$$\frac{\partial (XW)}{\partial W}$$

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

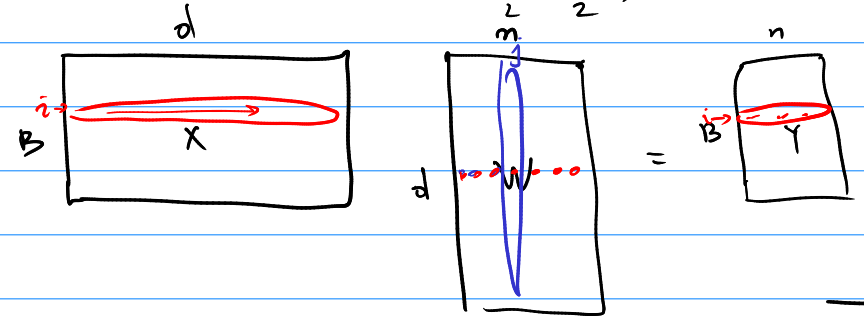
$$\frac{\partial c}{\partial x} = 0$$

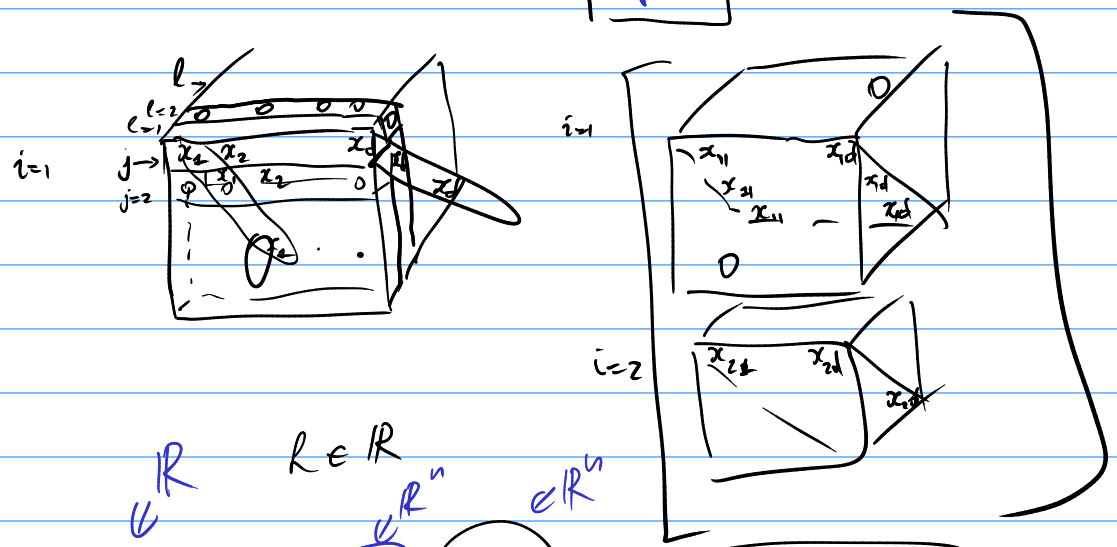
$$\frac{\partial}{\partial w_{kl}} \left(\sum_m x_{im} w_{mj} \right) = \sum_m x_{im} \frac{\partial w_{mj}}{\partial w_{kl}} = \sum_m x_{im} \delta_{mk} \delta_{jl}$$

$$\delta_{ij} = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases}$$

$$= x_{ik} \delta_{jl} \leftarrow$$

$X \otimes \delta$) 4





$R \in \mathbb{R}$

$\frac{\partial R}{\partial W_{kl}} \in \mathbb{R}$ $\quad \frac{\partial R}{\partial Y} \in \mathbb{R}^n$ $\quad \frac{\partial Y}{\partial W_{kl}} \in \mathbb{R}^n$

$$\frac{\partial R}{\partial W_{kl}} = \frac{\partial R}{\partial Y} \cdot \frac{\partial Y}{\partial W_{kl}}, \qquad Y = XW$$

B=1

i=1

B>1

$$= \sum_m \frac{\partial R}{\partial Y_m} \cdot \frac{\partial Y_m}{\partial W_{kl}}$$

$$\frac{\partial (XW)_{ij}}{\partial W_{kl}} = x_{ik} \delta_{jl}$$

$$= \sum_m \frac{\partial R}{\partial Y_m} \cdot x_k \delta_{ml}$$

$X \otimes \frac{\partial R}{\partial Y}$

$$= \frac{\partial R}{\partial Y_l} x_k \Rightarrow \frac{\partial R}{\partial W} = \begin{bmatrix} x \end{bmatrix}_{k} \begin{bmatrix} \frac{\partial R}{\partial Y} \end{bmatrix}_{l} \Rightarrow x_k \frac{\partial R}{\partial Y_l}$$

$\mathbb{R}^{d \times 1}$ $\quad \mathbb{R}^{1 \times n}$ $\quad \mathbb{R}^{d \times n}$

$$\mathcal{L} = \mathcal{L}(Y_1, \dots, Y_B) = \frac{1}{B}\sum_{i=1}^{B} \mathcal{L}_i \qquad \frac{\partial \mathcal{L}_i}{\partial Y_j} = \begin{cases} 0, & i \neq j \\ \frac{\partial \mathcal{L}_i}{\partial Y_i}, & i = j \end{cases}$$

$$\frac{\partial \mathcal{L}}{\partial W_{kl}} = \frac{1}{B}\sum_i \frac{\partial \mathcal{L}}{\partial Y_i}\frac{\partial Y_i}{\partial W_{kl}} = \sum_i \sum_m \frac{\partial \mathcal{L}}{\partial Y_{im}}\frac{\partial Y_{im}}{\partial W_{kl}}$$

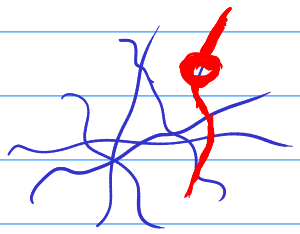
$$= \sum_i \sum_m \frac{\partial \mathcal{L}}{\partial Y_{im}} x_{ik}\,\delta_{ml}$$

$$= \sum_i \frac{\partial \mathcal{L}}{\partial Y_{il}} x_{ik}$$

$$= \sum_i \left(\frac{\partial \mathcal{L}}{\partial Y_{il}}\right)^T x_{ik} = (A^T B)_{ij}$$

$$= \left(\left(\frac{\partial \mathcal{L}}{\partial Y}\right)^T X\right)_{lk} = \left(X^T \frac{\partial \mathcal{L}}{\partial Y}\right)_{kl}$$

$$\Rightarrow \quad \frac{\partial \mathcal{L}}{\partial W} = X^T \frac{\partial \mathcal{L}}{\partial Y} \quad \Big\} = \begin{array}{l} X.T.dot(grad) \\ (grad.T.dot(X).T) \end{array}$$



0.85

3D CNN

Top50