심층 신경망

Deep Neural Network (DNN)

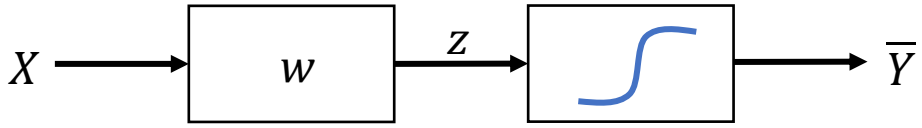
Agenda

- History of MLP

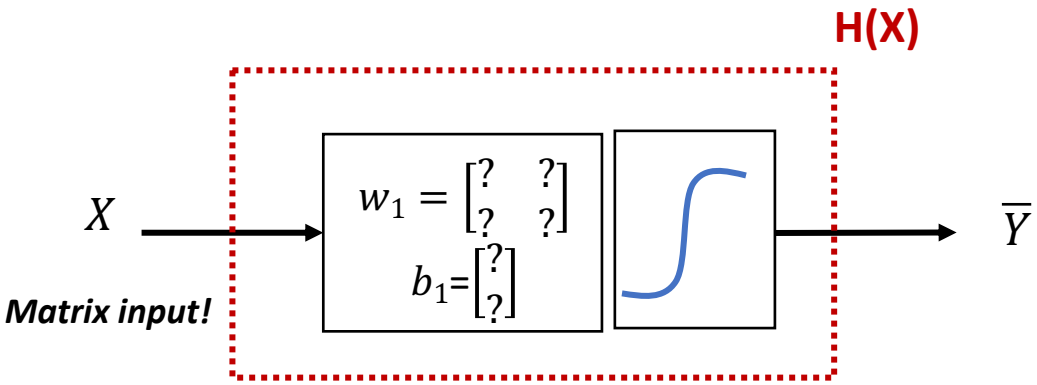
- Neural Net(NN)
  - Perceptron
  - Multi-Layer Perceptron (MLP)
  - Backpropagation

Problem: Perceptron

- One logistic regression unit cannot separate XOR



[실습] Perceptron for XOR



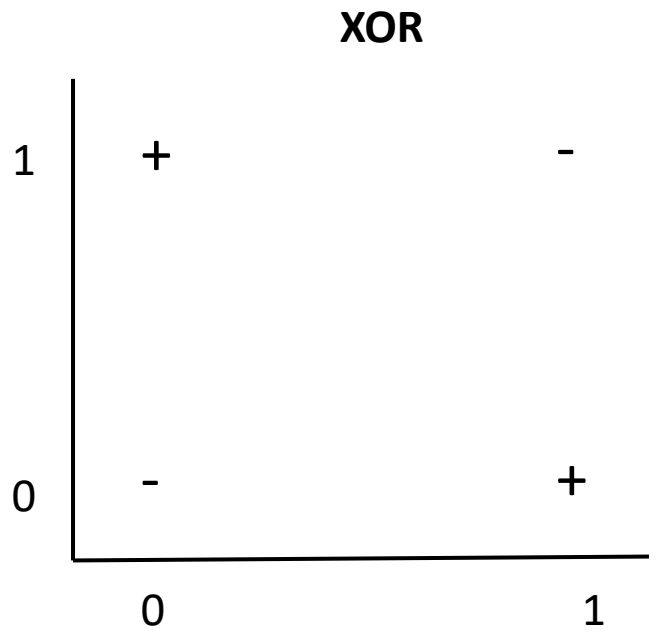
$\overline{Y}$ = H(X) = sigmoid(X·w_1+b_1)

