[dw 구하기]

$$Loss = \sum_{i=1}\sum_{j \neq y_i} \max(0, (S_i)_j - (S_i)_{y_i} + 1)$$

↓ W에 대해 미분, i는 무시

$$\max(0, S_j - S_{y_i} + 1)$$

① $\frac{\partial \max}{\partial W} = \frac{\partial \max}{\partial q} \cdot \frac{\partial q}{\partial W}$ ($q = S_j - S_{y_i} + 1$)

↳ $\frac{\partial \max(0,q)}{\partial q} = I(q>0)$

② $\frac{\partial q}{\partial W} = \frac{\partial S_j}{\partial W} - \frac{\partial S_y}{\partial W}$

$\frac{\partial S_j}{\partial W}$ → dw[:, j] += X[i]

$\frac{\partial S_y}{\partial W}$ → dw[:, y[i]] -= X[i]

$S = W^T \cdot x$

$S_j = \sum_k (W^T)_{jk} \cdot x_k$ ($j \neq y_i$) → true label 제외

$\frac{\partial S_j}{\partial W} = \begin{bmatrix} 0 & & 0 \\ 0 & \cdots & 0 \\ 0 & & \end{bmatrix}$ ($D \times C$)

↳ j번째 column의 x값만 남음

$\frac{\partial S_y}{\partial W} = \begin{bmatrix} 0\,0 & & 0 \\ 0\,0 & \cdots & 0 \\ 0\,0 & & 0 \end{bmatrix}$

↳ y번째 열의 x값만 남음

dW는 loss function을 W에 대해 미분한다는 것.

① $dW = \frac{dL}{dW} = \frac{1}{N} \frac{d \sum \max(0, S_j - S_{y_i} + 1)}{dW}$

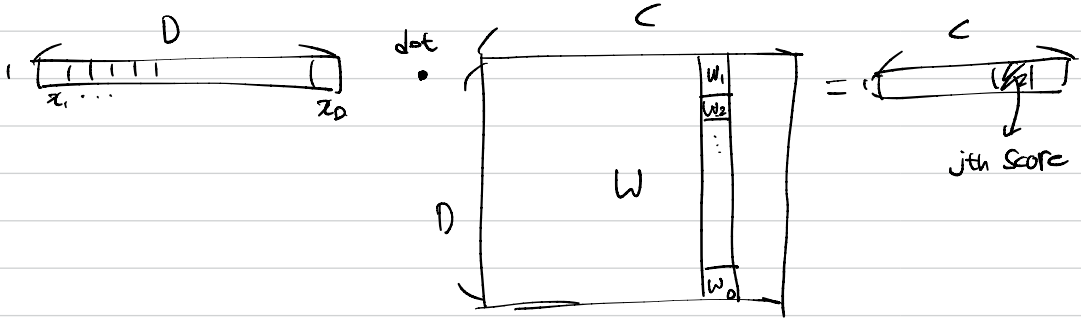
$S = W^T X + b$

② 
각 픽셀을 W에 대해 미분

③ 이는 max function을 W로 미분한 것.

$\frac{d \max(0, S_j - S_{y_i} + 1)}{dW}$ = j번째 class score 미분값을 더하고, y_i 정답 클래스의 미분값을 뺀다

④ image 1장 : X[i]



$\therefore S_j = W_1 \cdot x_1 + W_2 x_2 + \cdots + W_D \cdot x_D$

$\frac{dS_j}{dW} = x_1 + x_2 + \cdots + x_D$ → 즉, input pixels의 합과 같다