$a_k^{(l)} = g(\cancel{z}) = \frac{1}{1+e^{-z}}$$

$$D = [X \mid Y]$$

$$a_k^{(l)}(z_k^{(l)}) \qquad a_1^{(l-1)}$$

$$z_k^{(l)} = W_{k0} \cdot x_0 + W_{k1} \cdot a_1^{(l-1)} + \dots + W_{kj} a_j^{(l-1)} + \dots$$

$$\frac{\partial a_k^{(l)}(z)}{\partial z} = \left(\frac{1}{1+e^{-z}}\right)' =$$

$$= \frac{+e^{-z}}{(1+e^{-z})^2} =$$

$$= \left(1 - \frac{1}{1+e^{-z}}\right) \cdot \left(\frac{1}{1+e^{-z}}\right) =$$

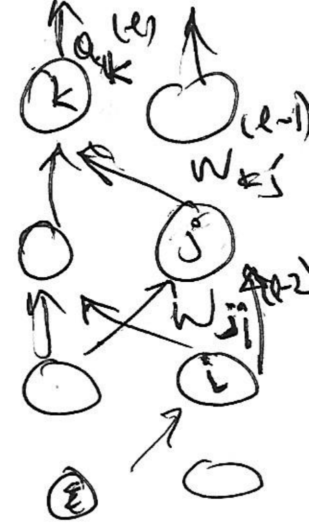
$$= \left(1 - a_k^{(l)}\right)\left(a_k^{(l)}\right)$$

$$\frac{\partial RSS}{\partial w_{ji}^{(l-2)}} = \frac{\partial RSS}{\partial a_j^{(l-1)}} \cdot \frac{\partial a_j^{(l-1)}}{\partial z_j^{(l-1)}} \cdot \underbrace{\frac{\partial z_j^{(l-1)}}{\partial w_{ji}^{(l-2)}}}_{\text{hidden}}$$

output(l)

$l-1$

$l-2$



$$= \frac{\partial RSS}{\partial a_j^{(l-1)}} \left(1-a_j^{(l-1)}\right) \cdot a_j^{(l-1)} \quad a_i^{(l-2)} =$$

$$= \frac{\partial RSS}{\partial a_k^{(l)}} \cdot \frac{\partial a_k(z_k^{(l)})}{\partial z_k^{(l)}} \cdot \frac{\partial z_k^{(l)}}{\partial a_j^{(l-1)}} \cdots$$

$$= -\sum_k \underbrace{\left(y_k - a_k^{(l)}\right) \cdot a_k^{(l)}\left(1-a_k^{(l)}\right)}_{\delta_k^{(l)}} \cdot w_{kj}^{(l-1)} \cdot \cdots$$

$$= -\underbrace{\sum_K \delta_k^{(l)} \cdot w_{kj}^{(l-1)} \cdot \left(1-a_j^{(l-1)}\right) \cdot a_j^{(l-1)}}_{\delta_j^{(l-1)}} \cdot a_i^{(l-2)} \quad \rightarrow$$

$\delta_j^{(l-1)} \cdot a_i$

$$\Rightarrow \boxed{\frac{\partial RSS}{\partial w_{ji}^{(l-2)}} = -\delta_j^{(l-1)} \cdot a_i^{(l-2)}}$$

$$RSS = \frac{1}{2} \sum_k \left(y_k - a_k^{(l)}\right)^2$$

$$z_j^{(l-1)} = w_{j0} \cdot x_0 + w_{j1}^{(l-2)} \cdot a_1^{(l-2)} + \cdots + w_{ji}^{(l-2)} \cdot a_i^{(l-2)}$$

$$z_k^{(l)} = w_{k0} \cdot x_0 + w_{k1}^{(l-1)} \cdot a_1^{(l-1)} + \cdots + w_{kj}^{(l-1)} \cdot a_j^{(l-1)}$$