

M4. 인공신경망 기본

인공신경망 기본개념 및 MNIST 과제소개

- 머신러닝/딥러닝의 차이
- FNN
- MNIST dataset
- 경사하강법
- Backpropagation
- Hyperparameter

딥러닝 개요

강아지와 고양이를 분류하는 모델을 만들자



VS.



전통적
ML

인간의 지식을 기계에게 주입하는 방식
(온톨로지 방식, 룰베이스 방식 등)

딤러닝 개요

엄청난 예외들...

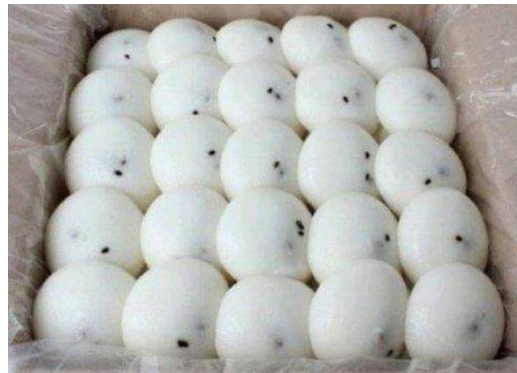


왜 기계는 사람이 정해진 규칙대로 강아지와 고양이를 구별하는 것에 어려움을 겪는가?

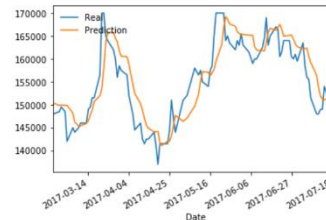
모라벡의 역설

사람에게 쉬운 것이 기계에게는 어렵다.

기계에게 쉬운 것은 사람에게 어렵다.



비숃프리제 강아지와 기정떡(증편)



$$\begin{aligned} F'(x) &= \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \left[\int_a^{x+h} f(t) dt - \int_a^x f(t) dt \right] \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \int_x^{x+h} f(t) dt \end{aligned}$$

영어 원문

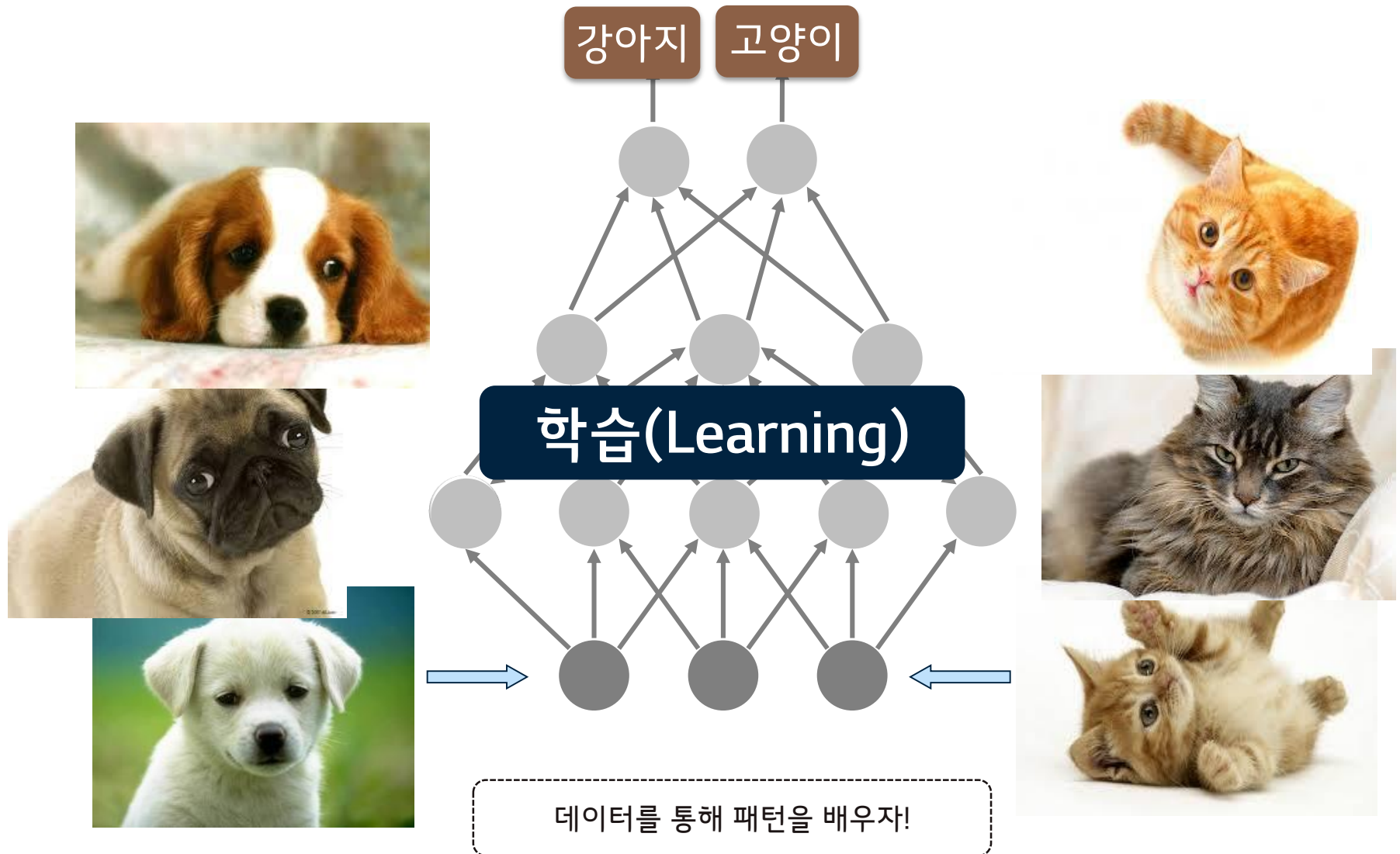
For samples with very high host nucleic acid contents (e.g., for certain tissues, such as spleen or blood samples with highly increased cell counts), use less than the maximum amount of sample recommended in the protocol or pretreatments.

