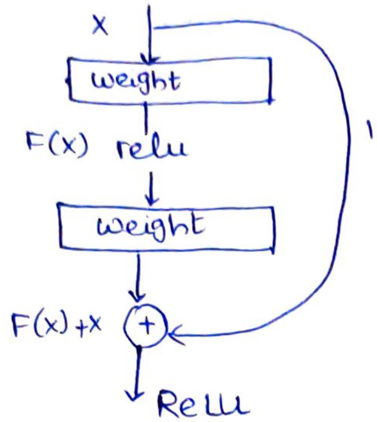
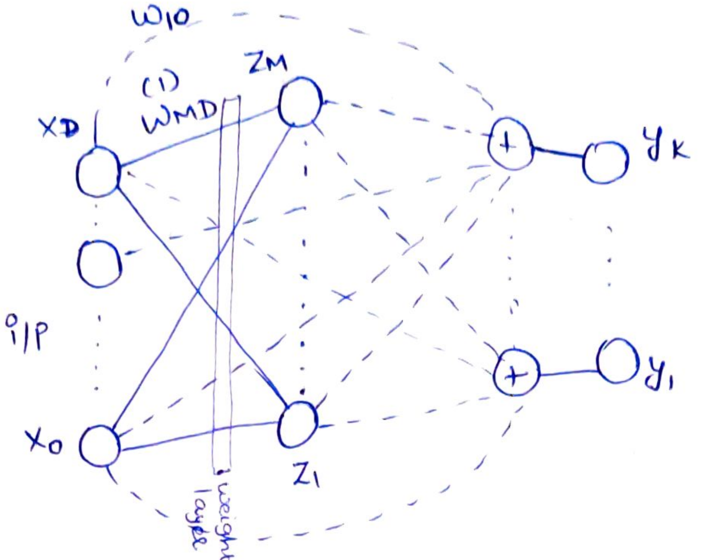
1. Skip connection looks like





Now, for the sake of simplicity, Assuming only one O/p exists (y_1)

we can get the output of i/p layer i.e. i/p to hidden layer as

$$a_i^{(1)} = \sum_{j=0}^{D} w_{ij}^{1} x_j$$

This would be converted by some activation fn (h) to $z_i^{(1)}$

$$z_i^{(1)} = h(a_i^{(1)})$$

The ~~o/p of hidden layer~~ ~~ie~~. input the skip connection becomes

$$a_i^{(2)} = \sum_{j=0}^{M} w_{ij} z_j + \sum_{k=0}^{D} w_{k0} x_k$$ [where w_k is the skip connection weight from i/p]

This would be given to the output node y_i

$y_i = h'(a_i^{(2)})$ But for simplicity purpose, Let's assume $h'()$ as an identity fn

$\Rightarrow y_i = a_i^{(2)}$

The loss (MSE loss) is calculated as $E = \frac{1}{2} \sum_{i=1}^{k} (y_i - t_i)^2$

The $\frac{\partial E}{\partial w_{10}}$ (~~the~~ w_{10} is the weight from X_D to 1st $\oplus$)

$$\frac{\partial E}{\partial w_{10}} = \frac{\partial E}{\partial y_1} \cdot \frac{\partial y_1}{\partial a_i} \cdot \frac{\partial a_i}{\partial w_{10}}$$

$$= \frac{\partial}{\partial y_1}\left(\frac{1}{2}\sum_{i=1}^{k}(y_i - t_i)^2\right) \cdot \frac{\partial}{\partial a_1}(a_1) \cdot \frac{\partial}{\partial w_{10}}\left(\sum_{j=0}^{M} w_{ij} z_j + \sum_{k=0}^{D} w_{k0} x_k\right)$$

$$= 1 \cdot 1 \cdot x_1$$

$$\frac{\partial E}{\partial w_{10}} = x_1$$

similarly for $\frac{\partial E}{\partial w_{20}} = x_2$, $\frac{\partial E}{\partial w_{30}} = x_3$ ...