Wednesday, December 11, 2019 3:39 PM

$x \in \mathbb{R}^D$

$w \in \mathbb{R}^D$

$p(y=1 \mid x,w) = \sigma(w^T x + b)$

$p(y=1 \mid x,w,b) = \sigma(w^T x + b)$

D

π $\sqrt{2}$

X	w	b
$N \times D$	D	1

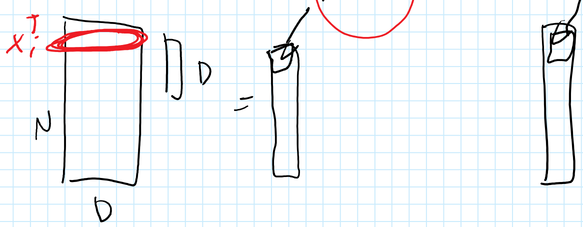
$\hat{y}$

N

$X \cdot w + b$

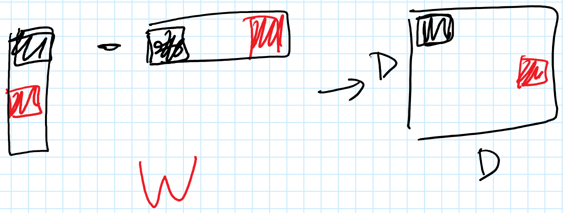
$x_i^T w$

$x_i^T w + b$



x_i

$d(x,y) = \sqrt{\sum_{i=1}^{D} (x_i - y_i)^2}$



W

$[N, N, D]$

$W_{ijk} = X_{ik} - X_{jk}$

$D_{ij} = \sqrt{\sum_k (x_{ik} - x_{jk})^2}$

$$= \sqrt{\sum_k W_{ijk}^2}$$

[D]

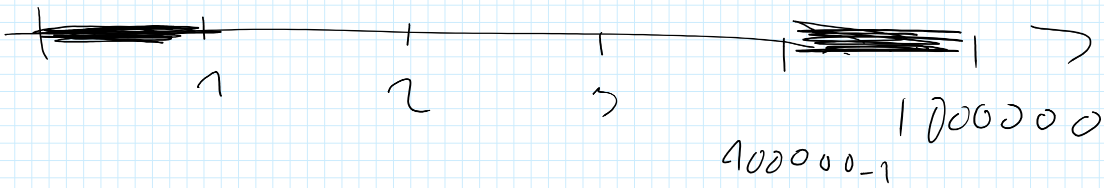
[1,D] [D,1]

2^{64}

double
float64

single
float32

float16



$x = [10, 5, -2, -3.5]$

↓

softmax(x):

$[0.8, 0.1, 0.06, 0.04]$

$$\mathrm{softmax}(x) = \mathrm{softmax}(x - a)$$

$$\frac{\exp(x_i - a)}{\sum_j \exp(x_j - a)} = \frac{\exp(x_i)}{\sum_j \exp(x_j)}$$

$$\overline{\sum_j \exp(x_j - a)} \mathrel{\hat{=}} \overline{\sum_j \exp(x_j)}$$

$$x = [0, 10]$$

$$e^0 = 1$$

$$e^{10} =$$

$$a = f(x)$$

$$b = g(a)$$

$$c = h(b, y)$$

$$\begin{tikzcd} x \arrow[r, "f"] & a \arrow[r, "g"] & b \arrow[r, "h"] & c \\ & & & y \arrow[u] \end{tikzcd}$$

$$\frac{\partial c}{\partial x} = \frac{\partial c}{\partial b} \frac{\partial b}{\partial a} \frac{\partial a}{\partial x}$$

a = f.forward(x)

b = g.forward(a)

c = h.forward(b, y)

$\frac{\partial c}{\partial b}, \frac{\partial c}{\partial y}$ = h.backward(1.)

