50.039 Theory and Practice of Deep Learning
Theory Homework 5

Joel Huang 1002530

March 1, 2019

1 Backpropagation

Neural network I

1.
$$\frac{\partial E}{\partial n_4} = \frac{\partial E}{\partial n_6} \cdot \frac{\partial n_6}{\partial n_4}$$

2.
$$\frac{\partial E}{\partial w_{2,5}} = \frac{\partial E}{\partial n_7} \cdot \frac{\partial n_7}{\partial n_5} \cdot \frac{\partial n_5}{\partial w_{5,2}}$$

3.
$$\begin{aligned}\frac{\partial E}{\partial (v_{1,1})_d} =& \frac{\partial E}{\partial n_6} \cdot \frac{\partial n_6}{\partial n_4} \cdot \frac{\partial n_4}{\partial n_1} \cdot \frac{\partial n_1}{\partial (v_{1,1})_d} \\ &+ \frac{\partial E}{\partial n_6} \cdot \frac{\partial n_6}{\partial n_3} \cdot \frac{\partial n_3}{\partial n_1} \cdot \frac{\partial n_1}{\partial (v_{1,1})_d}\end{aligned}$$

4.
$$\begin{aligned}\frac{\partial E}{\partial (x_2)_d} =& \frac{\partial E}{\partial n_6} \cdot \frac{\partial n_6}{\partial n_4} \cdot \frac{\partial n_4}{\partial n_2} \cdot \frac{\partial n_2}{\partial (x_2)_d} \\ &+ \frac{\partial E}{\partial n_7} \cdot \frac{\partial n_7}{\partial n_5} \cdot \frac{\partial n_5}{\partial n_2} \cdot \frac{\partial n_2}{\partial (x_2)_d}\end{aligned}$$

Neural network II

1.
$$\begin{aligned}\frac{\partial E}{\partial (v_{2,2})_d} =& \frac{\partial E}{\partial n_6} \cdot \frac{\partial n_6}{\partial n_4} \cdot \frac{\partial n_4}{\partial n_2} \cdot \frac{\partial n_2}{\partial (v_{2,2})_d} \\ &+ \frac{\partial E}{\partial n_8} \cdot \frac{\partial n_8}{\partial n_4} \cdot \frac{\partial n_4}{\partial n_2} \cdot \frac{\partial n_2}{\partial (v_{2,2})_d}\end{aligned}$$

2.
$$\begin{aligned}\frac{\partial E}{\partial w_{2,4}} =& \frac{\partial E}{\partial n_6} \cdot \frac{\partial n_6}{\partial n_4} \cdot \frac{\partial n_4}{\partial w_{2,4}} \\ &+ \frac{\partial E}{\partial n_8} \cdot \frac{\partial n_8}{\partial n_4} \cdot \frac{\partial n_4}{\partial w_{2,4}}\end{aligned}$$

3.
$$\begin{aligned}\frac{\partial E}{\partial n_1} =& \frac{\partial E}{\partial n_6} \cdot \frac{\partial n_6}{\partial n_3} \cdot \frac{\partial n_3}{\partial n_1} \\ &+ \frac{\partial E}{\partial n_7} \cdot \frac{\partial n_7}{\partial n_5} \cdot \frac{\partial n_5}{\partial n_1} \\ &+ \frac{\partial E}{\partial n_8} \cdot \frac{\partial n_8}{\partial n_3} \cdot \frac{\partial n_3}{\partial n_1}\end{aligned}$$

2 Number of products in a path

When calculating the gradient for a fixed neuron, the number of terms in a product corresponding to a single path from E to an arbitrary neuron in layer $k > 1$ is k. Consider the base case where $k = 2$, for neurons z_k in layers $k = 1, 2$ connected via a single path. Then, the gradient product has two terms:

$$\frac{\partial E}{\partial z_2} = \frac{\partial E}{\partial z_1} \cdot \frac{\partial z_1}{\partial z_2}$$

For any value of k, the gradient with respect to that neuron is given by a product of k partials:

$$\frac{\partial E}{\partial z_k} = \frac{\partial E}{\partial z_1} \cdot \frac{\partial z_1}{\partial z_2} \cdots \frac{\partial z_{k-1}}{\partial z_k}$$

Then for $k + 1$, a similar result is observed, with a product of $k + 1$ partials:

$$\frac{\partial E}{\partial z_{k+1}} = \frac{\partial E}{\partial z_1} \cdot \frac{\partial z_1}{\partial z_2} \cdots \frac{\partial z_k}{\partial z_{k+1}}$$

By induction, the number of gradient products is equal to k for all $k > 1$.

3 Convolutional parameters

For a convolutional layer with a square filter of size f, depth d and channels c, the number of parameters is $f^2 \times d \times c + c$.

1. $64 \times 64 \times 2 \times 96 + 96 = 786528$
2. $6 \times 6 \times 2 \times 96 + 96 = 7008$
3. $1 \times 1 \times 2 \times 96 + 96 = 288$