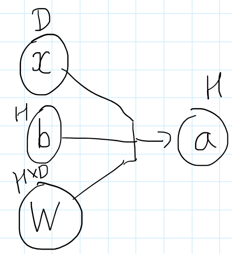
$$\frac{\partial E}{\partial a} \quad [1 \times H]$$




$$a = Wx + b$$

$$\frac{\partial E}{\partial W} \quad [1 \times H \times D]$$

$$[H \times D]$$

$$\begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_D} \\ \frac{\partial y_H}{\partial x_1} & \frac{\partial y_H}{\partial x_D} \end{bmatrix}$$

$$\frac{\partial E}{\partial W_{kl}} = \frac{\partial E}{\partial a_k} \cdot x_l$$

$$y = f(x)$$

$$\frac{\partial E}{\partial a} = \overset{H}{\left[\frac{\partial E}{\partial a_1} \quad \cdots \quad \frac{\partial E}{\partial a_H}\right]}$$

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_D \end{bmatrix} D$$

$$\Bigg[\underbrace{\underset{[H \times 1]}{\left(\frac{\partial E}{\partial a}\right)^T} \cdot \underset{[1 \times D]}{x^T}}_{H \times D}\Bigg]_{kl} = \left(\frac{\partial E}{\partial a}\right)_k \cdot (x_l)$$

$$\left(\frac{\partial E}{\partial a}\right)^T \cdot x^T$$

---

$X \quad [N \times D]$

$y \quad [N \times K] \qquad y_i = [0 \ 0 \ 1 \ 0] \qquad y_i = 3$

$W \quad [D \times K]$

$b \quad [K]$

$$a_i = W^T x_i + b \qquad [K]$$

$$p(y_i = k \mid x_i, W, b) = \text{softmax}(a_i)_k = \frac{e^{a_{ik}}}{\sum_c e^{a_{ic}}}$$

$$A = X \cdot W + \mathbb{1}_N b^T \qquad [N, K]$$

---

$$A[i,:] = a_i$$

$$F \quad [N, K] \qquad F_{nk} = \underline{\log f_k(x_n, W)} =$$

$$y \quad [N, K] \qquad\qquad [N, K]$$

$$-\sum_{n=1}^{N} \sum_{k=1}^{K} y_{nk} \cdot \log f_k(x_n, W) = -(y \circ F).\text{sum}()$$

softmax(A, axis = 1)

[N, K]

F = log- softmax(A, axis = 1)

$F_{nkc}$ log softmax$(a_n)_k$ = $\log e^{a_{nk}} - \log \sum_c e^{a_{nc}}$

A, B, C

np.dot(np.dot(A, B), C)

A@B@C

A.dot(B)

A.dot(B)

x + y

x.__add__(y)

A.shape = [m, n, q]

A[:, np.newaxis, :, :] [m, 1, n, q]

A[:, np.newaxis]

A[:, None, :, None, :] [m, 1, n, 1, q]

m, n, q = A.shape

A.reshape([m, 1, n, 1, q])

$\boxed{[n, q, m]}$

[-1, 1, n, 1, q]

~~[-1, 1, -1, 1, -1]~~

a = Wx + b  $\underset{H \times D}{W}$

$$\begin{bmatrix} a_1 \\ a_i \\ a_H \end{bmatrix} = \begin{bmatrix} W_{11} & W_{1D} \\ & \\ W_{H1} & W_{HD} \end{bmatrix} \begin{bmatrix} x_1 \\ x_D \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_H \end{bmatrix}$$

$$\frac{\partial E}{\partial W_{kl}} = \overbrace{\boxed{\frac{\partial E}{\partial a}} \cdot \frac{\partial a}{\partial W_{kl}}}$$

[1 × H]  [H × 1]

$$= \sum_i \left(\frac{\partial E}{\partial a}\right)_i \cdot \left(\frac{\partial a}{\partial W_{kl}}\right)_i$$

$$= \sum_i \frac{\partial E}{\partial a_i} \cdot \frac{\partial a_i}{\partial W_{kl}}$$

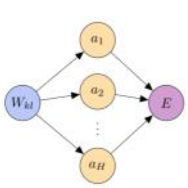
$$= \sum_i \frac{\partial E}{\partial a_i} \cdot \partial$$

