I. BACKPROP OF $Y = XW$

let $X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$ $N \times D$

$W = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{bmatrix}$ $D \times M$

$Y = \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \end{bmatrix}$ $N \times M$

L: loss : scalar

$$\frac{dL}{dY} = \begin{bmatrix} \frac{dL}{dy_{11}} & \frac{dL}{dy_{12}} & \frac{dL}{dy_{13}} \\ \frac{dL}{dy_{21}} & \frac{dL}{dy_{22}} & \frac{dL}{dy_{23}} \end{bmatrix}$$

$\underline{\underline{\frac{dL}{dX}}}$

$$\frac{dL}{dx_{11}} = \sum_{i,j} \frac{dL}{dy_{ij}} \frac{dy_{ij}}{dx_{11}} = \frac{dL}{dy_{11}} \cdot w_{11} + \frac{dL}{dy_{12}} \cdot w_{12} + \frac{dL}{dy_{13}} \cdot w_{13}$$

Illy $$\frac{dL}{dx_{21}} = \sum_{i,j} \frac{dL}{dy_{ij}} \frac{dy_{ij}}{dx_{21}} = \frac{dL}{dy_{21}} \cdot w_{11} + \frac{dL}{dy_{22}} \cdot w_{12} + \frac{dL}{dy_{23}} \cdot w_{13}$$

$$\therefore \quad \frac{dL}{dX} = \frac{dL}{dY} \times W^T$$

$N \times D$

$\frac{dL}{dW}$: $$\frac{dL}{dW_{11}} = \sum_{i,j} \frac{dL}{dy_{ij}} \frac{dy_{ij}}{dw_{11}} = \frac{dL}{dy_{11}} \cdot x_{11} + \frac{dL}{dy_{21}} \cdot x_{21}$$

$$\frac{dL}{dw_{22}} = \sum_{i,j} \frac{dL}{dy_{ij}} \frac{dy_{ij}}{dw_{22}} = \frac{dL}{dy_{12}} \cdot x_{12} + \frac{dL}{dy_{22}} \cdot x_{22}$$

$$\therefore \quad \frac{dL}{dW} = X^T \cdot \frac{dL}{dY}$$

$D \times M$

II. BACKPROP THROUGH SOFTMAX LAYER & CROSS ENTROPY LOSS.

$$Y = \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \end{bmatrix} \longrightarrow \text{SOFTMAX} \begin{bmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{23} \end{bmatrix} \longrightarrow \text{CROSS ENTROPY LOSS}$$

3 classes 2 inputs i.e $\underset{\substack{\uparrow \\ \text{batch size.}}}{N} \times C$

$$z_{ij} = \frac{e^{y_{ij}}}{\sum_{j=1}^{3} e^{y_{ij}}}$$

CROSS ENTROPY LOSS $= -\log\left(\frac{e^{\bar{y}_{ij}}}{\sum e^{y_{ij}}}\right)$

↳ log of ~~score~~ output of actual class.

let $L = -\log\left(\frac{e^{y_{12}}}{\sum_{j=1}^{3} e^{y_{1j}}}\right) + -\log\left(\frac{e^{y_{23}}}{\sum_{j=1}^{3} e^{y_{2j}}}\right) = -\log(z_{12}) - \log(z_{23})$

$$\frac{dL}{dy_{11}} = \sum_{ij} \frac{dL}{dz_{ij}} \cdot \frac{dz_{ij}}{dy_{11}} = \frac{dL}{dz_{12}} \cdot \frac{dz_{12}}{dy_{11}} + 0$$

$$\frac{dL}{dy_{12}} = \frac{dL}{dz_{12}} \cdot \frac{dz_{12}}{dy_{12}} \qquad \frac{dL}{dy_{13}} = \frac{dL}{dz_{12}} \cdot \frac{dz_{12}}{dy_{13}}$$

$$\frac{dL}{dz_{12}} = -\frac{1}{z_{12}} \quad ; \quad \frac{dz_{12}}{dy_{11}} = \frac{d}{dy_{11}}\left(e^{y_{12}} \times \left(e^{y_{11}} + e^{y_{12}} + e^{y_{13}}\right)^{-1}\right)$$

$$= -e^{y_{12}}\left(e^{y_{11}} + e^{y_{12}} + e^{y_{13}}\right)^{-2} + 0$$

$$= \frac{-e^{y_{12}} \cdot e^{y_{11}}}{\left(e^{y_{11}} + e^{y_{12}} + e^{y_{13}}\right)^2} = \frac{-z_{12} \cdot e^{y_{11}}}{e^{y_{11}} + e^{y_{12}} + e^{y_{13}}}$$

$$\therefore \frac{dL}{dy_{11}} = \frac{e^{y_{11}}}{e^{y_{11}} + e^{y_{12}} + e^{y_{13}}} =$$