$\frac{\partial c}{\partial a}$ = g.backward($\frac{\partial c}{\partial b}$)

$\frac{\partial c}{\partial x}$ = f.backward($\frac{\partial c}{\partial a}$) $\quad \frac{\partial c}{\partial a} \cdot \frac{\partial a}{\partial x} = \frac{\partial c}{\partial x}$

$$f(x, y) = x + y = z$$

$$\frac{\partial z}{\partial x} = \underbrace{[1\,1\,1\,1\,1]}_{D}$$

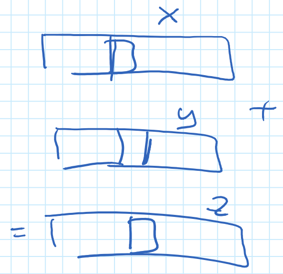
$$a = x \cdot y$$

$$\frac{\partial a}{\partial x} = y$$

$$\frac{\partial a}{\partial y} = x$$

$f(x,y) = x+y$

$\mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}^D$

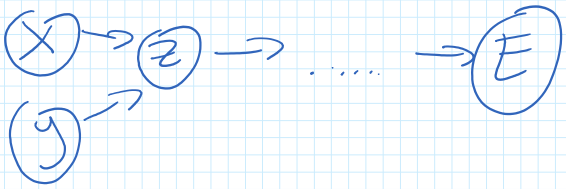


$z = f(x,y) = x+y$

$z_i = x_i + y_i$

$$\frac{\partial z}{\partial x} = \underbrace{\left[\frac{\partial z_i}{\partial x_j} \right]_D}_{D} = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$$

$$\frac{\partial z_i}{\partial x_j} = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i \neq j \end{cases}$$



$z = f.\text{forward}(x,y)$ $\qquad \frac{\partial E}{\partial \dots} \dots$

$\vdots$

$E = \dots$

$\frac{\partial E}{\partial z}$ $1 \times D$

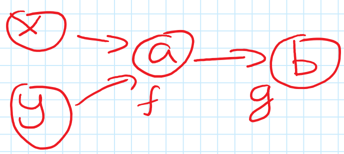
$$\underbrace{\frac{\partial E}{\partial x}}_{1\times D} = \boxed{\underbrace{\frac{\partial E}{\partial z}}_{1\times D} \cdot \underbrace{\frac{\partial z}{\partial x}}_{D\times D}}$$

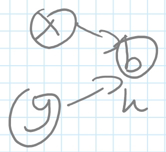
$$\frac{\partial E}{\partial x}, \frac{\partial E}{\partial y} = \underline{f.\text{backward}}\left(\frac{\partial E}{\partial z}\right)$$

$x, y \in \mathbb{R}^D$

$h(x,y) = x^T y \qquad b = h(x,y)$

$z = x^T y \qquad \frac{\partial z}{\partial x} = y \qquad \frac{\partial z}{\partial y} = x$





$[D]\ a = f(x,y) = x \circ y$

$[1]\ b = g(a) = \sum_{i=1}^{D} a_i$

input $\cdot \frac{\partial b}{\partial a}$

$$\boxed{\frac{\partial b}{\partial x} \quad [1,D]}$$

x, y

$a = f.\text{forward}(x,y)$

$b = g.\text{forward}(a)$

input

$\downarrow$

$\frac{\partial b}{\partial a} = g.\text{backward}(1.0)$

$\frac{\partial b}{\partial x}, \frac{\partial b}{\partial y} = f.\text{backward}\left(\frac{\partial b}{\partial a}\right)$

$$\frac{\partial b}{\partial x} = \boxed{\frac{\partial b}{\partial a}\;\frac{\partial a}{\partial x}}$$

f.backward(q):
    return $q\,\frac{\partial a}{\partial x}$, $q \cdot \frac{\partial a}{\partial y}$

