Deep Learning HW WR 3

2018320161 컴퓨터학과 송대선

December 19, 2020

1

l = Wx

s = softmax(l) = $\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{bmatrix} = \begin{bmatrix} \frac{exp(l_1)}{\sum_{i=1}^{n} exp(l_i)} \\ \frac{exp(l_2)}{\sum_{i=1}^{n} exp(l_i)} \\ \vdots \\ \frac{exp(l_m)}{\sum_{i=1}^{m} exp(l_i)} \end{bmatrix}$

L = c = NLL(s) = $\sum_{i=1}^{m} E_i^T(-log(s_i)) = -log(s_y)$

(Where E는 y-th element가 1이고, 나머지는 0인 column 벡터이다.)

$$L = c$$

$$\therefore \frac{dL}{dc} = 1$$

$$\frac{dc}{ds} = \begin{bmatrix} 0 \\ \vdots \\ -s_y^{-1} \\ \vdots \\ 0 \end{bmatrix}$$

$$\frac{dL}{ds} = \frac{dL}{dc}\frac{dc}{ds}$$

$$\frac{dL}{ds} = 1 \times \begin{bmatrix} 0 \\ \vdots \\ -s_y^{-1} \\ \vdots \\ 0 \end{bmatrix}$$

$$\therefore \frac{dL}{ds} = \begin{bmatrix} 0 \\ \vdots \\ -s_y^{-1} \\ \vdots \\ 0 \end{bmatrix}$$

$$\frac{ds}{dl} = \begin{bmatrix} \frac{exp(l_1)}{\sum_{i=1}^m exp(l_i)} \frac{1-exp(l_1)}{\sum_{i=1}^m exp(l_i))} & -\frac{exp(l_1)}{\sum_{i=1}^m exp(l_i)} \frac{exp(l_2)}{\sum_{i=1}^m exp(l_i)} & \cdots & -\frac{exp(l_1)}{\sum_{i=1}^m exp(l_i)} \frac{exp(l_m)}{\sum_{i=1}^m exp(l_i)} \\ -\frac{exp(l_1)}{\sum_{i=1}^m exp(l_i)} \frac{exp(l_2)}{\sum_{i=1}^m exp(l_i)} & \frac{exp(l_2)}{\sum_{i=1}^m exp(l_i)} \frac{1-exp(l_2)}{\sum_{i=1}^m exp(l_i))} & \cdots & -\frac{exp(l_2)}{\sum_{i=1}^m exp(l_i)} \frac{exp(l_m)}{\sum_{i=1}^m exp(l_i)} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{exp(l_1)}{\sum_{i=1}^m exp(l_i)} \frac{exp(l_m)}{\sum_{i=1}^m exp(l_i)} & -\frac{exp(l_2)}{\sum_{i=1}^m exp(l_i)} \frac{exp(l_m)}{\sum_{i=1}^m exp(l_i)} & \cdots & \frac{exp(l_m)}{\sum_{i=1}^m exp(l_i)} \frac{1-exp(l_m)}{\sum_{i=1}^m exp(l_i))} \end{bmatrix}$$

$$\frac{ds}{dl} = \begin{bmatrix} s_1(1-s_1) & -s_1 s_2 & \cdots & -s_1 s_m \\ -s_1 s_2 & s_2(1-s_2) & \cdots & -s_2 s_m \\ \vdots & \vdots & \ddots & \vdots \\ -s_1 s_m & -s_2 s_m & \cdots & s_m(1-s_m) \end{bmatrix}$$

$$\frac{dL}{dl} = \frac{dL}{ds}\frac{ds}{dl}$$

$$\frac{dL}{dl} = \begin{bmatrix} s_1(1-s_1) & -s_1 s_2 & \cdots & -s_1 s_m \\ -s_1 s_2 & s_2(1-s_2) & \cdots & -s_2 s_m \\ \vdots & \vdots & \ddots & \vdots \\ -s_1 s_m & -s_2 s_m & \cdots & s_m(1-s_m) \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ -s_y^{-1} \\ \vdots \\ 0 \end{bmatrix}$$

$$\frac{dL}{dl} = \begin{bmatrix} s_1(s_1-1) & s_1 s_2 & \cdots & s_1 s_m \\ s_1 s_2 & s_2(s_2-1) & \cdots & s_2 s_m \\ \vdots & \vdots & \ddots & \vdots \\ s_1 s_m & s_2 s_m & \cdots & s_m(s_m-1) \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ s_y^{-1} \\ \vdots \\ 0 \end{bmatrix}$$

$$\therefore \frac{dL}{dl} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_y - 1 \\ \vdots \\ s_m \end{bmatrix}$$

$$\frac{dl}{dW} = x^T$$

$$\frac{dL}{dW} = \frac{dL}{dl}\frac{dl}{dW}$$

$$\frac{dL}{dW} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_y - 1 \\ \vdots \\ s_m \end{bmatrix} \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$$

$$\therefore \frac{dL}{dW} = \begin{bmatrix} s_1x_1 & s_1x_2 & \cdots & s_1x_n \\ s_2x_1 & s_2x_2 & \cdots & s_2x_n \\ \vdots & \vdots & \ddots & \vdots \\ s_yx_1 - x_1 & s_yx_2 - x_2 & \cdots & s_yx_n - x_n \\ \vdots & \vdots & \ddots & \vdots \\ s_mx_1 & s_mx_2 & \cdots & s_mx_n \end{bmatrix}$$