X_1	X_2	$\overline{Y}$
0	0	0
0	1	1
1	0	1
1	1	0

Problem: Perceptron

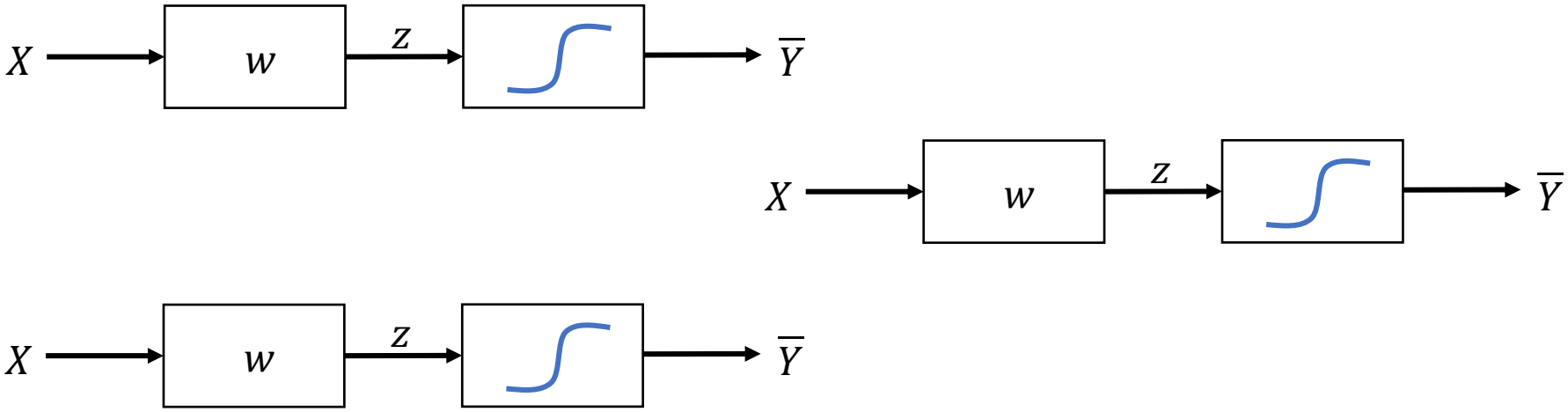
- XOR using NN

X1	X2	XOR	
0	0	0	-
0	1	1	+
1	0	1	+
1	1	0	-



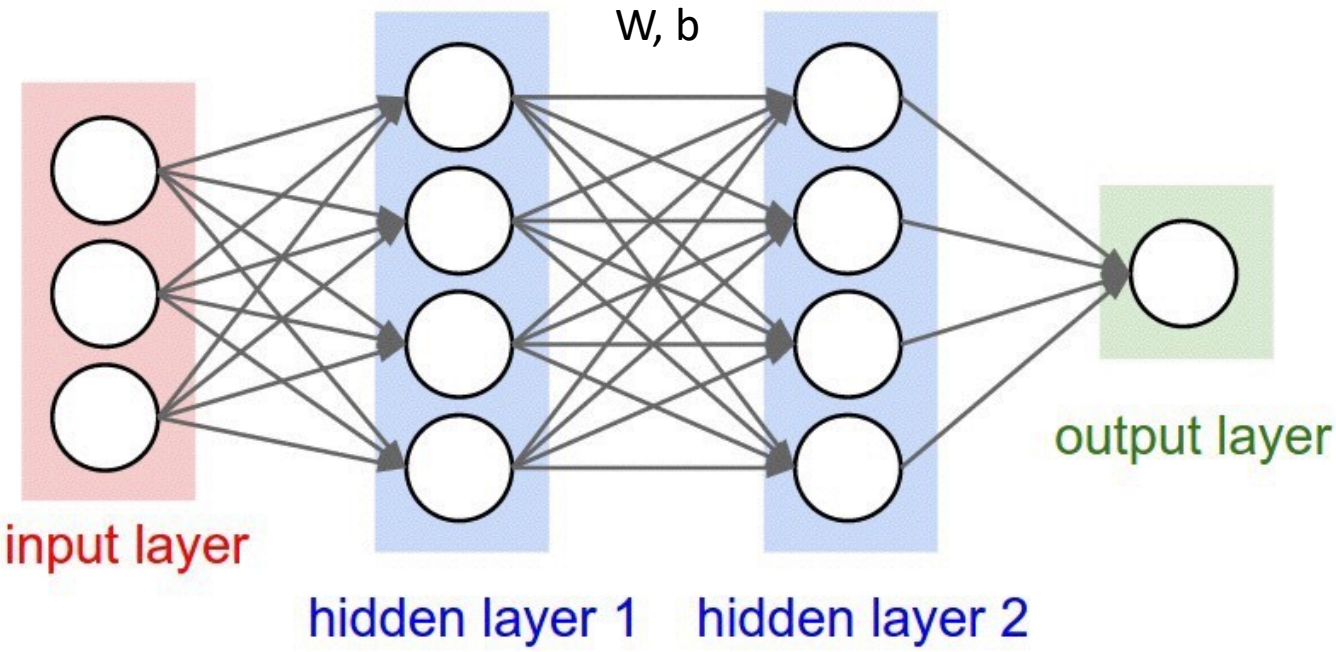
Solution: MLP

- Multiple logistic regression units can separate XOR



Solution: MLP

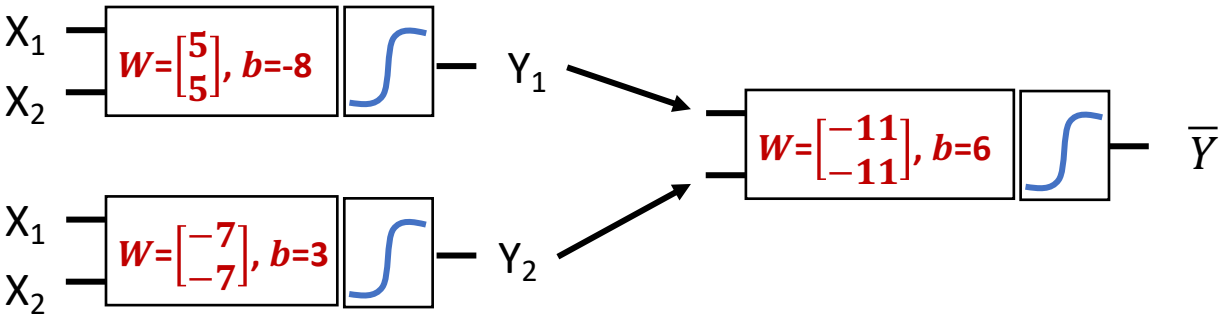
- Multiple logistic regression units can separate XOR
- But! "No one on earth had found a viable way to train"



