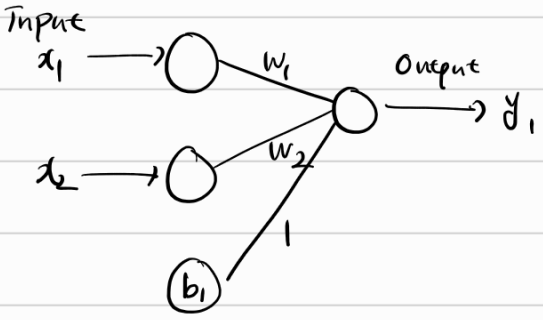
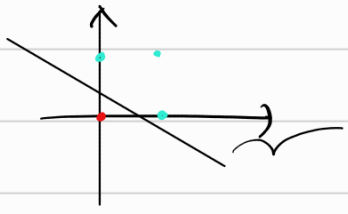
1. MLP?

Def 1.1 자코크조 관점에서 퍼셉트론



Note 1.2 And 게이트를 만들 수 있다

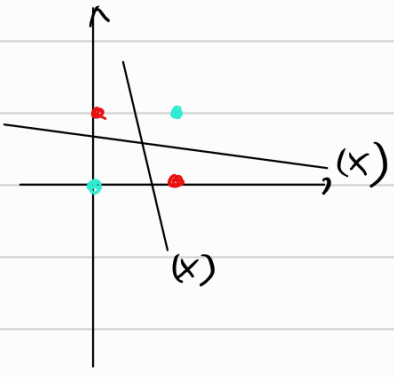


$$w_1x_1 + w_2x_2 + b_1 = y_1$$

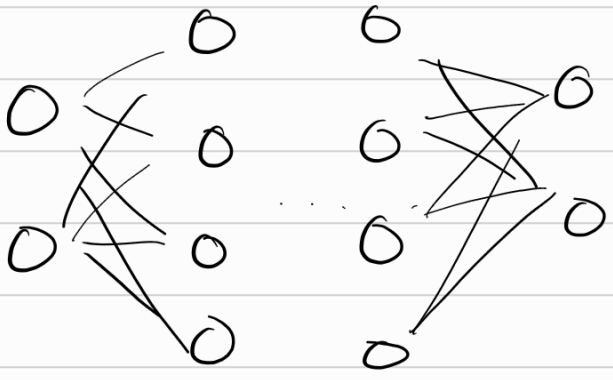
Rmk 1.3 XOR 나타내기 불가능 (2층 이상이 필요)

↳ 즉 논리는 XOR을 NAND, OR 로 만들 수 있다!

↳ Perceptron Convergence Theorem

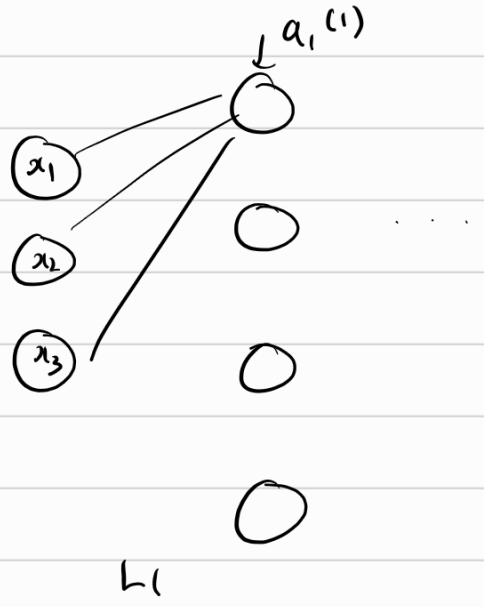


Def 1.4 MLP

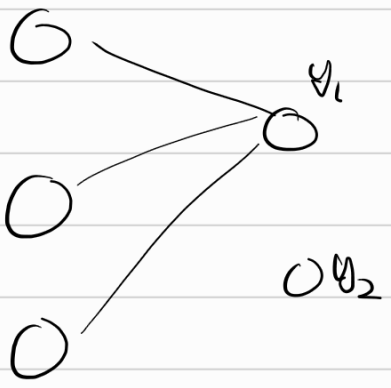


← Hidden layer →

2. Forward, 신호 전달



L_1



L_n

$$a_1 = h\left(\sum_i w_i x_i\right)$$

$$\Rightarrow \vec{a}^{(1)} = h\left(\underset{4\times3}{W^{(1)}}\underset{3\times1}{X^{(1)}} + \underset{4\times1}{B^{(1)}}\right)$$

batch size

$$\Rightarrow y_1 = \text{soft}\left(W^{(n)}X^{(n)} + B^{(n)}\right)$$

$$\Rightarrow \text{Softmax} : O \longrightarrow [0,1]$$

$$k \longmapsto \frac{e^{a_k}}{\sum_i e^{a_i}}$$

$$\frac{\exp(a_k + C')}{\sum_{i=1}^{n} \exp(a_i + C')}$$

(C'을 조절해서 분모가 과도하게 커지는 것 방지!)