자동 번역

매우 높은 숙주 핵산 함량을 갖는 샘플 (예를 들어 비장 또는 세포 수가 증가 된 혈액 샘플과 같은 특정 조직의 경우)은 프로토콜 또는 전처리에서 권장되는 최대 샘플 양을 사용하십시오.

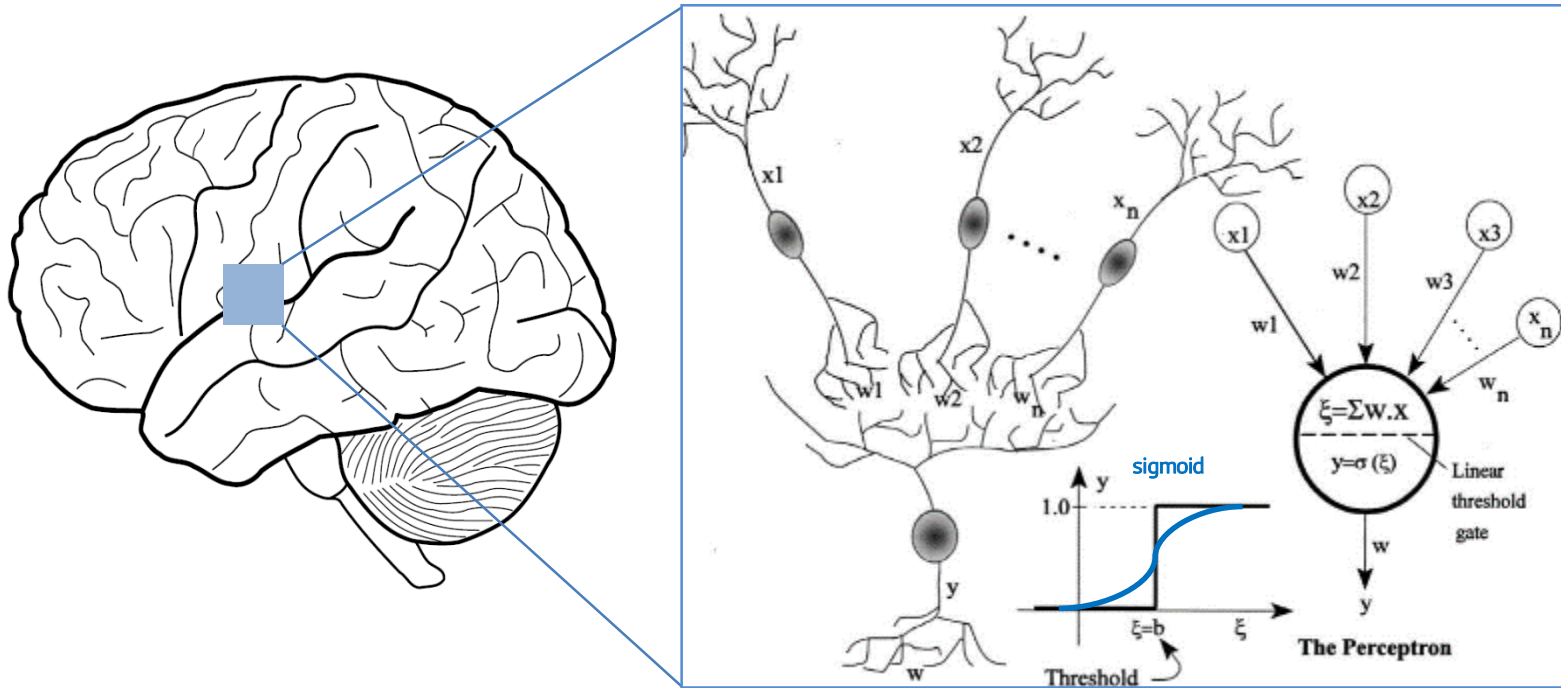
딥러닝 개요

사람은 어떻게 학습했을까?



인공신경망 개요

