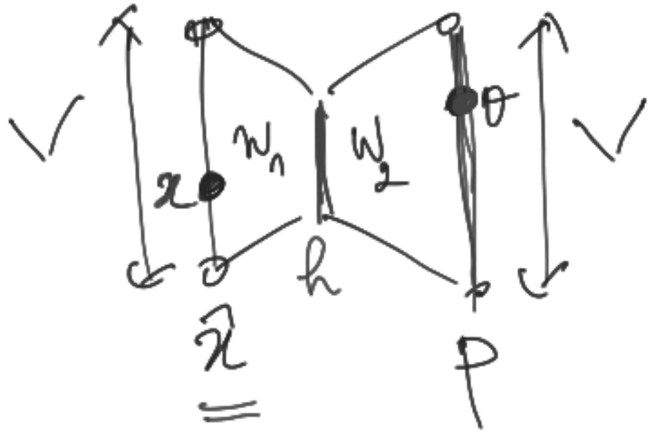


$$P_\theta = \sigma\left(W_1[x]^T W_2[\theta]\right)$$

$(1,D)\quad (D,1)\quad \theta \rightarrow (1,1)$

$$W_1 = \begin{bmatrix} \\ \\ \end{bmatrix}_x \qquad W_2 = \begin{bmatrix} \| \\ \| \\ \| \end{bmatrix}$$

Trick:

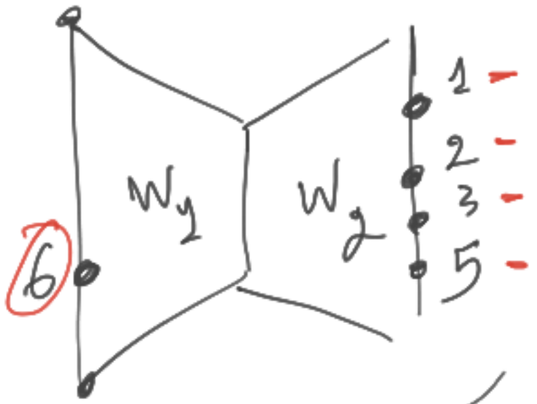
$$\begin{bmatrix} | & | & & | \\ c_1 & c_2 & \cdots & c_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} | \\ c_1 \\ | \end{bmatrix} + x_2 \begin{bmatrix} | \\ c_2 \\ | \end{bmatrix} + \cdots + x_n \begin{bmatrix} | \\ c_n \\ | \end{bmatrix}$$

Binary Classification loss

$x_i \xrightarrow{\text{Model}} p_i$: prediction

target t_i : 0 or 1

$$J = \frac{-1}{N} \sum_{i=1}^{N} \{ t_i \log p_i + \underbrace{(1-t_i)\log(1-p_i)}_{=0} \}$$



positive samples

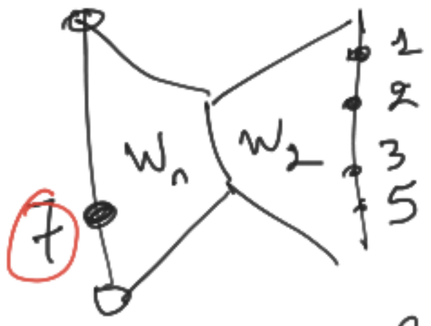
(6,1) (6,2) (6,3) (6,5) : true couple → target $t_i = 1$

$$J^+ = \frac{-1}{4} \sum_{i \in \{1,2,3,5\}} \log p_i \quad : \text{loss function}$$

$$= \frac{-1}{4} \sum_{i \in \{1,2,3,5\}} \log \sigma\left(W_1[6]^\top W_2[i]\right)$$

Binary Cross entropy : $J = -\frac{1}{N} \sum_{i=1}^{N} \{ t_i \log p_i + (1 - t_i) \log(1 - p_i) \}$

$t_i \log p_i = 0$



fake couples

(7, 1)
(7, 2)
(7, 3)
(7, 5)