$a_i = (Wx + b)_i$

$a_i = \left(\sum_j W_{ij} x_j\right) + b_i$

$\frac{\partial a_i}{\partial W_{kl}} =$

First, let's find $\frac{\partial E}{\partial W_{kl}}$ for some $k, l$




$$\begin{aligned}
\frac{\partial E}{\partial \boldsymbol{a}} \frac{\partial \boldsymbol{a}}{\partial W_{kl}} &= \sum_i \frac{\partial E}{\partial a_i} \frac{\partial a_i}{\partial W_{kl}} \\
&= \sum_i \frac{\partial E}{\partial a_i} \frac{\partial (\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b})_i}{\partial W_{kl}} \\
&= \sum_i \frac{\partial E}{\partial a_i} \frac{\partial (\boldsymbol{W}_{i:}\boldsymbol{x})}{\partial W_{kl}} \\
&= \sum_i \frac{\partial E}{\partial a_i} \frac{\partial (\sum_j W_{ij} x_j)}{\partial W_{kl}} \\
&= \sum_i \frac{\partial E}{\partial a_i} \frac{\partial (\sum_j W_{ij} x_j)}{\partial W_{kl}} \\
&= \sum_i \sum_j \frac{\partial E}{\partial a_i} \frac{\partial W_{ij} x_j}{\partial W_{kl}} = \frac{\partial E}{\partial a_k} x_l
\end{aligned}$$

$\partial W_{kl}$

$W_{i1} \cdot x_1 + \dots + \underbrace{W_{il} \cdot x_l} + \dots W_{iD} x_D$

$\sum_i \frac{\partial E}{\partial a_i} \cdot \frac{\partial a_i}{\partial W_{kl}}$ $\qquad$ $\frac{\partial a_i}{\partial W_{kl}} = \begin{cases} 0 & \text{if } i \neq k \\ x_l & \text{if } i = k \end{cases}$

$$\begin{bmatrix} a_1 \\ \boxed{a_i} \\ a_H \end{bmatrix} = \begin{bmatrix} W_{11} & \cancel{\cdot} & W_{1D} \\ \boxed{\quad W_{ij} \quad W_{kl}} \\ W_{H1} & & W_{HD} \end{bmatrix} \begin{bmatrix} \boxed{\begin{matrix} x_1 \\ x_j \\ x_D \end{matrix}} \end{bmatrix} + \begin{bmatrix} b_1 \\ \boxed{\vdots} \\ b_H \end{bmatrix}$$

$a_i = W_{kl} \cdot x_l + \dots$

$x \in \mathbb{R}^D, \; W \in \mathbb{R}^{H \times D} \quad b \in \mathbb{R}^H$

$a = W^T x W + b$

$D \times H \quad D \times 1 \quad H \times D$

$\boxed{\frac{\partial E}{\partial a}} \cdot \boxed{\frac{\partial a}{\partial x}}$

$a = x^T W x + b$

$1 \times D \quad H \times D \quad D \times 1 + b$

$\underbrace{1 \times D \quad D \times D \quad D \times 1}_{1 \times 1} \quad + b$

$\cancel{b}$ $\quad 1$

$x \quad [D]$

$b \quad [H]$

$W \quad \cancel{[H, D]} \quad [D, D]$

$M, N$

$A \; [M \times N] \qquad A_{ij}$

$\frac{\partial a}{\partial W} \qquad [H] \quad [H \times 1] \quad [H \times 1 \times 1]$

$[H \times D]$

$\frac{\partial a_i}{\partial W_{jk}}$ $\qquad$ $i \quad 1, \dots, H$; $\quad j \quad 1, \dots, H$; $\quad k \quad 1, \dots, D$

$\left(\frac{\partial a}{\partial W}\right)_{ijk} = \frac{\partial a_i}{\partial W_{jk}}$

$f : \mathbb{R}^{H \times D} \to \mathbb{R}^H$

$a = f(W)$

$\frac{\partial E}{\partial W} \quad [1 \times H \times D]$

$H \cdot D$

$H \cdot H \cdot D$

$\frac{\partial E}{\partial W} = \frac{\partial E}{\partial a} \cdot \frac{\partial a}{\partial W}$

$[1 \times H \times D] \quad [1 \times H] \quad [H \times H \times D]$

$\sum_i \frac{\partial E}{\partial a_i} \cdot \frac{\partial a_i}{\partial W} \quad 1 \times H \times D$

$[1 \times H \times D]$ $[1 \times H]$ $[H \times H \times D]$

$\sum_i \underbrace{\frac{\partial E}{\partial a_i}}_{1 \times 1} \cdot \underbrace{\frac{\partial a_i}{\partial W}}_{1 \times H \times D}$ $1 \times H \times D$

$[1 \times H \cdot D] = [1 \times H] \cdot [\underline{H} \times \underbrace{H \cdot D}]$

$W$ $[H \times D]$

$V$ $[M \times N]$

$$\left(\frac{\partial W}{\partial V}\right)_{ijkl} = \frac{\partial W_{ij}}{\partial V_{kl}}$$

$[H \times D \times M \times N]$