*Marvin Minsky

Solution: MLP

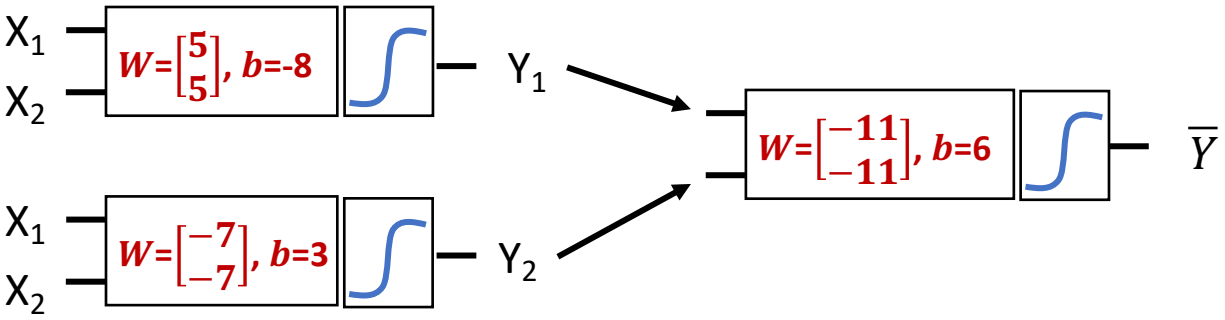
- XOR using NNs
  - toy-examples



X_1	X_2	Y_1	Y_2	$\overline{Y}$	XOR
0	0				0
0	1				1
1	0				1
1	1				0

Solution: MLP

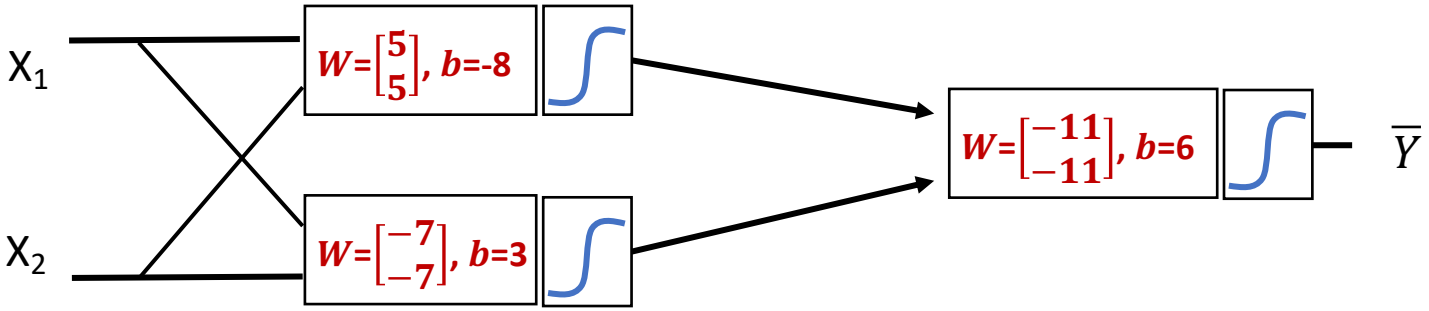
- XOR using NNs



X_1	X_2	Y_1	Y_2	$\overline{Y}$	XOR
0	0	0	1	0	0
0	1	0	0	1	1
1	0	0	0	1	1
1	1	1	0	0	0

