Error Propagation:

1)(a) $$\frac{\partial E_n(w)}{\partial a_1^{(4)}} = \frac{\partial\left(\frac{1}{2}(h(a_1^{(4)})-t_n)^2\right)}{\partial a_1^{(4)}}$$

$$= \frac{\partial\left(\frac{1}{2}(a_1^{(4)}-t_n)^2\right)}{\partial a_1^{(4)}} \quad \left[\because h(a)=a \text{ for o/p layer}\right]$$

$$= \frac{1}{\cancel{2}} \times \cancel{2}\,(a_1^{(4)}-t_n) \times 1$$

$$\frac{\partial E_n(w)}{\partial a_1^{(4)}} = a_1^{(4)} - t_n \equiv \delta_1^{(4)}$$

(b) Calculate

$$\frac{\partial E_n(w)}{\partial w_{12}^{(3)}} = \frac{\partial E_n(w)}{\partial a_1^{(4)}} \frac{\partial a_1^{(4)}}{\partial w_{12}^{(3)}}$$

$$= \delta_1^{(4)} \frac{\partial a_1^{(4)}}{\partial w_{12}^{(3)}}$$

$$\frac{\partial a_1^{(4)}}{\partial w_{12}^{(3)}} = \frac{\partial}{\partial w_{12}^{(3)}}\left(w_{11}^{(3)} z_1^{(3)} + w_{12}^{(3)} z_2^{(3)} + w_{13}^{(3)} z_3^{(3)} + w_{14}^{(3)} z_4^{(3)}\right)$$

$$= z_2^{(3)}$$

$$\frac{\partial a_1^{(4)}}{\partial w_{12}^{(3)}} = z_2^{(3)} = h(a_2^{(3)}) = \frac{1}{1+e^{-a_2^{(3)}}}$$

$$\therefore \frac{\partial E_n(w)}{\partial w_{12}^{(3)}} = \delta_1^{(4)} \frac{1}{1+e^{-a_2^{(3)}}}$$

$$= (a_1^{(4)} - t_n) \frac{1}{1+e^{-a_2^{(3)}}}$$

2) a) penultimate layer:

$$\frac{\partial E_n(w)}{\partial a_1^{(3)}} = \frac{\partial E_n(w)}{\partial a_1^{(4)}} \frac{\partial a_1^{(4)}}{\partial z_1^{(3)}} \frac{\partial z_1^{(3)}}{\partial a_1^{(3)}} \rightarrow 1$$

$$\frac{\partial z_1^{(3)}}{\partial a_1^{(3)}} = \frac{\partial (h(a_1^{(3)}))}{\partial a_1^{(3)}} = \frac{\partial \left(\frac{1}{1+e^{-a_1^{(3)}}}\right)}{\partial a_1^{(3)}}$$

$$= \frac{e^{a_1^{(3)}}}{(1+e^{a_1^{(3)}})^2} \rightarrow 2$$

$$\frac{\partial a_1^{(4)}}{\partial z_1^{(3)}} = \frac{\partial}{\partial z_1^{(3)}}\left(w_{11}^{(3)} z_1^{(3)} + \underbrace{\cancel{w_{12}^{(3)} z_2^{(3)}}}_{0} + \underbrace{\cancel{w_{13}^{(3)}} z_3^{(3)}}_{0} + \underbrace{\cancel{w_{14}^{(3)} z_4^{(3)}}}_{0}\right)$$

$$= w_{11}^{(3)} \rightarrow 3$$

From 1, applying question 1) a), 2, 3 we get

$$\frac{\partial E_n(w)}{\partial a_1^{(3)}} = \delta_1^{(4)} \cdot w_{11}^{(3)} \cdot \frac{e^{a_1^{(3)}}}{(1+e^{a_1^{(3)}})^2} \equiv \delta_1^{(3)}$$

