1. $Y_{1,i,j} = X_{i+1,j} - X_{i,j}$

$Y_{2,i,j} = X_{i,j+1} - X_{i,j}$

$$W_{1,:,:} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$W_{2,:,:} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$Y_{1,i,j} = \sum_{\alpha=1}^{3} \sum_{\beta=1}^{3} W_{1,\alpha,\beta} \cdot X_{i+(\alpha-2),\, j+(\beta-2)}$$

$W_{1,2,2}$, $W_{1,3,2}$ 만 0이 아니므로

$$= W_{1,2,2} X_{i,j} + W_{1,3,2} X_{i+1,j} = -X_{i,j} + X_{i+1,j}$$

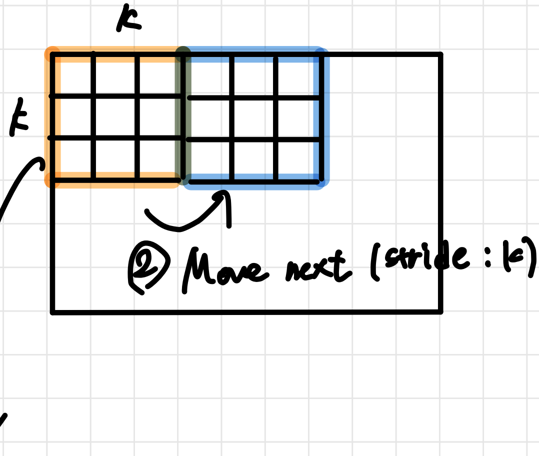
마찬가지로

$$Y_{2,i,j} = \sum_{\alpha=1}^{3} \sum_{\beta=1}^{3} W_{2,i,j} \cdot X_{i+(\alpha-2),\, j+(\beta-2)}$$

$$= W_{2,2,2} X_{i,j} + W_{2,2,3} X_{i,j+1}$$

$$= -X_{i,j} + X_{i,j+1}$$

2. Avg-Pool2d operation 과정은 다음 ①과 ②의 반복이다.



① Take the average over k squared elements

① 로 인해서 $W = \frac{1}{k^2}\begin{bmatrix} 1 & 1 & 1 & \cdots \\ 1 & 1 & \ddots & \\ \vdots & & & \ddots \end{bmatrix}$

② 로 인해서 stride $= k$

3. $W_{1,1,1} = 0.299, \; W_{2,1,1} = 0.587, \; W_{3,1,1} = 0.114$

$$Y_{i,j} = \sum_{r=1}^{3} W_{r,1,1} \cdot X_{r,i,j}$$

$$= 0.299 X_{1,i,j} + 0.587 X_{2,i,j} + 0.114 X_{3,i,j}$$

4. max pool operation 에서 한번 포장이

적용되는 영역 A 안에서만 생각해보자.

nondecreasing 의 정의에 의해 $b \geq a$ 이면

$\sigma(b) \geq \sigma(a)$ 이다.

maxpool operation 에서 어떤 max 값 x_k 를

뽑았다고 해보자. $x_k \geq x_i$ i = 영역안의 모든 index

그러면 $\sigma(x_k) \geq \sigma(x_i)$ 이다.

즉, σ 를 쓰고 후 maxpool operation 해도, $\sigma(x_k)$ 가 뽑힘.

즉, $\sigma(\ell(A)) = \sigma(x_k)$

$\ell(\sigma(A)) = \sigma(x_k)$

임의의 영역 A에 대해 성립하므로 $\sigma(\ell(X)) = \ell(\sigma(X))$

6. (a) $y_L = A_L y_{L-1} + b_L$

$$\therefore \frac{\partial y_L}{\partial b_L} = I, \quad \frac{\partial y_L}{\partial y_{L-1}} = A_L$$