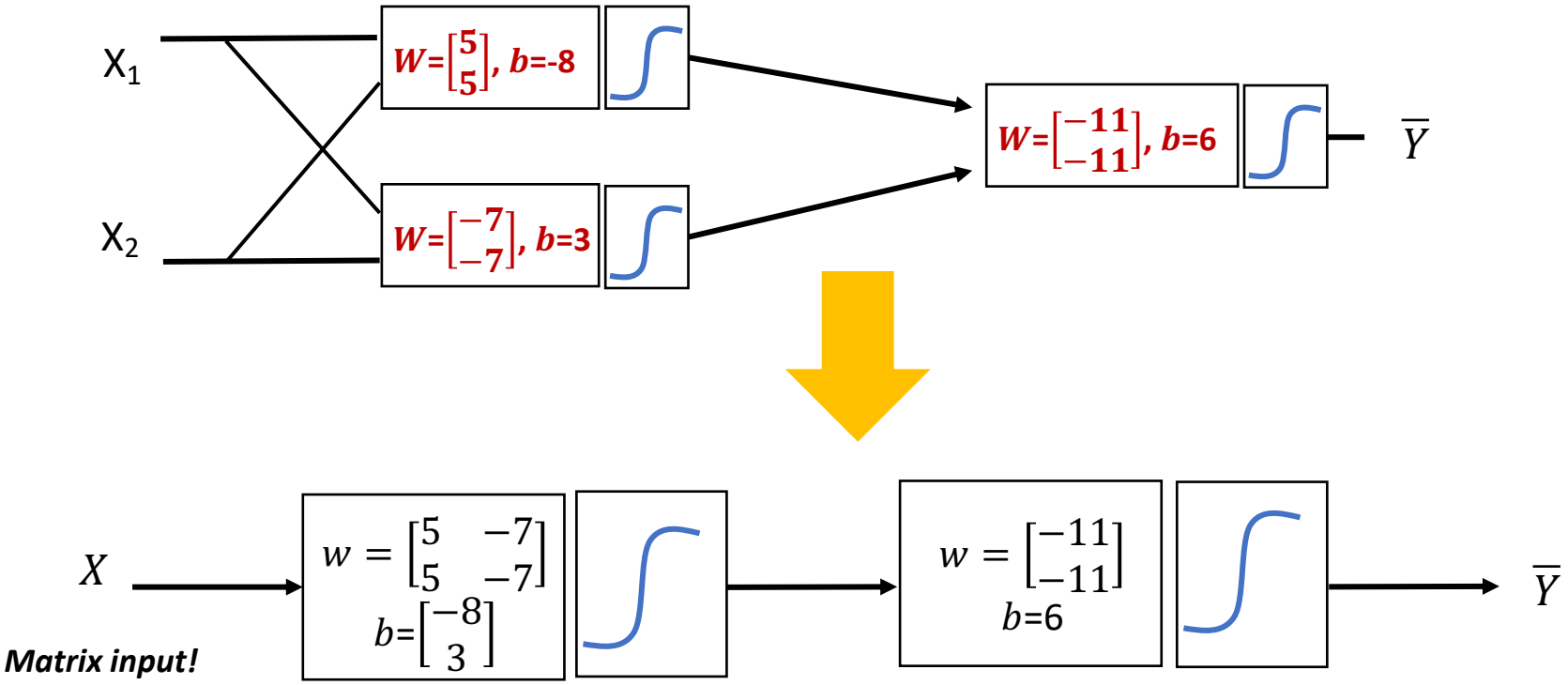
Solution: MLP

- Forward Propagation을 통한 XOR 문제 풀이 가능 검증 완료

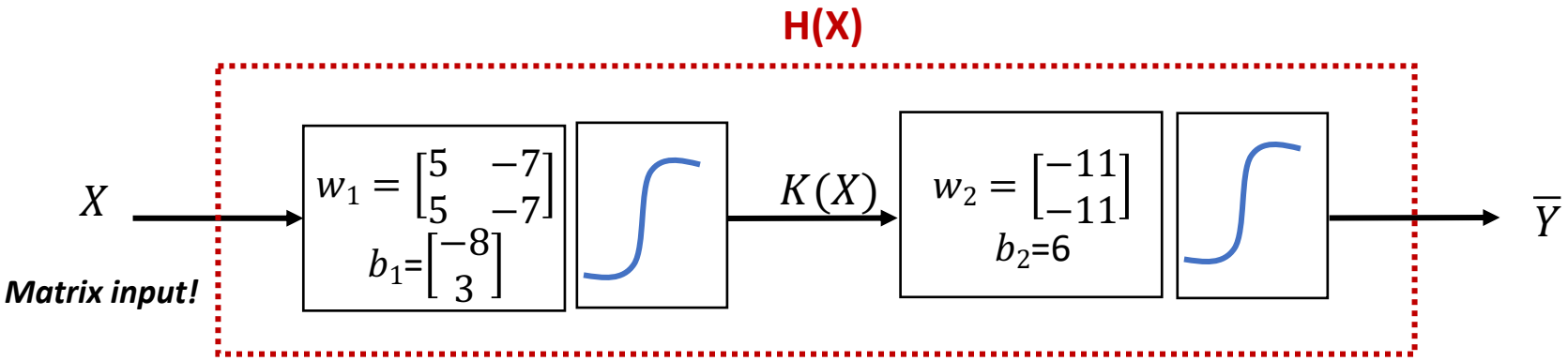


Solution: MLP

- Forward Propagation을 통한 XOR 문제 풀이 가능 검증 완료