2) b) Calculate $\frac{\partial E_n(w)}{\partial w_{11}^{(2)}}$

$$\frac{\partial E_n(w)}{\partial w_{11}^{(2)}} = \underbrace{\frac{\partial E_n(w)}{\partial a_1^{(3)}}}_{\text{known from previous answer}} \frac{\partial a_1^{(3)}}{\partial w_{11}^{(2)}}$$

$$\frac{\partial a_1^{(3)}}{\partial w_{11}^{(2)}} = \frac{\partial}{\partial w_{11}^{(2)}} \left(w_{11}^{(2)} z_1^{(2)} + \cancel{w_{12}^{(2)} z_2^{(2)}} + \cancel{w_{13}^{(2)} z_3^{(2)}} + \cancel{w_{14}^{(2)} z_4^{(2)}} \right)$$

$$\frac{\partial a_1^{(3)}}{\partial w_{11}^{(2)}} = z_1^{(2)} = h(a_1^{(2)}) = \frac{1}{1+e^{-a_1^{(2)}}}$$

$$\therefore \frac{\partial E_n(w)}{\partial w_{11}^{(2)}} = \delta_1^{(3)} \cdot \frac{1}{1+e^{-a_1^{(2)}}}$$

$$= \delta_1^{(4)} \cdot w_{11}^{(3)} \cdot \frac{e^{a_1^{(3)}}}{(1+e^{a_1^{(3)}})^2} \cdot \frac{1}{1+e^{-a_1^{(2)}}}$$

3) a) From the input layers

$$\frac{\partial E_n(w)}{\partial a_1^{(2)}} = \sum_{k=1}^{3} \frac{\partial E_n(w)}{\partial a_k^{(3)}} \frac{\partial a_k^{(3)}}{\partial a_j^{(2)}}$$

$$\sum_{k=1}^{3} \frac{\partial E_n(w)}{\partial a_k^{(3)}} = \sum_{k=1}^{3} \delta_k^{(3)}$$

$$\frac{\partial a_k^{(3)}}{\partial a_1^{(2)}} = \frac{\partial}{\partial a_1^{(2)}} \left(\sum_{m=1}^{3} \sum_{k=1}^{3} w_{km}^{(2)} z_m^{(2)} \right)$$

$$= h'(a_1^{(2)}) \sum_{k=1}^{3} w_{k1}^{(2)}$$

$$= \frac{e^{a_1^{(2)}}}{\left(1+e^{a_1^{(2)}}\right)^2} \sum_{k=1}^{3} w_{k1}^{(2)}$$

$$\therefore \frac{\partial E_n(w)}{\partial a_1^{(2)}} = \frac{e^{a_1^{(2)}}}{\left(1+e^{a_1^{(2)}}\right)^2} \sum_{k=1}^{3} w_{k1}^{(2)} \delta_k^{(3)} \equiv \delta_1^{(2)}$$

(b)

$$\frac{\partial E_n(w)}{\partial w_{11}^{(1)}} = \underbrace{\frac{\partial E_n(w)}{\partial a_1^{(2)}}}_{\text{known from previous answer}} \frac{\partial a_1^{(2)}}{\partial w_{11}^{(1)}}$$

$$\frac{\partial a_1^{(2)}}{\partial w_{11}^{(1)}} = \frac{\partial}{\partial w_{11}^{(1)}} \left(w_{11}^{(1)} z_1^{(1)} + \cancelto{0}{w_{12}^{(1)} z_2^{(1)}} + \cancelto{0}{w_{13}^{(1)} z_3^{(1)}} + \cancelto{0}{w_{14}^{(1)} z_4^{(1)}} \right)$$

$$z_1^{(1)} \Rightarrow h(a_1^{(1)})$$

$$\frac{\partial a_1^{(2)}}{\partial w_{11}^{(1)}} = \frac{1}{1 + e^{-a_1^{(1)}}}$$

$$\therefore \frac{\partial E_n(w)}{\partial w_{11}^{(1)}} = \frac{e^{a_1^{(2)}}}{\left(1 + e^{a_1^{(2)}}\right)^2} \sum_{k=1}^{3} w_{k1}^{(2)} \delta_k^{(3)} \frac{1}{1 + e^{-a_1^{(1)}}}$$

Fine Tuning a pre-trained network:

The model is trained on CIFAR10 dataset. The test/validation set is also taken from the same dataset.

The dataset is trained with 10 epochs. For 3rd, 6th and 10th epoch, we validate the CIFAR10 test dataset and store the testing error and the accuracy scores. The best model is obtained when we add L2 regularization with an accuracy of around 65.35%.