두뇌 세포를 모사한 인공신경망(ANN; Artificial Neural Network) 알고리즘



Neural
Network

인간이 생각하는 방식을 기계에게 주입

딥러닝이 최근 주목 받는 이유

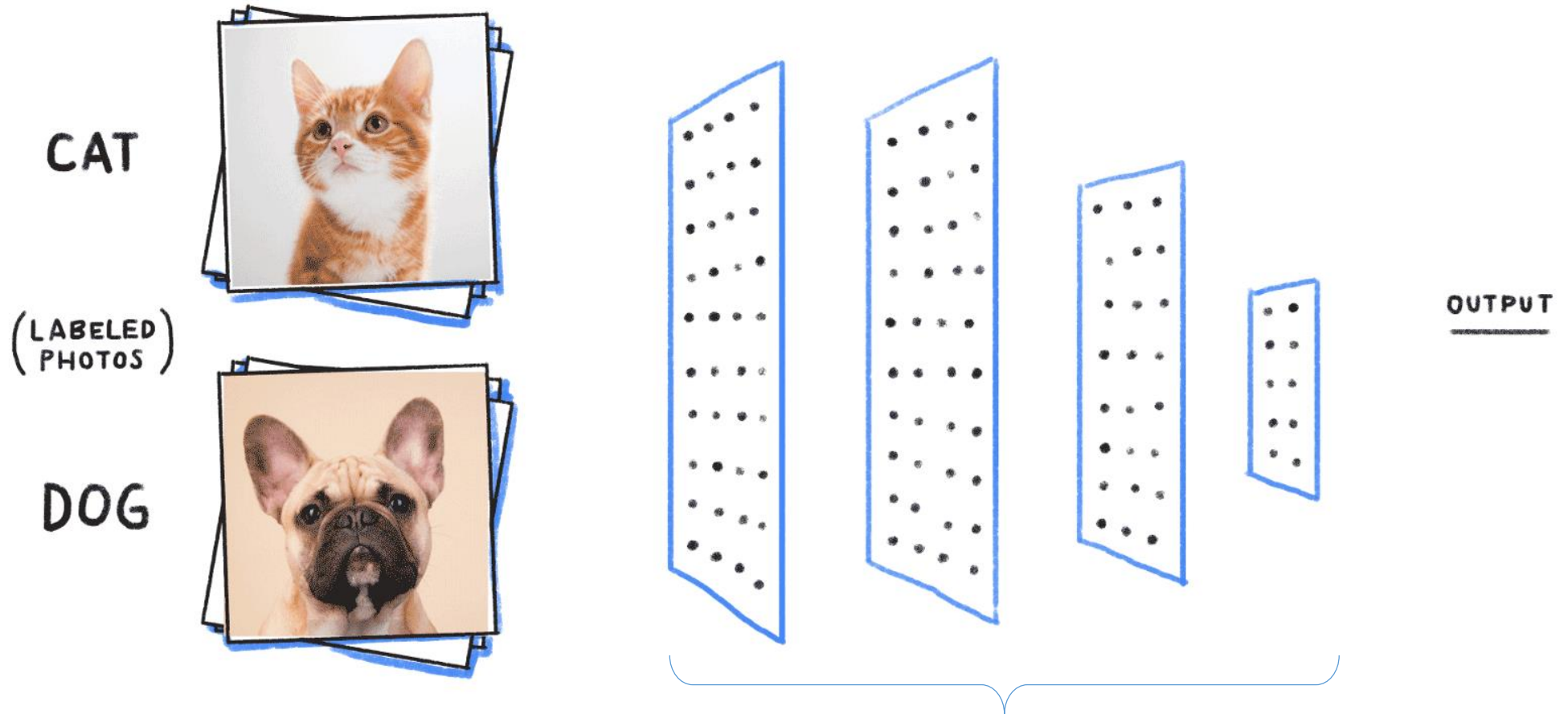
- ✓ 많은 데이터를 학습할 수록 정확도 개선
- ✓ 고성능 시스템 및 하드웨어로 더 많이, 더 빨리 학습 가능



