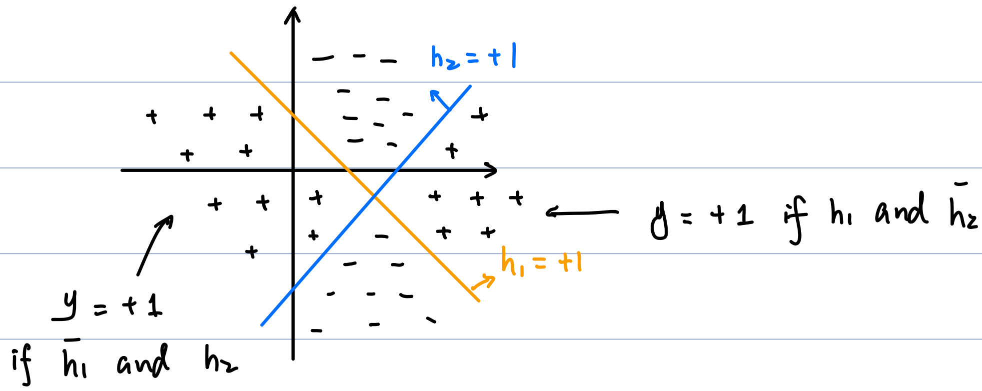
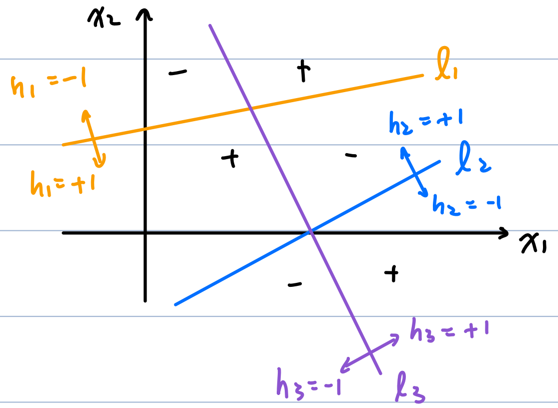
Lec 14.1 MLP continued

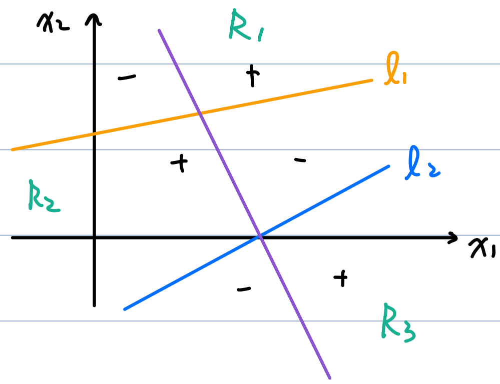
1. MLP continued

↳ Recall:



↳ what if:





↳ fact: Any decision rule that can be expressed by iteration of hyper planes can be interruped using a 3-layer MLP

- R_1 : $h_1 = -1$, $h_2 = +1$, $h_3 = +1$, $\equiv$ AND $(\bar{h}_1, h_2, h_3)$

  R_2 : $h_1 = +1$, $h_2 = +1$, $h_3 = -1$, $\equiv$ AND $(h_1, h_2, \bar{h}_3)$

  R_3 : $h_1 = -1$, $h_2 = +1$, $h_3 = +1$, $\equiv$ AND $(h_1, \bar{h}_2, h_3)$

- $y = +1$ if $\underline{x} \in R_1 \cup R_2 \cup R_3$

  $\therefore y = $ OR $[$AND$(\bar{h}_1, h_2, h_3)$, AND$(h_1, h_2, \bar{h}_3)$, AND$(h_1, \bar{h}_2, h_3)]$

  ↑ 3-layer MLP

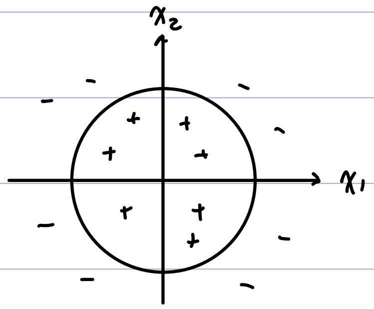
• Hidden layer 1: compute h_1, h_2, h_3

2: AND function

3: OR function

↳ <u>what if non-linear ?</u>

• e.g.



