Week 4 Theory HW

Naik Hiong Chiang

Task 1:

figure 1:

1.

$$\frac{\partial E}{\partial n_4} = \frac{\partial E}{\partial n_6}\frac{\partial n_6}{\partial n_4}$$
$$= \frac{\partial \sum_{i\in data}(n_6 - y_6^{(i)})^2 + (n_7 - y_7^{(i)})^2}{\partial n_6}\frac{\partial g(b + \sum_k w_{k,6}n_k)}{\partial n_4}$$
$$= \Big(\sum_{i\in data} 2(n_6 - y_6^{(i)})\Big)\frac{\partial g(b + w_{3,6}n_3 + w_{4,6}n_4)}{\partial n_4}$$
$$= \Big(2\sum_{i\in data}(n_6 - y_6^{(i)})\Big)g'(b + w_{3,6}n_3 + w_{4,6}n_4)\cdot w_{4,6}$$

2.

$$\frac{\partial E}{\partial w_{2,5}} = \frac{\partial E}{\partial n_5}\frac{\partial n_5}{\partial w_{2,5}}$$
$$= \frac{\partial E}{\partial n_7}\frac{\partial n_7}{\partial n_5}\frac{\partial n_5}{\partial w_{2,5}}$$
$$= \Big(\sum_{i\in data} 2(n_7 - y_7^{(i)})\Big)\frac{\partial g(b + \sum_k w_{k,7}n_k)}{\partial n_5}\frac{\partial g(b + \sum_k w_{k,5}n_k)}{\partial w_{2,5}}$$
$$= \Big(2\sum_{i\in data}(n_7 - y_7^{(i)})\Big)\frac{\partial g(b + w_{5,7}n_5)}{\partial n_5}\frac{\partial g(b + w_{2,5}n_2)}{\partial w_{2,5}}$$
$$= \Big(2\sum_{i\in data}(n_7 - y_7^{(i)})\Big)g'(b + w_{5,7}n_5)\cdot w_{5,7}\cdot g'(b + w_{2,5}n_2)\cdot n_2$$

3.

$$\frac{\partial E}{\partial (v_{1,1})_d} = \frac{\partial E}{\partial n_1}\frac{\partial n_1}{\partial (v_{1,1})_d}$$
$$= \Big(\frac{\partial E}{\partial n_3}\frac{\partial n_3}{\partial n_1} + \frac{\partial E}{\partial n_4}\frac{\partial n_4}{\partial n_1}\Big)\frac{\partial g(b + \sum_l v_{l,i}x_l)}{\partial (v_{1,1})_d}$$
$$= \Big(\frac{\partial E}{\partial n_6}\frac{\partial n_6}{\partial n_3}\frac{\partial n_3}{\partial n_1} + \frac{\partial E}{\partial n_6}\frac{\partial n_6}{\partial n_4}\frac{\partial n_4}{\partial n_1}\Big)\frac{\partial g(b + v_{1,1}x_1)}{\partial (v_{1,1})_d}$$
$$= \Big(\frac{\partial E}{\partial n_6}\frac{\partial n_6}{\partial n_3}\frac{\partial n_3}{\partial n_1} + \frac{\partial E}{\partial n_6}\frac{\partial n_6}{\partial n_4}\frac{\partial n_4}{\partial n_1}\Big)g'(b + v_{1,1}x_1)\cdot (x_1)_d$$

4.

$$\frac{\partial E}{\partial (x_2)_d} = \frac{\partial E}{\partial n_2}\frac{\partial n_2}{\partial (x_2)_d}$$
$$= \Big(\frac{\partial E}{\partial n_4}\frac{\partial n_4}{\partial n_2} + \frac{\partial E}{\partial n_5}\frac{\partial n_5}{\partial n_2}\Big)\frac{\partial n_2}{\partial (x_2)_d}$$
$$= \Big(\frac{\partial E}{\partial n_6}\frac{\partial n_6}{\partial n_4}\frac{\partial n_4}{\partial n_2} + \frac{\partial E}{\partial n_7}\frac{\partial n_7}{\partial n_5}\frac{\partial n_5}{\partial n_2}\Big)\frac{\partial n_2}{\partial (x_2)_d}$$

figure 2:

1.

$$\frac{\partial E}{\partial (v_{2,2})_d} = \frac{\partial E}{\partial n_2}\frac{\partial n_2}{\partial (v_{2,2})_d}$$
$$= \frac{\partial E}{\partial n_4}\frac{\partial n_4}{\partial n_2}\frac{\partial g(b + \sum_l v_{l,i}x_l)}{\partial (v_{2,2})_d}$$
$$= \Big(\frac{\partial E}{\partial n_6}\frac{\partial n_6}{\partial n_4} + \frac{\partial E}{\partial n_8}\frac{\partial n_8}{\partial n_4}\Big)\frac{\partial n_4}{\partial n_2}\frac{\partial g(b + v_{2,2}x_2)}{\partial (v_{2,2})_d}$$
$$= \Big(\frac{\partial E}{\partial n_6}\frac{\partial n_6}{\partial n_4} + \frac{\partial E}{\partial n_8}\frac{\partial n_8}{\partial n_4}\Big)\frac{\partial n_4}{\partial n_2}g'(b + v_{2,2}x_2)\cdot (x_2)_d$$

2.

$$\frac{\partial E}{\partial w_{2,4}} = \frac{\partial E}{\partial n_4}\frac{\partial n_4}{\partial w_{2,4}}$$
$$= \Big(\frac{\partial E}{\partial n_6}\frac{\partial n_6}{\partial n_4} + \frac{\partial E}{\partial n_8}\frac{\partial n_8}{\partial n_4}\Big)\frac{\partial n_4}{\partial w_{2,4}}$$

3.

$$\frac{\partial E}{\partial n_1} = \frac{\partial E}{\partial n_3}\frac{\partial n_3}{\partial n_1} + \frac{\partial E}{\partial n_5}\frac{\partial n_5}{\partial n_1}$$
$$= \Big(\frac{\partial E}{\partial n_6}\frac{\partial n_6}{\partial n_3} + \frac{\partial E}{\partial n_8}\frac{\partial n_8}{\partial n_3}\Big)\frac{\partial n_3}{\partial n_1} + \frac{\partial E}{\partial n_7}\frac{\partial n_7}{\partial n_5}\frac{\partial n_5}{\partial n_1}$$

Task 2:

Assuming K terms made along a single path to the loss. One term will be the differential of the loss term and the preceeding layer of the loss. Then there will be K-1 terms that are the differential from neuron to neuron.

Task 3:

1.

$$64 \times 64 \times 96 \times 2 + 96 = 786528$$

786528 parameters learnt

2.

$$6 \times 6 \times 96 \times 2 + 96 = 7008$$

7008 parameters learnt

3.

$$96 \times 2 + 96 = 288$$

288 parameters learnt