Wilbert Aristo Guntoro (1003742)

Date

1. Randomly initialize the weight: $w_0 = \begin{bmatrix} 1 \\ -1 \\ 2 \\ -1 \end{bmatrix}$

Using perceptron algorithm:

When $x_1 = \begin{bmatrix} -5 \\ 1 \\ 3 \\ 1 \end{bmatrix}$, $y_1 = +1$

When $x_2 = \begin{bmatrix} 2 \\ 2 \\ -3 \\ 1 \end{bmatrix}$, $y_2 = -1$

$\therefore y_1 (w_0 \cdot x_1) = (1) \left(\begin{bmatrix} 1 \\ -1 \\ 2 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} -5 \\ 1 \\ 3 \\ 1 \end{bmatrix} \right) = (1)(-5-1+6-1) = -1 < 0$

Update weight:

$$w_1 = w_0 + y_1 x_1$$
$$= \begin{bmatrix} 1 \\ -1 \\ 2 \\ -1 \end{bmatrix} + \begin{bmatrix} -5 \\ 1 \\ 3 \\ 1 \end{bmatrix}$$
$$= \begin{bmatrix} -4 \\ 0 \\ 5 \\ 0 \end{bmatrix}$$

$\therefore y_2 (w_1 \cdot x_2) = (-1) \left(\begin{bmatrix} -4 \\ 0 \\ 5 \\ 0 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ -3 \\ 1 \end{bmatrix} \right) = (-1)(-8-15) = 23 > 0$

↳ don't need update!

$\therefore$ The weights for this linear classifier $= \begin{bmatrix} -4 \\ 0 \\ 5 \\ 0 \end{bmatrix}$ //

Date

2. $z = w_1x_1 + w_2x_2 - w_0$, $g(z) = \begin{cases} z & \text{if } |z| \le 1 \\ \text{sign}(z) & \text{otherwise} \end{cases}$

Positive Cases ($g(z) \ge 0$)

Substitute $x_1 = 0$ & $x_2 = 0$,

$z = -w_0$

$\therefore g(-w_0) \ge 0$

$\therefore$ Let $w_0 = -1$

Substitute $x_1 = 0$ & $x_2 = -1$

$z = -w_2 - w_0 \rightarrow g(-w_2 - w_0) \ge 0$

From above, we let $w_0 = -1$

$g(-w_2 + 1) \ge 0$

$\therefore$ Let $w_2 = 0$

Negative Cases ($g(z) < 0$)

Substitute $x_1 = 1$ & $x_2 = 0$,

$z = w_1 - w_0 \rightarrow g(w_1 - w_0) < 0$

From above, we let $w_0 = -1$

$g(w_1 + 1) < 0$

$\therefore$ Let $w_1 = -2$

Substitute $x_1 = 1$ & $x_2 = -1$,

$z = w_1 - w_2 - w_0 \rightarrow g(w_1 - w_2 - w_0) < 0$

From above, we let $w_0 = -1$, $w_1 = -2$, $w_2 = 0$ $\boxed{\text{CORRECT!}}$

$\therefore g(-2 - 0 + 1) = g(-1) = -1 < 0$

$\therefore$ Set of weights $= \begin{bmatrix} -1 \\ -2 \\ 0 \end{bmatrix}$

$\hookrightarrow w_0 = -1$, $w_1 = -2$, $w_2 = 0$ //

3. (1) $s(a) = \frac{\exp(a)}{1+\exp(a)} = \frac{1}{\exp(-a)+1}$

$$\log(s(a)) = \log\left(\frac{1}{\exp(-a)+1}\right) = \log(1) - \log(\exp(-a)+1)$$
$$= -\log(\exp(-a)+1)$$

$$\frac{\partial \log(s(a))}{\partial(a)} = \frac{\partial(-\log(\exp(-a)+1))}{\partial(a)}$$

$$= -\frac{1}{e^{-a}+1} \cdot \frac{\partial}{\partial a}(e^{-a}+1)$$

$$= -\frac{\frac{\partial}{\partial a}(e^{-a}) + \frac{\partial}{\partial a}(1)}{e^{-a}+1}$$

$$= -\frac{e^{-a} \cdot \frac{\partial}{\partial a}(-a)}{e^{-a}+1}$$

$$= -\frac{-e^{-a}}{e^{-a}+1} = \frac{e^{-a}}{e^{-a}+1}$$

$$= 1 - \frac{1}{e^{-a}+1}$$

$$= 1 - s(a) \quad // \qquad \text{(PROVEN!)}$$

$$\log(1-s(a)) = \log\left(\frac{e^{-a}}{e^{-a}+1}\right)$$
$$= \log(e^{-a}) - \log(e^{-a}+1)$$

$$\frac{\partial \log(1-s(a))}{\partial(a)} = \frac{\partial \log(e^{-a})}{\partial a} + \underbrace{\frac{\partial(-\log(e^{-a}+1))}{\partial a}}_{\text{Part 1!}}$$

$$= \frac{\partial(-a)}{\partial a} + (1-s(a))$$

$$= -1 + 1 - s(a)$$

$$= -s(a) \quad // \qquad \text{(PROVEN)}$$

3 (a) $L = (-1) \cdot \sum_{i=1}^{n} y_i \log(h(x_i)) + (1-y_i)\log(1-h(x_i))$

$= (-1) \cdot \sum_{i=1}^{n} y_i \log(S(w \cdot x_i)) + (1-y_i)\log(1-S(w \cdot x_i))$

