Let $z = f(y)$

$$\frac{\partial z}{\partial y} = \frac{\partial f(y)}{\partial y}$$

$$\frac{\partial z}{\partial x} = \frac{\partial f(g(x))}{\partial g(x)} \cdot \frac{\partial (g(x))}{\partial x}$$

$$\frac{\partial z}{\partial x_i} = \sum_j \frac{\partial z}{\partial y_j} \cdot \frac{dy_j}{\partial x_i}$$

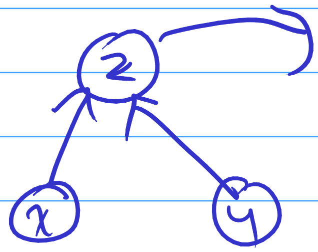
$$\nabla_x z = \frac{\partial z}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x} \\ \vdots & & \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_m} \end{bmatrix}^T \begin{bmatrix} \frac{\partial z}{\partial y_1} \\ \vdots \\ \frac{\partial z}{\partial y_m} \end{bmatrix} = \left[\frac{\partial y}{\partial x}\right]^T \nabla_y z$$

If x is not a vector, but of a higher order (matrix or sth)
then people see as 'x' as a vector as a whole
→ write the matrix as follows and actually expand this internally

[This is brainfuck sh]

To ease us in this process, we use a computational graph, which is a discrete structure

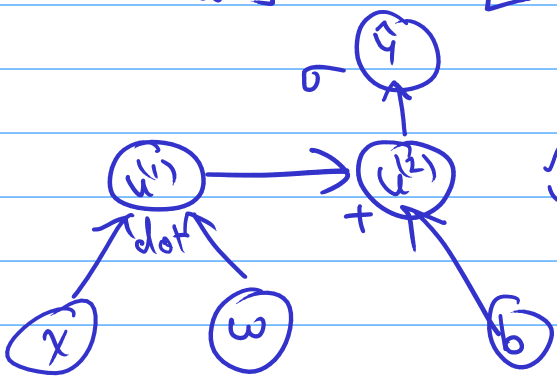
Ex Computational graph of xy



Operation = $\cdot$, multiplication

$z = xy$

$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \qquad W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}$

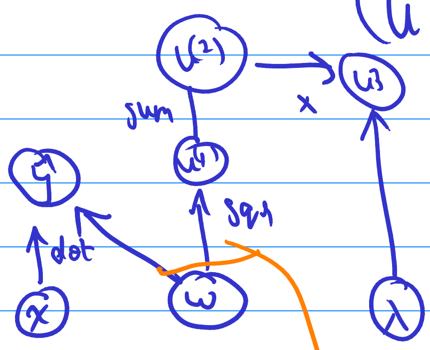


$\hat{y} = \sigma(x^T w + b)$

Two operations on 1 variable

Weight decay (i) $\hat{y}$

(ii) weight decay penalty



(d)

[two]

connections