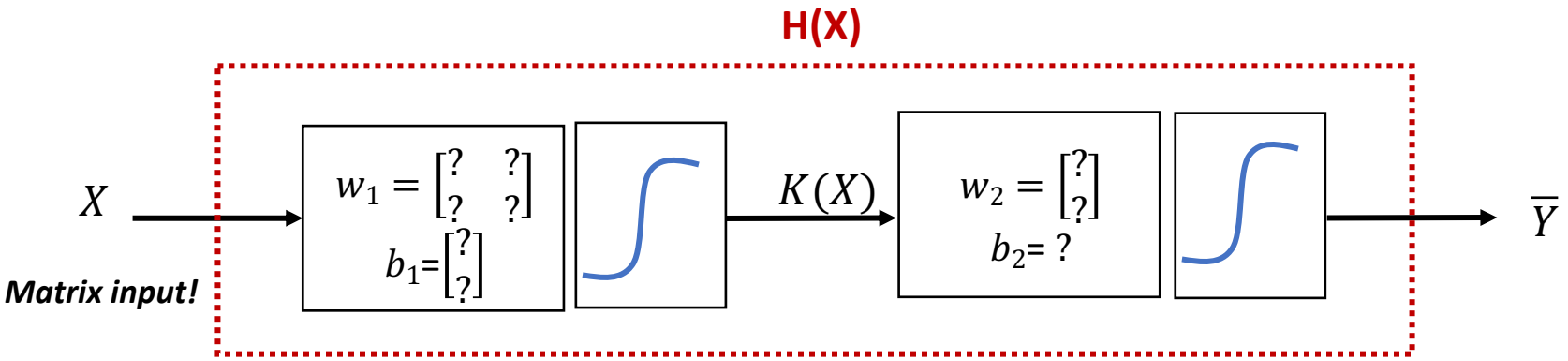
Solution: MLP for XOR



$K(X) = \text{sigmoid}(X \cdot W_1 + b_1)$

$\overline{Y} = H(X) = \text{sigmoid}(K(x) \cdot w_2 + b_2)$

How can we learn W1, W2, B1, B2 from training data?