$$\nabla_w L = \sum_{i=1}^{n} y_i\left[-x_i(1-S(w \cdot x_i))\right] + (1-y_i)\left[-x_i(-S(w \cdot x_i))\right]$$

$$= \sum_{i=1}^{n} -x_i y_i + x_i y_i S(w \cdot x_i) + x_i S(w \cdot x_i) - x_i y_i S(w \cdot x_i)$$

$$= \sum_{i=1}^{n} x_i\left[S(w \cdot x_i) - y_i\right]$$

$$= \sum_{i=1}^{n} x_i(h(x_i) - y_i) \quad \text{(PROVEN!)}$$

4. Gradient loss of $n_2 = \frac{\partial L}{\partial n_2}$

Gradient loss of $n_3 = \frac{\partial L}{\partial n_3}$

Gradient loss of n_4

$$= \frac{\partial L}{\partial n_2}\frac{\partial n_2}{\partial n_4} + \frac{\partial L}{\partial n_3}\frac{\partial n_3}{\partial n_4}$$

Gradient loss of n_5

$$= \frac{\partial L}{\partial n_4}\frac{\partial n_4}{\partial n_5}$$

$$= \left(\frac{\partial L}{\partial n_2}\frac{\partial n_2}{\partial n_4} + \frac{\partial L}{\partial n_3}\frac{\partial n_3}{\partial n_4}\right)\frac{\partial n_4}{\partial n_5}$$

Gradient loss of n_6

$$= \frac{\partial L}{\partial n_4}\frac{\partial n_4}{\partial n_6} + \frac{\partial L}{\partial n_3}\frac{\partial n_3}{\partial n_6}$$

$$= \left(\frac{\partial L}{\partial n_2}\frac{\partial n_2}{\partial n_4} + \frac{\partial L}{\partial n_3}\frac{\partial n_3}{\partial n_4}\right)\frac{\partial n_4}{\partial n_6} + \frac{\partial L}{\partial n_3}\frac{\partial n_3}{\partial n_6}$$

Gradient loss of n_7

$$= \frac{\partial L}{\partial n_5}\frac{\partial n_5}{\partial n_7} + \frac{\partial L}{\partial n_6}\frac{\partial n_6}{\partial n_7}$$

$$= \left[\left(\frac{\partial L}{\partial n_2}\frac{\partial n_2}{\partial n_4} + \frac{\partial L}{\partial n_3}\frac{\partial n_3}{\partial n_4}\right)\frac{\partial n_4}{\partial n_5}\right]\frac{\partial n_5}{\partial n_7} + \left[\left(\frac{\partial L}{\partial n_2}\frac{\partial n_2}{\partial n_4} + \frac{\partial L}{\partial n_3}\frac{\partial n_3}{\partial n_4}\right)\frac{\partial n_4}{\partial n_6} + \frac{\partial L}{\partial n_3}\frac{\partial n_3}{\partial n_6}\right]\frac{\partial n_6}{\partial n_7}$$

Gradient Loss of x

$$= \frac{\partial L}{\partial n_7}\frac{\partial n_7}{\partial x}$$

$$= \left\{\left[\left(\frac{\partial L}{\partial n_2}\frac{\partial n_2}{\partial n_4} + \frac{\partial L}{\partial n_3}\frac{\partial n_3}{\partial n_4}\right)\frac{\partial n_4}{\partial n_5}\right]\frac{\partial n_5}{\partial n_7} + \left[\left(\frac{\partial L}{\partial n_2}\frac{\partial n_2}{\partial n_4} + \frac{\partial L}{\partial n_3}\frac{\partial n_3}{\partial n_4}\right)\frac{\partial n_4}{\partial n_6} + \frac{\partial L}{\partial n_3}\frac{\partial n_3}{\partial n_6}\right]\frac{\partial n_6}{\partial n_7}\right\}\frac{\partial n_7}{\partial x}$$

Date

5. (i) $C_{j,k} = \sum_{i} A_{ijk} b_i$ $\boxed{\text{Matrix-Vector Multiplication}}$

$\boxed{C_{jk} = A_{ijk} b_i}$ → torch.einsum('ijk,i->jk', [a,b]) //

(ii) $A_{ik} = \sum_{j,l} A_{ijkl}$ $\boxed{\text{Column / Row Sum}}$

$\boxed{A_{ik} = A_{ijkl}}$ → torch.einsum('ijkl->ik', [a]) //

(iii) $A_{ki} = \sum_{j,l} A_{ijkl}$ $\boxed{\text{Column / Row Sum}}$

$\boxed{A_{ki} = A_{ijkl}}$ → torch.einsum('ijkl->ki', [a]) //

(iv.) $C_i = \sum_{j,k} A_{ijk} A_{ijk}$ $\boxed{\text{Matrix - Matrix Multiplication}}$

$\boxed{C_i = A_{ijk} A_{ijk}}$ → torch.einsum('ijk,ijk->i', [a,a]) //

(v.) $\boxed{C = AG^T B}$ → torch.einsum('de,fe,fl->dl', [a,g,b])

$\boxed{\text{Matrix - Matrix Multiplication}}$ //