⇒ can use many tangent lines for better argument

Target function:

$y = \text{sign}(1 - x_1^2 - x_2^2)$

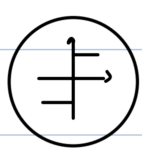
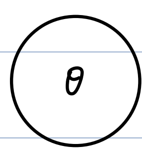
• Fact: A 3-layer MLP with sufficient many hidden units in each layer can approxmate any smooth decision boundary

Problem: hard to train the machine

<u>Lec 14.2 Neural Network</u>

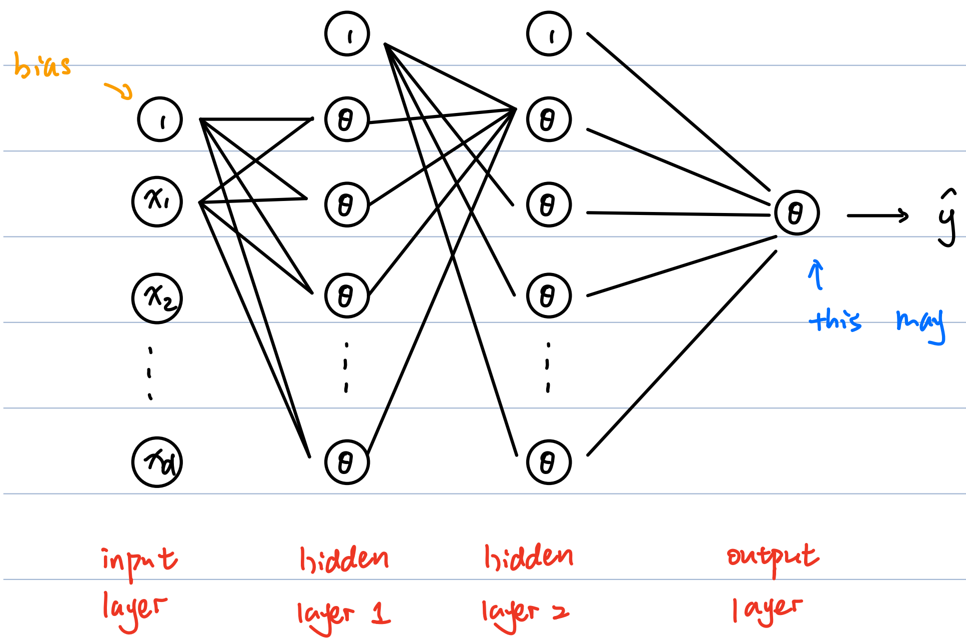
1. <u>General idea:</u>

↳ Usually use a 'softer' func than sign(·) for easier opitimization / learning.

from  to 

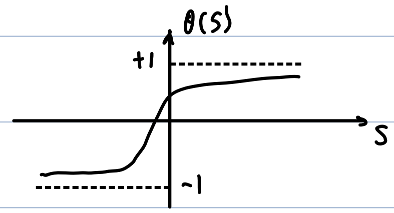
↳ eg:

// $\theta(\cdot)$: activation function



↳ $\theta(\cdot)$: activation function

e.g. $\theta(s) = \tanh(s) = \dfrac{e^s - e^{-s}}{e^s + e^{-s}}$



// " Universal approxmiation theorem "

↳ Properties of $\theta(s)$:

- $\theta(s) \approx s$ if $|s| \approx 0$
- straturation if $|s| \gg 1$
- $\theta'(s) = 1 - \theta^2(s)$