→ target $t_i = 0$

$$J = \frac{-1}{4} \sum_{i \in \{1,2,3,5\}} \log(1 - p_i)$$

$$= \frac{-1}{4} \sum_{i \in \{1,2,3,5\}} \log\left(1 - \sigma\left(W_1[7]^T W_2[i]\right)\right)$$

$$J^+ = \frac{-1}{4} \sum_{i \in \{1,2,3,5\}} \log\left(\sigma\left(w_1[6]^T w_2[i]\right)\right)$$

parameters $w_1[6]$; $w_2[i]$ $\forall i \in \{1,2,3,5\}$

$$\nabla_{w_1[6]} \left\{ \log\left(\sigma\left(w_1[6]^T w_2[i]\right)\right) \right\} = \nabla_{w_1[6]} g\left(\overbrace{w_1[6]^T w_2[i]}\right)$$

$$= g'\left(w_1[6]^T w_2[i]\right) \cdot \nabla_{w_1[6]} \left\{ w_1[6]^T w_2[i] \right\}$$

$$= \left(1 - \sigma\left(w_1[6]^T w_2[i]\right)\right) \cdot w_2[i]$$

$g(x) = \log(\sigma(x))$ $\forall x \in \mathbb{R}$

$\sigma : x \mapsto \frac{1}{1+e^{-x}}$; $\boxed{g'(x) = 1 - \sigma(x)}$

$\forall z, a \in \mathbb{R}^d$

$$\nabla_z \left(z^T a\right) = a$$

$$J^+ = -\frac{1}{4} \sum_{i \in \{1,2,3,5\}} \log\left(\sigma\left(w_1[6]^T w_2[i]\right)\right)$$

$$\nabla_{w_1[6]} J^+ = -\frac{1}{4} \sum_{i \in \{1,2,3,5\}} \nabla_{w_1[6]} \left\{ \log\left(\sigma\left(w_1[6]^T w_2[i]\right)\right) \right\}$$

$$= -\frac{1}{4} \sum_{i \in \{1,2,3,5\}} \left(1 - \sigma\left(w_1[6]^T w_2[i]\right)\right) w_2[i]$$

$$= \frac{1}{4} \sum_{i \in \{1,2,3,5\}} \left(\sigma\left(w_1[6]^T w_2[i]\right) - 1\right) w_2[i]$$

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

$$J^+ = -\frac{1}{4} \sum_{i \in \{1,2,3,5\}} \log\left(\sigma\left(w_1[6]^T w_2[i]\right)\right)$$

$$\nabla_{w_1[6]} J^+ = \frac{1}{4} \sum_{i \in \{1,2,3,5\}} \left(\sigma\left(w_1[6]^T w_2[i]\right) - 1\right) w_2[i]$$

$$\nabla_{w_2[i]} J^+ = \frac{-1}{4} \sum_{k \in \{1,2,3,5\}} \nabla_{w_2[i]} \left\{\log\left(\sigma\left(w_1[6]^T w_2[k]\right)\right)\right\}$$

$= 0 \quad \forall k \neq i$

$$= -\frac{1}{4} \nabla_{w_2[i]} \left\{\log\left(\sigma\left(w_1[6]^T w_2[i]\right)\right)\right\}$$

$$= +\frac{1}{4} \left(-1 + \sigma\left(w_1[6]^T w_2[i]\right)\right) w_1[6]$$

$$J^- = -\frac{1}{4} \sum_{i \in \{1,2,3,4\}} \underbrace{\log\left(1 - \sigma\left(W_1[\exists]^T W_2[i]\right)\right)}_{(*)}$$

$$\nabla_{W_1[\exists]} \overset{(*)}{=} g'\left(W_1[\exists]^T W_2[i]\right) \nabla_{W_1(\exists)} \left\{ W_1[\exists]^T W_2[i] \right\}$$

$$= -\sigma\left(W_1[\exists]^T W_2[i]\right) W_2[i]$$

$$\begin{cases} g(x) = \log(1 - \sigma(x)) \\ g'(x) = -\sigma(x) \end{cases}$$

$$(g \circ f)' = g' \circ f \times f'$$

$$\nabla_z (z^T a) = a .$$

$$\nabla_{W_1[7]} J^- = +\frac{1}{4} \sum_{i \in \{1,2,3,5\}} + \sigma\left(W_1[7]^T W_2[i]\right) W_2[i]$$