2

where

$$x = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 0 \\ 0 & -1 & -2 \end{bmatrix}, W = \begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix},$$

$k = CONV(filter = W, input = x), y = ReLU(k)$이라고 하자.
$k_{11} = 1(-1) + 2(1) + 1(1) + 1(2) = 4$
$k_{12} = 2(-1) + 3(1) + 1(1) + 0(2) = 2$
$k_{21} = 1(-1) + 1(1) + 0(1) - 1(2) = -2$
$k_{22} = 1(-1) + 0(1) - 1(1) - 2(2) = -6$

$$k = \begin{bmatrix} 4 & 2 \\ -2 & -6 \end{bmatrix}$$

$ReLU(x) = max(0, x)$

$$y = \begin{bmatrix} 4 & 2 \\ 0 & 0 \end{bmatrix}$$

Given $\frac{dL}{dy} = \begin{bmatrix} -3 & 2 \\ 0 & 1 \end{bmatrix}$,

if $k_{ij} > 0$ then $\frac{dy}{dk_{ij}} = 1$
if $k_{ij} \leq 0$ then $\frac{dy}{dk_{ij}} = 0$

$$\frac{dL}{dk} = \begin{bmatrix} -3(1) & 2(1) \\ 0(0) & 1(0) \end{bmatrix}$$

$$\therefore \frac{dL}{dk} = \begin{bmatrix} -3 & 2 \\ 0 & 0 \end{bmatrix}$$

$\frac{dL}{dW}$ = CONV(filter = $\frac{dL}{dk}$, input = x)

$\frac{dL}{dW_{11}} = 1(-3) + 2(2) + 1(0) + 1(0) = 1$
$\frac{dL}{dW_{12}} = 2(-3) + 3(2) + 1(0) + 0(0) = 0$
$\frac{dL}{dW_{21}} = 1(-3) + 1(2) + 0(0) - 1(0) = -1$
$\frac{dL}{dW_{22}} = 1(-3) + 0(2) - 1(0) - 2(0) = -3$

$$\therefore \frac{dL}{dW} = \begin{bmatrix} 1 & 0 \\ -1 & -3 \end{bmatrix}$$

3

K_i : i-th layer's number of filters,
S_i : i-th layer's stride,
H_i : i-th layer's height,
W_i : i-th layer's Weight,
P_i : i-th layer's Padding,
C_i : i-th output's number of Channel,

$(C, H, W) = (4, 37, 37)$

$C_a = K_a = 8$
$H_a = (H - F_a + 2P_a)/S_a + 1 = 18$
$W_a = (W - F_a + 2P_a)/S_a + 1 = 18$
$(C_a, H_a, W_a) = (8, 18, 18)$

$C_b = C_a = 8$
$H_b = (H_a - F_b)/S_b + 1 = 9$
$W_b = (W_a - F_b)/S_b + 1 = 9$
$(C_b, H_b, W_b) = (8, 9, 9)$

$C_c = K_c = 16$
$H_c = (H_b - F_c + 2P_c)/S_c + 1 = 4$
$W_c = (W_b - F_c + 2P_c)/S_c + 1 = 4$
$(C_c, H_c, W_c) = (16, 4, 4)$

$C_d = K_d = 16$
$H_d = (H_c - F_d + 2P_d)/S_d + 1 = 4$
$W_d = (W_c - F_d + 2P_d)/S_d + 1 = 4$

$(C_d, H_d, W_d) = (16, 4, 4)$

$C_e = C_d = 16$
$H_e = (H_d - F_e)/S_e + 1 = 2$
$W_e = (W_d - F_e)/S_e + 1 = 2$
$(C_e, H_e, W_e) = (16, 2, 2)$

따라서, 최종 output tensor의 shape은 $(16, 2, 2)$이다.

4

(a) number of parameter : $1 \times 1 \times 64 = 64$
(b) number of parameter : $1 \times 1 \times 96 + 3 \times 3 \times 128 = 1248$
(c) number of parameter : $1 \times 1 \times 16 + 5 \times 5 \times 32 = 816$
(d) number of parameter : $0 + 1 \times 1 \times 32 = 32$
number of parameters in the Inception module : $64 + 1248 + 816 + 32 = 2160$