3. Loss

Def 3.1 MSE, Cross Entropy

① $\frac{1}{2}\sum_i (y_i - t_i)^2$

② $-\sum_i t_i \log y_i$ (mini batch: $-\frac{1}{N}\sum_n \sum_i t_i \log y_i$)

Note 3.2

if Loss = accuracy, $\frac{dL}{dw} = 0$ (often) so learn ✗

Def 3.3 Numerical gradient

$\lim_{h \to 0} \frac{f(x+h) - f(x-h)}{2h} \simeq \frac{\partial f}{\partial x}$ (take $h = 10^{-4}$)

Def 3.4 Learn

$$w_{ij} = w_{ij} - \alpha \frac{\partial L}{\partial w_{ij}} \quad \forall_{ij}$$

4. Back Propagation

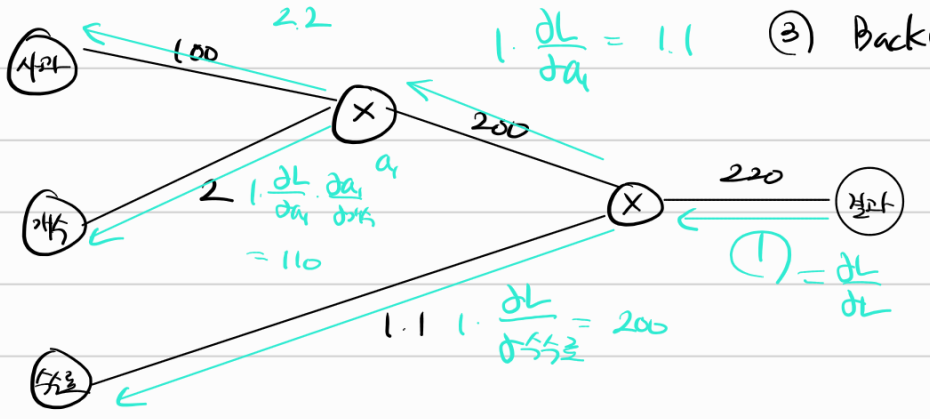
Note

① Chain Rule

② 곱의 미분

③ Backward is locally!

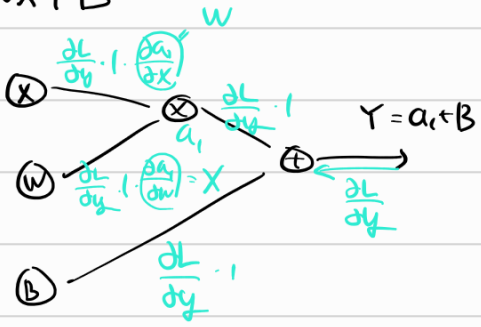
Def 4.1 cal graph



Thm 4.2 Backward for each Layer

① Affine

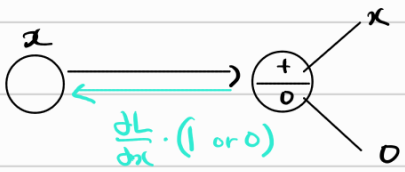
$Y = WX + B$



ex) $W_{4\times3}$, $X_{3\times50}$, $B_{4\times50}$

$$\Rightarrow \begin{cases} \underset{3\times50}{\frac{\partial L}{\partial X}} = \underset{3\times4}{W^T} \times \underset{4\times50}{\frac{\partial L}{\partial Y}} \\ \underset{4\times3}{\frac{\partial L}{\partial W}} = \underset{4\times50}{\frac{\partial L}{\partial Y}} \times \underset{50\times3}{X^T} \\ \underset{4\times50}{\frac{\partial L}{\partial B}} = \frac{\partial L}{\partial Y} \end{cases}$$