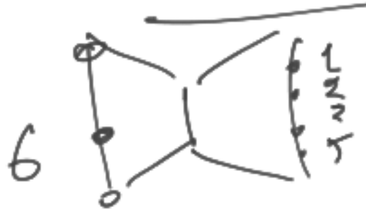
$$\nabla_{W_2[i]} J^- = -\frac{1}{4} \sum_{k \in \{1,2,3,5\}} \underbrace{\nabla_{W_2[i]} \left\{ \log\left(1 - \sigma\left(W_1[7]^T W_2[k]\right)\right) \right\}}_{= 0 \quad \forall k \neq i}$$

$$= -\frac{1}{4} \nabla_{W_2[i]} \left\{ \log\left(1 - \sigma\left(W_1[7]^T W_2[i]\right)\right) \right\}$$

$$= -\frac{1}{4} \left(-\sigma\left(W_1[7]^T W_2[i]\right) \right) W_1[7]$$

$$= \frac{1}{4} \sigma\left(W_1[7]^T W_2[i]\right) W_1[7]$$

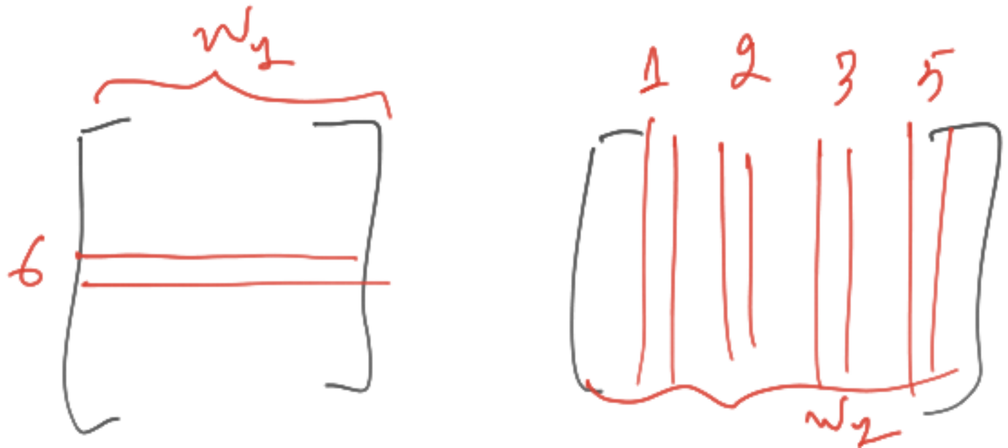
Positive Batch



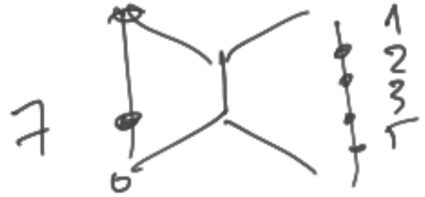
$$W_1[6] \leftarrow W_1[6] - \eta \nabla_{W_1[6]} J^+$$

$\forall \theta \in \{1, 2, 3, 5\}$.

$$W_2[\theta] \leftarrow W_2[\theta] - \eta \nabla_{W_2[\theta]} J^+$$



Negative Batch.



$$W_1[7] \leftarrow W_1[7] - \eta \nabla_{W_1[7]} J^-$$

$\forall \theta \in \{1, 2, 3, 5\}$

$$W_2[\theta] \leftarrow W_2[\theta] - \eta \nabla_{W_2[\theta]} J^-$$