X	$[N, D]$
W	$[D, K]$
b	$[K]$
y	$[N, K]$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ & \vdots & \end{bmatrix}$$

$$y_{nk} = \begin{cases} 1 & \text{if sample } n \text{ belongs to class } k \\ 0 & \text{else} \end{cases}$$

$$A = \underset{N\times D}{X} \cdot \underset{D\times K}{W} + \underset{N\times 1}{1_N} \cdot \underset{1\times K}{b^T}$$

$$N\times K \qquad N\times K$$

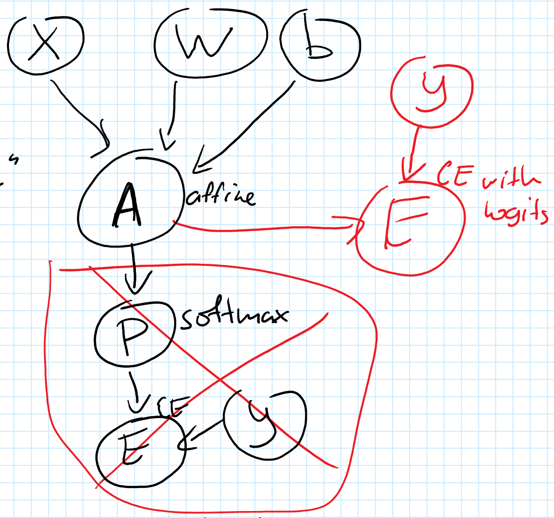
$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} b_1 & \dots & b_K \end{bmatrix} \quad \begin{bmatrix} b_1 & \dots & b_K \\ b_1 & & b_K \\ b_1 & & b_K \\ \vdots & & \vdots \end{bmatrix}$$

$$\underset{N\times K}{A} = \text{Affine.forward}(X, W, b)$$

A_{nk} = "unnormalized probability of n in class k"

$$\underset{N\times K}{P_{nk}} = p(y_{nk} = 1 \mid x_n, W, b)$$

$$P_{nk} = \frac{\exp(A_{nk})}{\sum_c \exp(A_{nc})}$$



$$CE(y, P) = -\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K} y_{nk} \cdot \log p(y_{nk} = 1 \mid x_n, W, b)$$

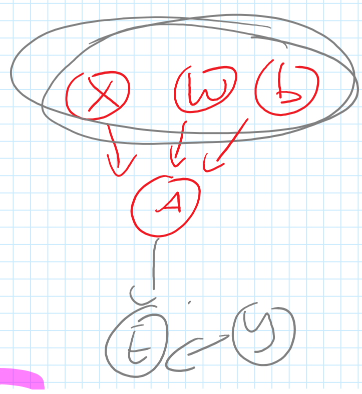
$$= -\frac{1}{N}\sum_n \sum_k y_{nk} \cdot \log P_{nk}$$

$$CE_{w/\,logits}(y, A) = \dots$$

$$A = \text{Affine.forward}(X, W, b)$$

$$E = \text{CE.forward}(A, y)$$

$$\left(\boxed{\underset{N\times K}{\frac{\partial E}{\partial A}}}, \frac{\partial E}{\partial y}\right) = \text{CE.backward}(1.)$$



Goal: find $\frac{\partial E}{\partial W}$, $\frac{\partial E}{\partial b}$

∂A

$N \times L$

$$\frac{\partial E}{\partial X}, \frac{\partial E}{\partial W}, \frac{\partial E}{\partial b} = \text{Affine.backward}\left(\boxed{\frac{\partial E}{\partial A}}\right)$$

$E \leftarrow y$

$$z = g(x) = \sum_i x_i \qquad \leftarrow g.\text{back}\left(\frac{\partial E}{\partial z}\right)$$

$$g: \mathbb{R}^D \to \mathbb{R}$$

$$\underset{1 \times D}{\frac{\partial z}{\partial x}} = [\ \boxed{1}\ \ ] \qquad \frac{\partial z}{\partial x} = [\underbrace{1\ 1\ 1\ 1\ 1}_{D}]$$

$\uparrow$

$\frac{\partial z}{\partial x_i}$