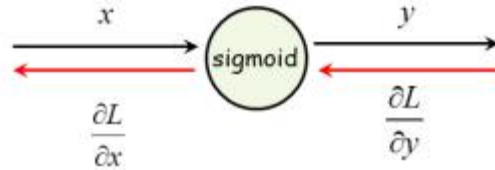


모든 문제는 풀이 과정을 단계별로 명시하세요. (단, log 는 자연로그를 뜻합니다. )

### 1. (sigmoid 함수의 backpropagation)

sigmoid 층으로 data batch 묶음  $X = \begin{pmatrix} \log 2 & \log 3 & \log 4 \\ \log 5 & \log 6 & \log 7 \end{pmatrix}$  가 입력되고,  $\frac{\partial L}{\partial Y} = \begin{pmatrix} 3^2 & 4^2 & 5^2 \\ 6^2 & 7^2 & 8^2 \end{pmatrix}$

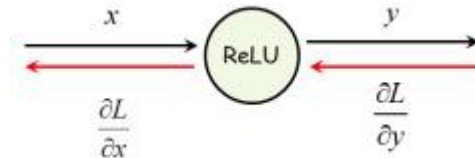
일 때  $\frac{\partial L}{\partial X}$  를 구하세요.



### 2. (relu 함수의 backpropagation)

ReLU 층으로 data batch 묶음  $X = \begin{pmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \end{pmatrix}$  가 입력되고,  $\frac{\partial L}{\partial Y} = \begin{pmatrix} 1 & -2 & -3 \\ 4 & 5 & -6 \end{pmatrix}$  일 때

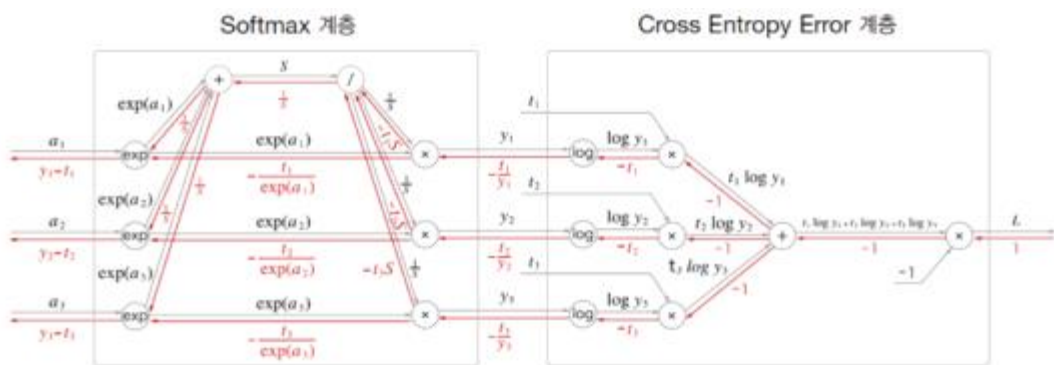
$\frac{\partial L}{\partial X}$  를 구하세요.



### 3. (softmaxwithloss 의 backpropagation)

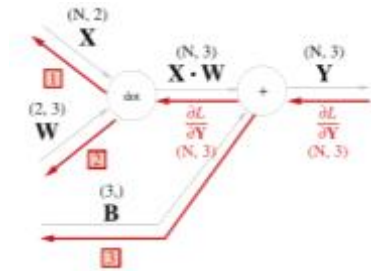
softmaxwithloss 층으로 data batch 묶음  $X = (\log 2 \ \log 3)$  가 입력되고, target label 이

$\begin{pmatrix} 1 & 0 \end{pmatrix}$  일 때  $\frac{\partial L}{\partial X}$  를 구하세요.



4. (affine layer 의 backpropagation)  $X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$ ,  $W = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ ,  $B = (7 \ 8 \ 9)$  으로 주어지

있고,  $\frac{\partial L}{\partial Y} = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$  일 때, Affine 층의 계산 그래프를 이용하여  $\frac{\partial L}{\partial X}$ ,  $\frac{\partial L}{\partial W}$ ,  $\frac{\partial L}{\partial B}$  을 구하세요.



5. (대칭변환)  $X=(x_1, x_2, x_3)$  가  $Y=(x_2, x_3, x_1)$  으로 변환되는 layer가 있습니다.  $\frac{\partial L}{\partial Y} = (d_1, d_2, d_3)$

일 때  $\frac{\partial L}{\partial X}$  를 구하세요. (힌트:  $X \times \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = Y$ )

6. (Momentum algorithm) 이변수 함수  $f(x, y) = x^2 + xy$  를 momentum 방법으로 최적화하고자 합니다. 초기 위치  $x_0 = (1, 1)$  에서 출발하여 3 step 진행할 때,  $x_1, x_2, x_3$  를 구하세요. (단 learning rate  $\eta = 1$ , momentum 계수  $\alpha = 1$ )

7. (NAG algorithm) 이변수 함수  $f(x, y) = x^2 + xy$  를 NAG 방법으로 최적화하고자 합니다. 초기 위치  $x_0 = (1, 1)$  에서 출발하여 3 step 진행할 때,  $x_1, x_2, x_3$  를 구하세요. (단 learning rate  $\eta = 1$ , momentum 계수  $\alpha = 1$ )

8. (AdaGrad algorithm) 이변수 함수  $f(x, y) = x^2 + xy$  를 AdaGrad 방법으로 최적화하고자 합니다. 초기 위치  $x_0 = (1, 1)$  에서 출발하여 2 step 진행할 때,  $x_1, x_2$  를 구하세요. (단 learning rate  $\eta = \frac{1}{2}$ )

9. (RMSProp algorithm) 이변수 함수  $f(x, y) = x^2 + xy$  를 RMSProp 방법으로 최적화하고자 합니다. 초기 위치  $x_0 = (1, 1)$  에서 출발하여 2 step 진행할 때,  $x_1, x_2$  를 구하세요. (단 learning rate  $\eta = \frac{1}{2}$ , forgetting factor  $\gamma = \frac{8}{9}$ )

10. (Adam algorithm) 이변수 함수  $f(x, y) = x^2 + xy$  를 Adam 방법으로 최적화하고자 합니다. 초기 위치  $x_0 = (1, 1)$  에서 출발하여 2 step 진행할 때,  $x_1, x_2$  를 구하세요. (단 learning rate  $\eta = 1$ ,  $\beta_1 = \beta_2 = \frac{1}{2}$ )