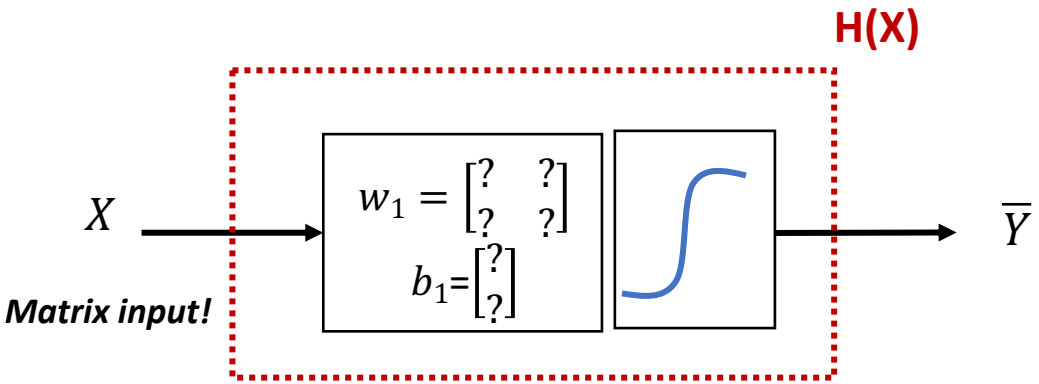
$K(X) = \text{sigmoid}(X \cdot W_1 + b_1)$

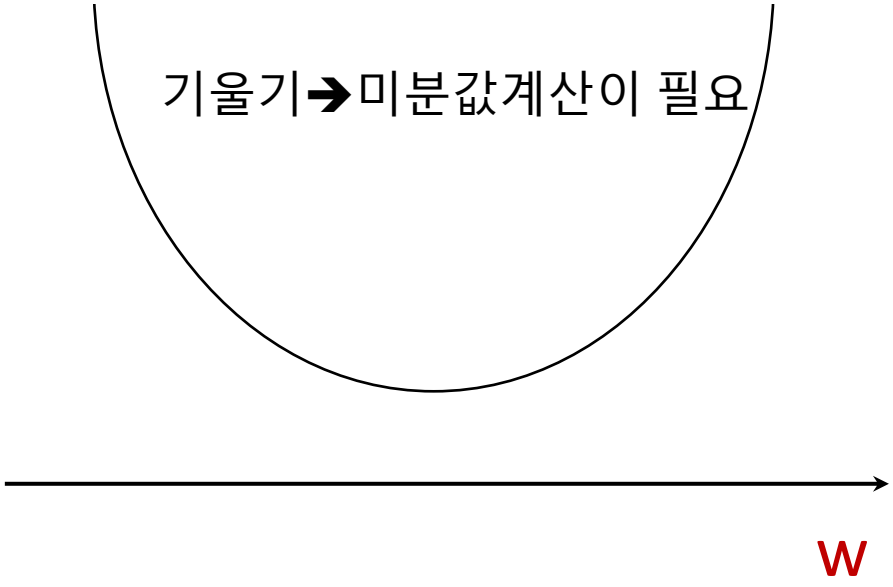
$\overline{Y} = H(X) = \text{sigmoid}(K(x) \cdot w_2 + b_2)$