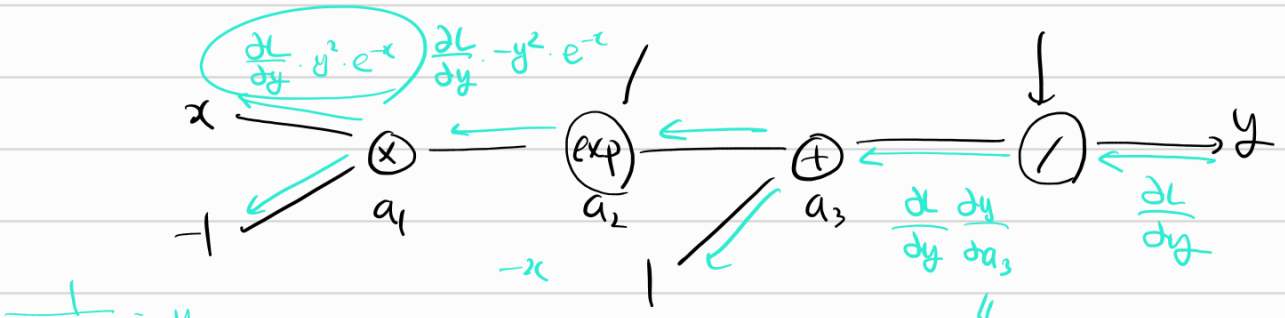
② Relu



③ Sigmoid $\frac{1}{1+e^{-x}}$

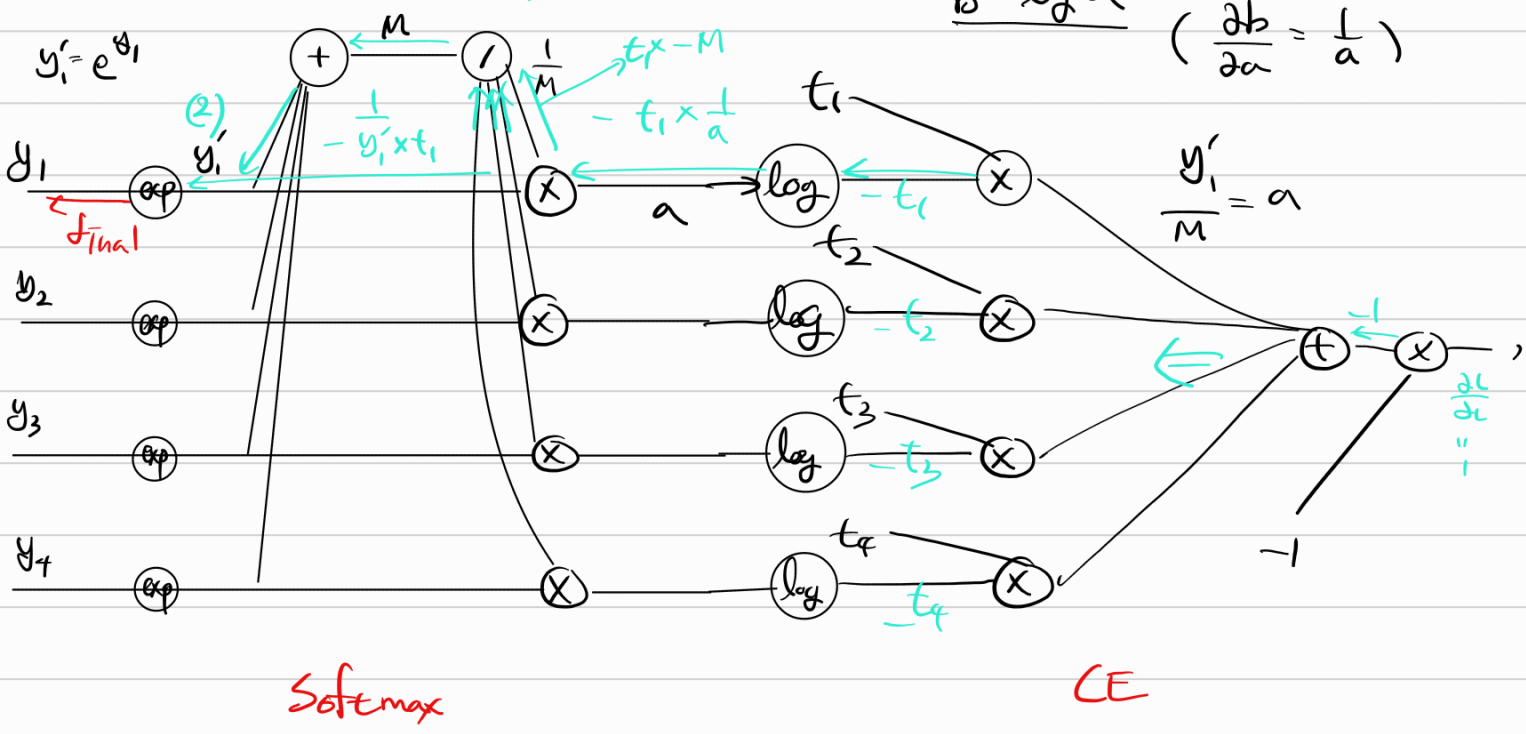
$\left(\frac{1}{y}-1\right) \cdot y^2 \cdot \frac{\partial L}{\partial y} = y(1-y)\frac{\partial L}{\partial y}$

$(a_2 = e^{a_1},\ a_2' = e^{a_1})$ $\quad (y = \frac{1}{x} \Rightarrow y' = -\frac{1}{x^2})$



④ Softmax with Loss $\quad y_k = \frac{e^{y_k}}{\sum_i e^{y_i}}$ ... (1)

$b = \log a \quad (\frac{\partial b}{\partial a} = \frac{1}{a})$



④-(1) : $-M(t_1+t_2+t_3+t_4) \Rightarrow -M \times -\frac{1}{M^2} = \frac{1}{M}$

$\Vert$

④-(2) : $\left(\frac{1}{M} - \frac{1}{y_1'}t_1\right)$

final : $\left(\frac{1}{M} - \frac{1}{y_1'}t_1\right) \times y_1'$

$= \frac{y_1'}{M} - t_1 = a_1 - t_1$ ★