2. Notation:

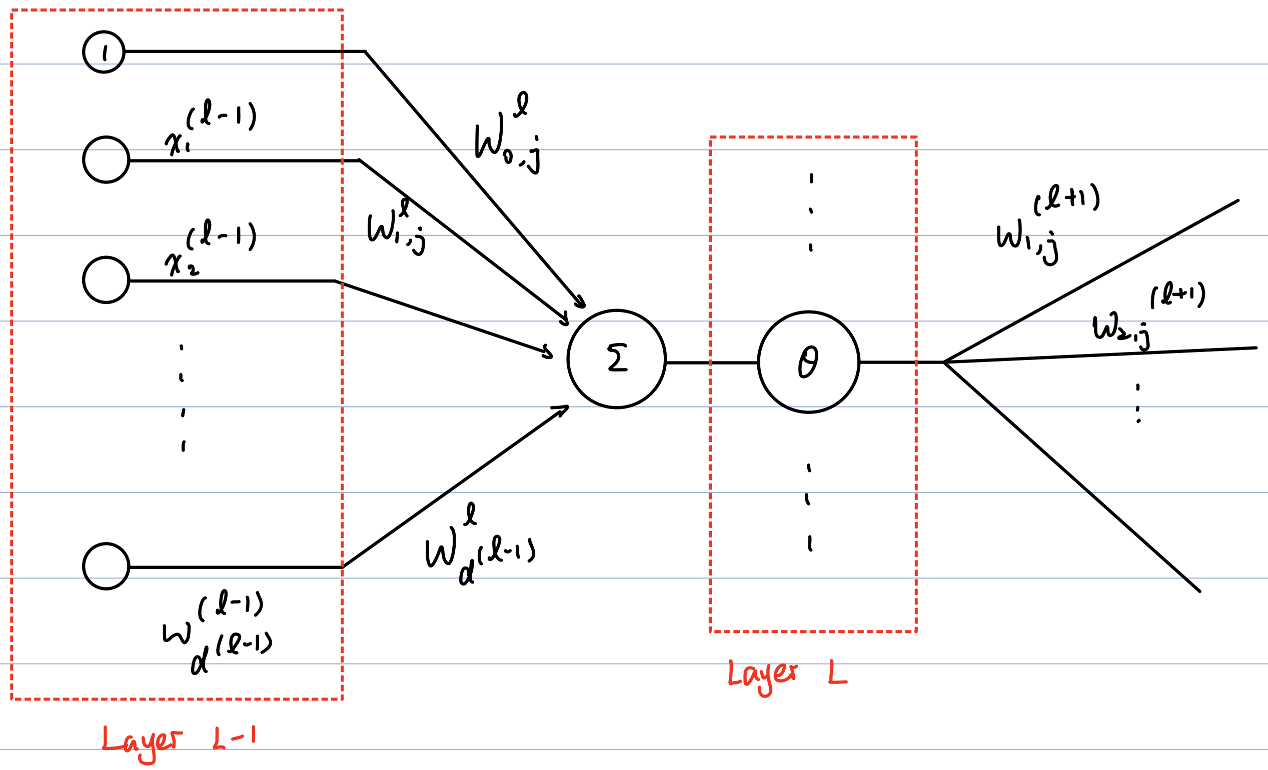
↳ input layer : $l = 0$

hidden layer : $1 \leq l \leq L-1$

output layer : $l = L$

↳ $d^{(\ell)}$: No. of nodes in layer ℓ (excluding bias) , $0 \le \ell \le L$

$W_{i,j}^{(\ell)}$: weight connection node i in layer $\ell - 1$ to node j in layer ℓ.



Input to node j in layer ℓ: $S_j^{(\ell)}$

Output to node j in layer ℓ: $x_j^{(\ell)}$

$$x_j^{(\ell)} = \theta\left(s_j^{(\ell)}\right) \quad ; \quad S_j^{(\ell)} = W_{0,j}^{(\ell)} + \sum_{i=1}^{d^{(\ell-1)}} W_{i,j}^{(\ell)} \cdot x_j^{(\ell-1)}$$

(chaper 7.1)