First of all, calculate the back propagation formula:

- ~~1st~~ 2nd Hidden layer:

$$\frac{\partial RSS}{\partial w_{i1}^{(3)}} = \frac{\partial RSS}{\partial a_1^{(4)}} \cdot \frac{\partial a_1^{(4)}}{\partial z_1^{(4)}} \cdot \frac{\partial z_1^{(4)}}{\partial w_{i1}^{(3)}} = -2(y_1 - a_1) z_1^{(4)}(1 - z_1^{(4)}) a_i^{(3)}$$

$$\Rightarrow W_{i1}^{(3)} = W_{i1}^{(3)} + 2(y_1 - a_1^{(4)}) z_1^{(4)}(1 - z_1^{(4)}) a_i^{(3)} \cdot \eta$$

- 1st layer:

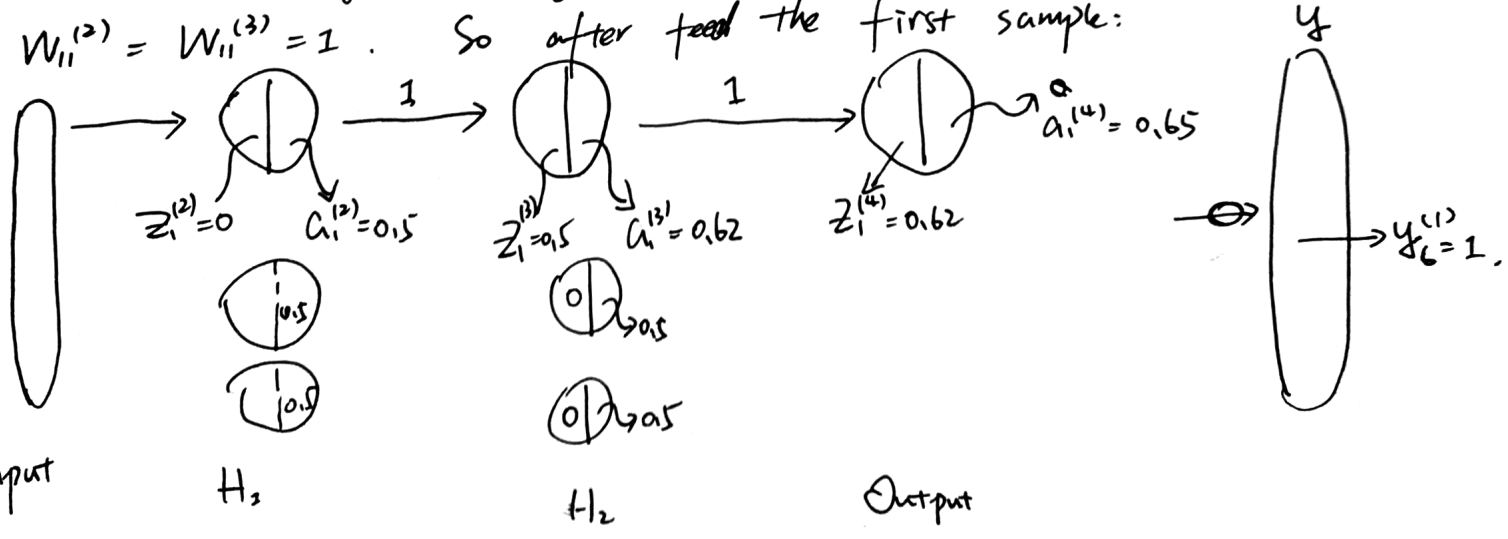
$$\frac{\partial RSS}{\partial w_{i1}^{(2)}} = \sum_{k=1}^{10} \frac{\partial RSS}{\partial a_k^{(4)}} \cdot \frac{\partial a_k^{(4)}}{\partial z_k^{(4)}} \cdot \frac{\partial z_k^{(4)}}{\partial a_1^{(3)}} \cdot \frac{\partial a_1^{(3)}}{\partial z_1^{(3)}} \cdot \frac{\partial z_1^{(3)}}{\partial w_{i1}^{(2)}}$$

$$= -2 z_1^{(3)}(1 - z_1^{(3)}) a_i^{(2)} \sum_{k=1}^{10} (y_k - a_k^{(4)}) z_k^{(4)}(1 - z_k^{(4)}) W_{1k}^{(3)}$$

$$\Rightarrow W_{i1}^{(2)} = W_{i1}^{(2)} + 2 z_1^{(3)}(1 - z_1^{(3)}) a_i^{(2)} \sum_{k=1}^{10} (y_k - a_k^{(4)}) z_k^{(4)}(1 - z_k^{(4)}) W_{1k}^{(3)} \cdot \eta$$

Then set learning rate: $\eta = 1$ and all the weights to be zero except:

$W_{11}^{(2)} = W_{11}^{(3)} = 1$. So after ~~feed~~ the first sample:



1st layer:

$$\Rightarrow W_{i1}^{(3)} = W_{i1}^{(3)} + 2(0 - 0.65) \cdot 0.62(1 - 0.62) \cdot 0.62 = W_{i1}^{(3)} - 0.19.$$

$$\Rightarrow W_{11}^{(3)} = \cancel{0.19}\ 0.81, \quad W_{i1}^{(3)} = -0.19 \quad (i = 0, 2, 3)$$

2nd layer:

$$W_{i1}^{(2)} = W_{i1}^{(2)} + 2 \cdot 0.5(1 - 0.5) \cdot 0.5 \sum_{k=1}^{10} (y_k - a_k^{(4)}) z_k^{(4)}(1 - z_k^{(4)}) W_{1k}^{(3)}$$

~~Ok~~ Only $W_{11}^{(3)} = 1$:

$$W_{i1}^{(2)} = W_{i1}^{(2)} + 0.25 (y_1 - a_1^{(4)}) z_1^{(4)}(1 - z_1^{(4)}) W_{11}^{(3)} = W_{i1}^{(2)} + 0.25(0 - 0.65) 0.62 \cdot 0.38$$

$$= W_{i1}^{(2)} - 0.021 \Rightarrow W_{11}^{(2)} = 0.979, \quad W_{i1}^{(2)} = -0.021 \ (i = 0, 2, 3).$$