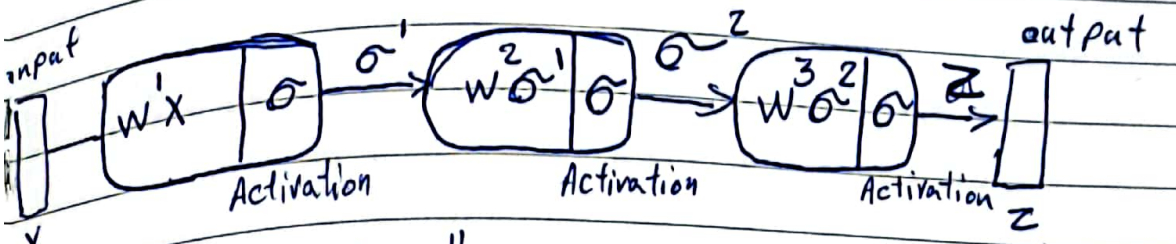

if we consider the same min squar error loss function since we have

$$\left.\begin{aligned} z &= \sigma^*(W^3\sigma^2) \\ \sigma^2 &= \sigma^*(W^2\sigma^1) \\ \sigma^1 &= \sigma^*(W^1X) \end{aligned}\right\} \Rightarrow z = \sigma^*(W^3\sigma^*(W^2\sigma^*(W^1X)))$$

we considered min squar error as loss function so:

$$J(W) = \frac{1}{2m}\sum_{s\in Q}\sum_{k=1}^{c} t_k^{(s)} - z_k^{(s)}$$

$$\frac{\partial J}{\partial W_3} = \frac{\partial J}{\partial z} \times \frac{\partial z}{\partial W_3}$$

$$\frac{\partial J}{\partial W_2} = \frac{\partial J}{\partial z} \times \frac{\partial z}{\partial W_3} \times \frac{\partial W_3}{\partial \sigma^2} \times \frac{\partial \sigma^2}{\partial W^2}$$

$$\frac{\partial J}{\partial W_1} = \frac{\partial J}{\partial z} \times \frac{\partial z}{\partial W^3} \times \frac{\partial W^3}{\partial \sigma^2} \times \frac{\partial \sigma^2}{\partial W^2} \times \frac{\partial W^2}{\partial \sigma^1} \times \frac{\partial \sigma^1}{\partial W_1}$$

---

* σ is element wise sigmoid of corresponding entry