$$y_l = \sigma(A_l y_{l-1} + b_l)$$

$$\frac{\partial y_l}{\partial b_l} = \begin{bmatrix} \frac{\partial y_{l1}}{\partial b_{l1}} & \frac{\partial y_{l1}}{\partial b_{l2}} & \cdots \\ \frac{\partial y_{l2}}{\partial b_{l1}} & \ddots & \\ \vdots & & \end{bmatrix} = \begin{bmatrix} \sigma'(A_l y_{l-1} + b_l)\cdot 1 & 0 & 0 & 0 & \cdots \\ 0 & \sigma'(A_l y_{l-1} + b_l)\cdot 1 & 0 & \cdots & \\ 0 & & & & \\ \vdots & & & & \end{bmatrix}$$

$$= \mathrm{diag}(\sigma'(A_l y_{l-1} + b_l))$$

$$\frac{\partial y_l}{\partial y_{l-1}} = \begin{bmatrix} \frac{\partial y_{l1}}{\partial y_{l-1\,1}} & \frac{\partial y_{l1}}{\partial y_{l-1\,2}} & \cdots \\ \frac{\partial y_{l2}}{\partial y_{l-1\,1}} & \ddots & \\ \vdots & & \end{bmatrix} = \begin{bmatrix} \sigma'(A_l y_{l-1} + b_l) A_l & 0 & 0 & \cdots \\ 0 & \sigma'(A_l y_{l-1} + b_l) A_l & 0 & \cdots \\ \vdots & \vdots & & \end{bmatrix}$$

$$= \mathrm{diag}(\sigma'(A_l y_{l-1} + b_l)) \cdot A_l$$

(h) $A_L \in \mathbb{R}^{1 \times n_{L-1}}$

$$\left(\frac{\partial y_L}{\partial A_L}\right)_{ij} = \frac{\partial y_L}{\partial (A_L)_{1j}} = y_{L-1\,j}$$

($y_L = \sum_{i=1}^{n_{L-1}} (A_L)_{ij} \cdot y_{L-1\,j}$ 이므로)

$$\therefore \frac{\partial y_L}{\partial A_L} = y_{L-1}^T$$

$$\frac{\partial y_L}{\partial (A_l)_{ij}} = \frac{\partial y_L}{\partial y_l} \cdot \frac{\partial y_l}{\partial (A_l)_{ij}}$$

y_l에 $(A_l)_{ij}$ 가 어떤영향을 미치는지 알아본다.

$y_l = \sigma(A_l y_{l-1} + b_l)$

y_l의 k번째원소는 A_l의 k th row 한테만 영향을 받는다.

$$\frac{\partial y_{lk}}{\partial (A_l)_{ij}} = \begin{cases} \sigma'(A_l y_{l-1} + b_l)_k \cdot y_{l-1\,j} & \text{if } i=k \\ 0 & \text{if } i \neq k \end{cases}$$

y_L는 스칼라이고 y_l은 n_l차원 vector 이므로

$\frac{\partial y_L}{\partial y_l}$은 n_l차원 row vector 이다.

$$\frac{\partial y_L}{\partial A_{l_i}}$$

$$=$$

$$\frac{\partial y_L}{\partial y_{l_1}} \begin{array}{c} n_{l-1} \\ \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & & \\ \vdots & & & \end{bmatrix} \end{array} + \cdots + \frac{\partial y_L}{\partial y_{l_i}} \begin{bmatrix} 0 & \cdots & & 0 \\ 0 & \cdots & & 0 \\ \sigma'(A_l y_{l-1} + b_l)_i \cdot y_{l-1}^T & & & \\ 0 & \cdots & - & 0 \\ 0 & \cdots & & 0 \end{bmatrix} + \cdots$$

$$= \frac{\partial y_L}{\partial y_{l_i}} \cdot \sigma'(A_l y_{l-1} + b_l)_i \begin{bmatrix} 0 \\ \vdots \\ y_{l-1}^T \\ \vdots \\ 0 \end{bmatrix}$$

$$= \mathrm{diag}\left(\sigma'(A_l y_{l-1} + b_l)\right) \left(\frac{\partial y_L}{\partial y_l}\right)^T y_{l-1}^T$$