Pattern Recognition Homework 4

Pisit Wajanasara 5931042721

T1.

T1. $x_1 = ReLU(x_0 * w_0 + b_0)$

$y_1 = x_1 * w_1 + b_1$

$z = ReLU(y_1 + x_0)$

Let $x_0 = 1.0$, $w_0 = 0.5$, $w_1 = -0.2$,

$b_0 = 0.1$, $b_1 = 0.15$

Forward Pass

$$x_1 = ReLU(x_0 w_0 + b_0)$$
$$= ReLU(1(0.5) + 0.1)$$
$$= ReLU(0.6) = 0.6$$

$$y_1 = x_1 w_1 + b_1$$
$$= 0.6(-0.2) + 0.15$$
$$= -0.12 + 0.15 = 0.03$$

$$z = ReLU(y_1 + x_0)$$
$$= ReLU(0.03 + 1.0)$$
$$= ReLU(1.03) = 1.03$$

Backward Pass

$$\frac{\partial z}{\partial z} = 1$$

$$\frac{\partial z}{\partial y_1} = \frac{\partial z}{\partial (y_1 + x_0)} \cdot \frac{\partial (y_1 + x_0)}{\partial y_1} = 1 \times 1 = 1 \quad ; y_1 + x_0 \geq 0$$

$$\frac{\partial z}{\partial w_1} = \frac{\partial z}{\partial y_1} \times \frac{\partial y_1}{\partial w_1} = 1 \times \frac{\partial}{\partial w_1}(x_1 w_1 + b_1) = x_1$$

$$\frac{\partial z}{\partial b_1} = \frac{\partial z}{\partial y_1} \times \frac{\partial y_1}{\partial b_1} = 1 \times \frac{\partial}{\partial b_1}(x_1 w_1 + b_1) = 1$$

$$\frac{\partial z}{\partial x_1} = \frac{\partial z}{\partial y_1} \cdot \frac{\partial y_1}{\partial x_1} = \frac{\partial}{\partial x_1}(x_1 w_1 + b_1) = w_1$$

$$\frac{\partial z}{\partial w_0} = \frac{\partial z}{\partial x_1} \cdot \frac{\partial x_1}{\partial (x_0 w_0 + b_0)} \cdot \frac{\partial (x_0 w_0 + b_0)}{\partial w_0}$$

$$= w_1 (1)(x_0) \quad ; \; x_0 w_0 + b_0 \geq 0$$

$$= w_1 x_0$$

$$\frac{\partial z}{\partial b_0} = \frac{\partial z}{\partial x_1} \cdot \frac{\partial x_1}{\partial (x_0 w_0 + b_0)} \cdot \frac{\partial (x_0 w_0 + b_0)}{\partial b_0}$$

$$= w_1 (1)(1) = w_1 \quad ; \; x_0 w_0 + b_0 \geq 0$$

$\therefore$ Gradient of z

$$\frac{\partial z}{\partial w_1} = x_1 = 1 \qquad \frac{\partial z}{\partial w_0} = w_1 x_0 = -0.2$$

$$\frac{\partial z}{\partial b_1} = 1 \qquad \frac{\partial z}{\partial b_0} = w_1 = -0.2$$

T2. Size of A is 98x98x5
Size of B is 96x96910
Size of C is 46x46x10
Size of D is 96x96x10
Size of E is 96x96x5

OT1. Amount to multiplies to compute A is (9*98*98 + 6*2*98 + 3*2*98 + 4 + 2*2 + 1) * 5 * 3 = 1,323,135

Amount to multiplies to compute B is (9*96*96 + 6*2*96 + 3*2*96 + 4 + 2*2 + 1) * 10 * 5 = 4,234,050

Amount to multiplies to compute D is 240100 * 3 * 10 = 7,203,000

You can see that, amount of multiply to D is 7,203,000 is greater than sum of the case of A and B which is 5,557,185.

Amount of parameter in the path to A is 3*3*5 = 45
Amount of parameter in the path to B is 3*3*10 = 90
There are 45 + 90 = 135 parameters for the path to A and B
Amount of parameter in the path to D is 5*5*10 = 250

You can see that, sum of amount of parameters in the path to A and B is 135 which is less than the case if D which is 250.

T3.

T3. Let $L = -\Sigma_j y_j \log P(y=j)$

where $P(y=j) = \frac{\exp(h_j)}{\Sigma_k \exp(h_k)}$ is a softmax layer,

y_j is 1 if y is class j, and 0 otherwise

Prove that $\frac{\partial L}{\partial h_i} = P(y=i) - y_i$

Assume j is the correct class

Case $j=i$:

$$\frac{\partial P(y=j)}{\partial h_i} = \frac{\partial P(y=j)}{\partial h_j} = \frac{\partial}{\partial h_j}\left[\frac{e^{h_j}}{\Sigma_k e^{h_k}}\right]$$

$$= \frac{[\Sigma_k e^{h_k}]\cdot e^{h_j} - e^{h_j}[e^{h_j}]}{(\Sigma_k e^{h_k})^2}$$

$$= \frac{e^{h_j}}{\Sigma_k e^{h_k}} - \left[\frac{e^{h_j}}{\Sigma_k e^{h_k}}\right]^2$$

$$= P(y=j) - P^2(y=j)$$

$$\frac{\partial L}{\partial h_i} = \frac{\partial L}{\partial h_j} = \frac{\partial L}{\partial P(y=j)} \cdot \frac{\partial P(y=j)}{\partial h_j}$$

$$= \frac{\partial}{\partial P(y=j)}\left[-\Sigma_k y_k \log P(y=k)\right] \cdot [P(y=j) - P^2(y=j)]$$

$$= \frac{-1}{P(y=j)} \times P(y=j)[1 - P(y=j)]$$

$$= -1 + P(y=j)$$

$$\therefore \frac{\partial L}{\partial h_i} = P(y=j) - 1 \quad \text{where } i=j$$

Case $j \neq i$:

$$\frac{\partial P(y=j)}{\partial h_i} = \frac{\partial}{\partial h_i}\left[\frac{e^{h_j}}{\Sigma_k e^{h_k}}\right]$$

$$= e^{h_j} \frac{\partial}{\partial h_i} [\Sigma_k e^{h_k}]^{-1}$$

$$= e^{h_j} (-[\Sigma_k e^{h_k}]^{-2})(e^{h_i})$$

$$= \frac{-e^{h_j} e^{h_i}}{[\Sigma_k e^{h_k}]^2}$$

$$= -P(y=j)P(y=i)$$

$$\frac{\partial L}{\partial h_i} = \frac{\partial L}{\partial P(y=j)} \cdot \frac{\partial P(y=j)}{\partial h_i}$$

$$= \frac{\partial}{\partial P(y=j)}\left[-\Sigma_k y_k \log P(y=k)\right] \cdot (-P(y=j)P(y=i))$$

$$= -\frac{1}{P(y=j)} (-P(y=j)P(y=i))$$

$$= P(y=i)$$

$$\therefore \frac{\partial L}{\partial h_i} = P(y=i) \text{ where } i \neq j$$

$\therefore$ Therefore

$$\frac{\partial L}{\partial h_i} = \begin{cases} P(y=i)-1 ; & \text{if it is class } i \\ P(y=i) ; & \text{otherwise} \end{cases}$$