למידה עמוקה תרגיל מס' 1

יובל פנסו
208254128

Q1 Given $x \in \mathbb{R}^d$

$$\text{softmax}(x)_{[i]} = \frac{\exp(x_{[i]})}{\sum_j \exp(x_{[j]})}$$

we will try to ~~prove~~ show that $\text{softmax}(x+c) = \text{softmax}(x)$ for every constant c

$$\text{softmax}(x+c) = \frac{\exp(x_{[i]}+c)}{\sum_j \exp(x_{[j]}+c)} = \frac{e^{x_{[i]}+c}}{\sum e^{x_{[j]}+c}} =$$

$$= \frac{e^{c} \cdot e^{x_{[i]}}}{\sum_j e^{c} \cdot e^{x_{[j]}}} = \frac{\cancel{e^{c}} \cdot e^{x_{[i]}}}{\cancel{e^{c}} \sum_j e^{x_{[j]}}} = \frac{e^{x_{[i]}}}{\sum_j e^{x_{[j]}}}$$

$$= \text{softmax}(x)$$

Q2 Let $x \in \{0,1\}^2$ input vector.

The model: $f(x) = w^T h + b_2$

$h = \max(Ux + b_1, 0)$

In order to achive our goals such that sign(f(x)) is actully a xor, we would like that given $x = \{^1_0\}$ and $x = \{^0_1\}$ $f(x)$ would be positive and give $\{^0_0\}$ and $\{^1_1\}$ $f(x)$ negative.

$$f(x) = (w_1 \; w_2) \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + b_2$$

$h = \max(Ux + b_1, 0) \rightarrow h_1 = \max(Ux_0 + b_{1,0}, 0)$

$\rightarrow h_2 = \max(Ux_1 + b_{1,1}, 0)$

Suppose define

$$U = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad b_1 = 0$$

$$W = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad b_2 = -0.5$$

$x = \begin{Bmatrix} 0 \\ 1 \end{Bmatrix}$ $\quad f(x) = (1, 1)\left(\max\left[\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right]\right) - 0.5 =$

$= (1,1)\left(\max\left(\begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right)\right) - 0.5 = (1,1) \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} - 0.5$

$= 1 - 0.5 = 0.5 > 0 \Rightarrow \text{sign}\left(f\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = 1$

$x = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix}$ $\quad f(x) = (1,1)\left[\max\left(\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)\right] - 0.5 =$

$= (1,1)\left[\max\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right)\right] - 0.5 = (1,1)\begin{pmatrix} 1 \\ 0 \end{pmatrix} - 0.5 = 0.5$

$\Rightarrow \text{sign}\left(f\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = 1$

$x = \begin{Bmatrix} 1 \\ 1 \end{Bmatrix}$ $\quad f(x) = (1,1)\left[\max\left(\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right)\right] - 0.5 =$

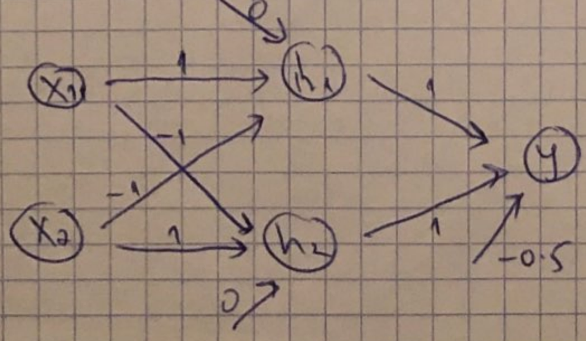
$= (1,1)\left[\max\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right)\right] - 0.5 = (1,1)\begin{pmatrix} 0 \\ 0 \end{pmatrix} - 0.5 = -0.5$

$\Rightarrow \text{sign}\left(f\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) = 0$

$x = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$ $\quad f(x) = (1,1)\left[\max\left(\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right)\right] - 0.5 =$

$= (1,1)\left[\max\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right)\right] - 0.5 = (1,1)\begin{pmatrix} 0 \\ 0 \end{pmatrix} - 0.5 = -0.5$

$\Rightarrow \text{sign}\left(f\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) = 0$

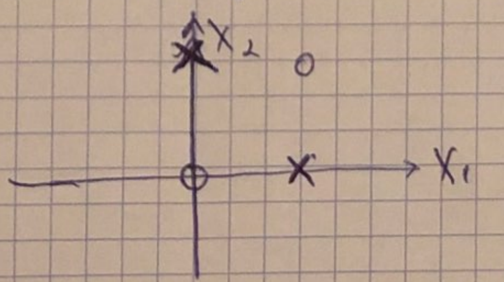


2. In the given model, if we discard the max operation/Layer, achieving XOR Function <u>isn't possible</u>.

In order to achieve XOR we need to have non-linearity to make the non-linear transform to a place where then we can do linear seperation.

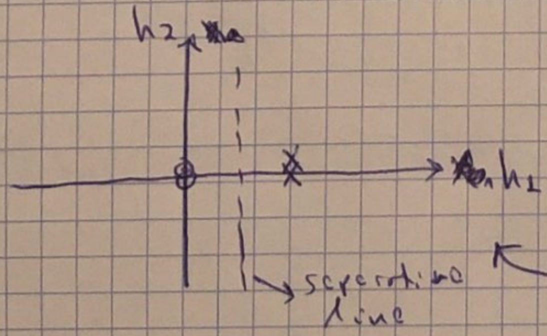
Without the non linearity there is no way to transform $\begin{Bmatrix}1\\0\end{Bmatrix}$ & $\begin{Bmatrix}0\\1\end{Bmatrix}$ to one area and $\begin{Bmatrix}1\\1\end{Bmatrix}$ & $\begin{Bmatrix}0\\0\end{Bmatrix}$ to the second area such that those areas are seperatable by a linear line.

This phenomena best shown visually:



no line can seperate the dots to one side of the line and the X's to the other side

⇓ non-linear transform



non linear transform like the max layer as shown above cause the dots and X's to be seperatable by a line as shown in the graph