Scanned with CamScanner

IIIly $$\frac{dL}{dy_{13}} = \frac{+e^{y_{13}}}{e^{y_{11}}+e^{y_{12}}+e^{y_{13}}} = z_{13}$$

$$\frac{dL}{dy_{21}} = \frac{e^{y_{21}}}{e^{y_{21}}+e^{y_{22}}+e^{y_{23}}} = z_{21} \qquad \frac{dL}{dy_{22}} = \frac{e^{y_{22}}}{e^{y_{21}}+e^{y_{22}}+e^{y_{23}}} = z_{22}$$

$$\frac{dL}{dy_{12}} = \frac{dL}{dz_{12}} \cdot \frac{dz_{12}}{dy_{12}} = \frac{-1}{z_{12}} \cdot \frac{dz_{12}}{dy_{12}}$$

$$\frac{d}{dy_{12}}\left[\frac{e^{y_{12}}}{e^{y_{11}}+e^{y_{12}}+e^{y_{13}}}\right] = -e^{y_{12}}\left(e^{y_{11}}+e^{y_{12}}+e^{y_{13}}\right)^{-2} \cdot e^{y_{12}} + e^{y_{12}} \cdot \left(e^{y_{11}}+e^{y_{12}}+e^{y_{13}}\right)^{-1}$$

$$= \frac{-e^{y_{12}}}{e^{y_{11}}+e^{y_{12}}+e^{y_{13}}}\left[\frac{e^{y_{12}}}{e^{y_{11}}+e^{y_{12}}+e^{y_{13}}} - 1\right]$$

$$= -z_{12}\left[\frac{e^{y_{12}}}{e^{y_{11}}+e^{y_{12}}+e^{y_{13}}} - 1\right]$$

$$\therefore \frac{dL}{dy_{12}} = -\frac{1}{\cancel{z_{12}}} \times -\cancel{z_{12}}\left[z_{12} - 1\right] = z_{12} - 1$$

IIIly $$\frac{dL}{dy_{23}} = z_{23} - 1$$

. BACKPROP OF MULTI-CLASS SVM LOSS

$y_i = 2$

$$\begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \end{bmatrix} \longrightarrow \text{MULTI CLASS SVM LOSS} \; : \quad L = \sum_{\substack{j=1 \\ j \neq y_i}}^{3} \max(0, s_j - s_{y_i} + 1) + \sum_{j=1}^{3} \max(0, s_j - s_{y_{2j}} + 1)$$

i.e let the correct class for 1st be id = y_{12}

2nd be id = y_{23}

SVM

$\therefore$ Loss $= \max(0, y_{11} - y_{12} + 1) + \max(0, y_{13} - y_{12} + 1) +$

$\max(0, y_{21} - y_{23} + 1) + \max(0, y_{22} - y_{23} + 1)$

$$\frac{dL}{dy_{11}} = 1(y_{11} - y_{12} + 1 > 0) \qquad \frac{dL}{dy_{21}} = 1(y_{21} - y_{23} + 1 > 0)$$

$$\frac{dL}{dy_{13}} = 1(y_{13} - y_{12} + 1 > 0) \qquad \frac{dL}{dy_{22}} = 1(y_{22} - y_{23} + 1 > 0)$$

1 : Indicato
fen

$$\frac{dL}{dy_{12}} = -1(y_{11} - y_{12} + 1 > 0) + -1(y_{13} - y_{12} + 1 > 0)$$

Aly $$\frac{dL}{dy_{23}} = -1(y_{21} - y_{23} + 1 > 0) + -1(y_{21} - y_{23} + 1 > 0).$$

Scanned with CamScanner