X_1	X_2	$\overline{Y}$
0	0	0
0	1	1
1	0	1
1	1	0

Remind

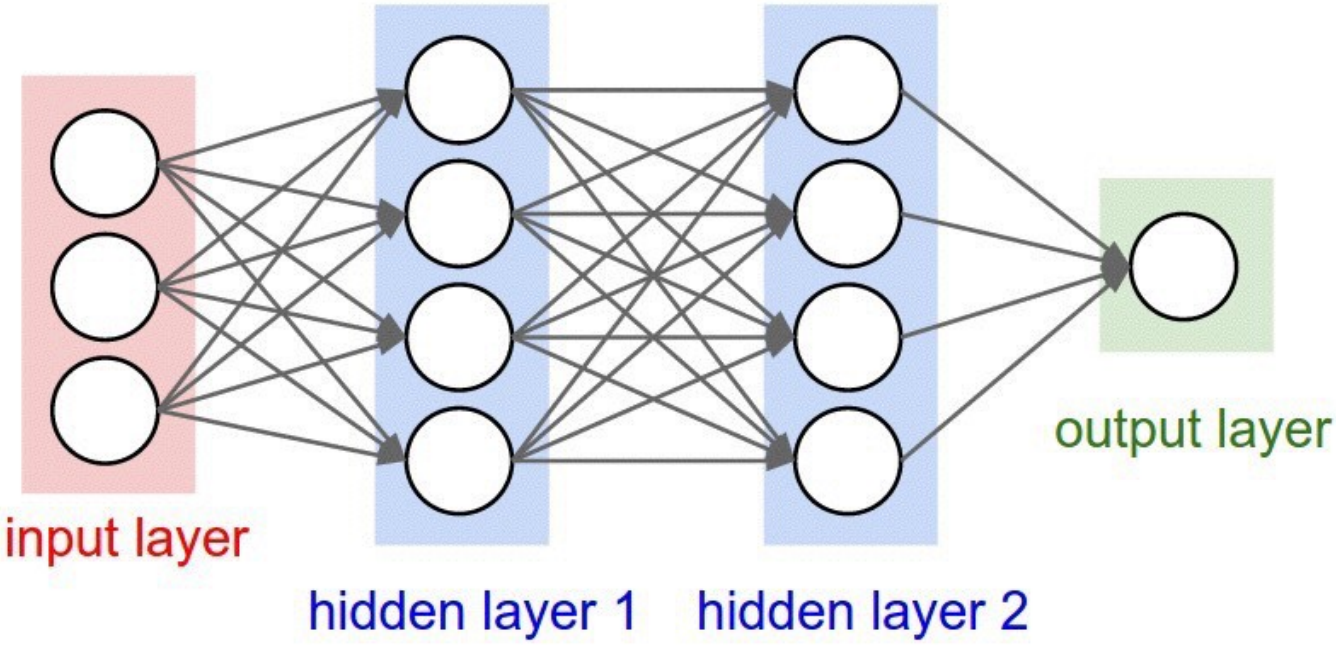
- We try to find a minimum value using a gradient descent method.





Remind

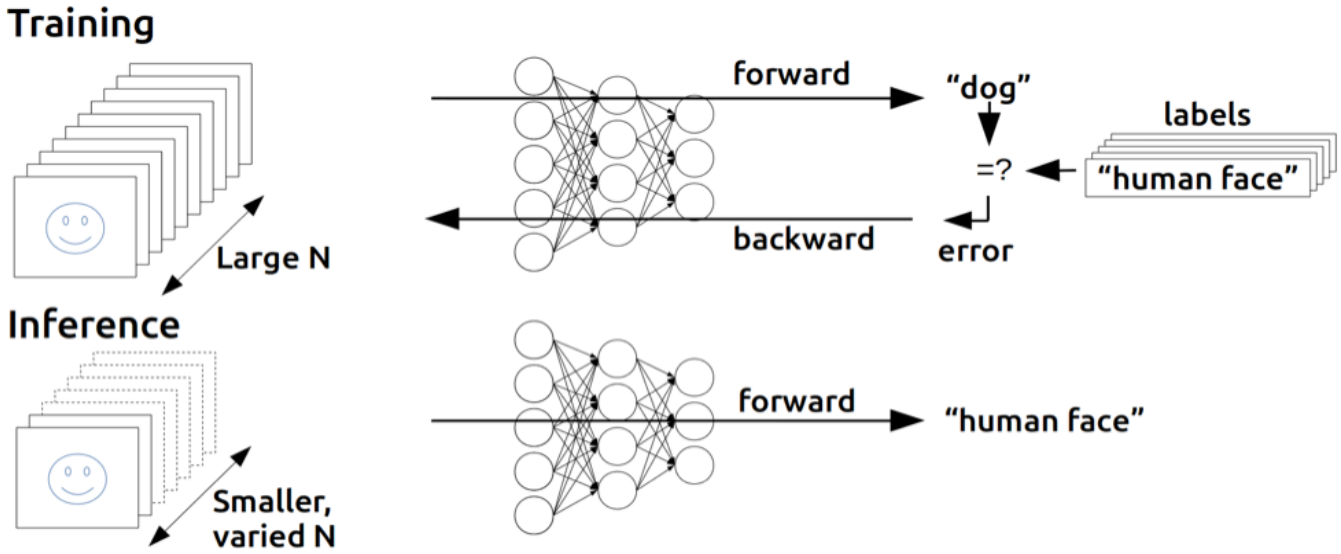
- But, **MLP** is too complex to apply gradient descent method for finding optimal parameters.



입력과 출력에 영향을 미치는 각각의 W, b 값을 구하는게 어렵다.

Remind

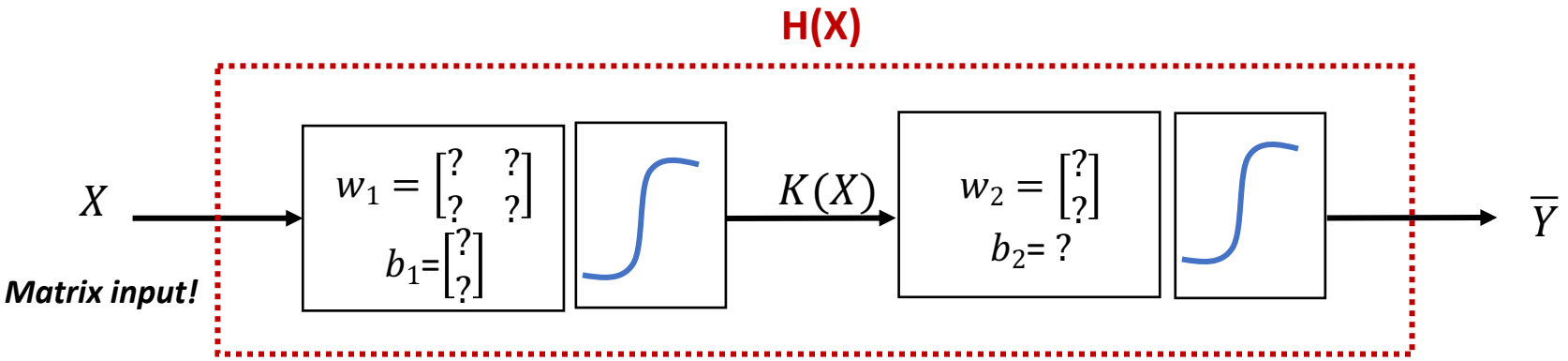
- **Backpropagation** (1974, 1982 by Paul Werbos, 1986 by Hinton)



1) W, b 초기화
2) Forward 계산 & Error 측정
3) **Error 값 을 Backward 하면서 w, b 업데이트**

[실습] MLP for XOR

How can we learn W1, W2, B1, B2 from training data?



$K(X) = \text{sigmoid}(X \cdot W_1 + b_1)$

$\overline{Y} = H(X) = \text{sigmoid}(K(x) \cdot w_2 + b_2)$

X_1	X_2	$\overline{Y}$
0	0	0
0	1	1
1	0	1
1	1	0

END