즉, 데이터는 많아지고 하드웨어는 발전하고 있기 때문

인공신경망 개요

두뇌 세포를 모사한 인공신경망(ANN; Artificial Neural Network) 알고리즘

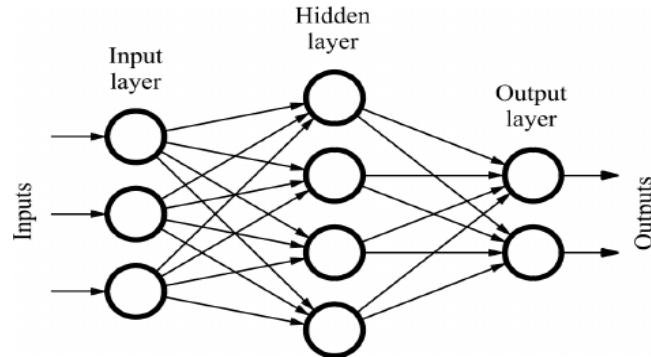


비정형 데이터를 처리할 수 있는 Neural Network

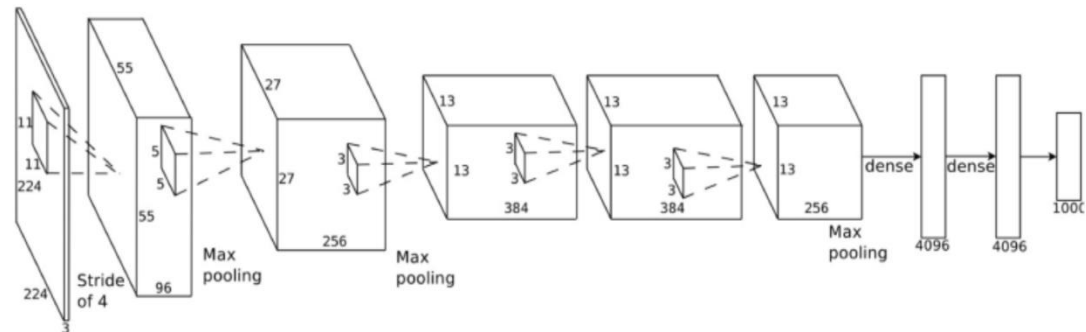
구조에 따른 인공신경망 분류

인공 뉴런 연결 방식에 따른 딥러닝 모델 기본 아키텍처 (FNN / CNN / RNN)

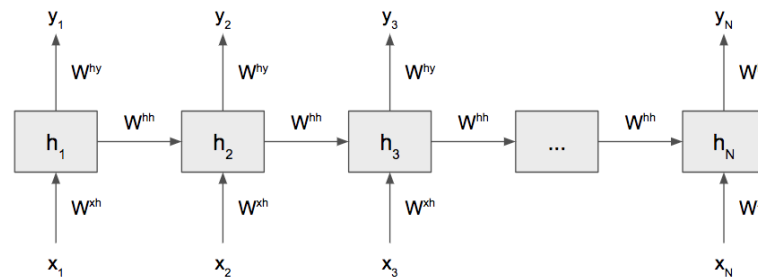
Feedforward
Neural Network



Convolutional
Neural Network



Recurrent
Neural Network



LG CNS에 입사할 것인가?

성장 가능성이
있는가?

팀원 중에
이상한 사람이 있나?

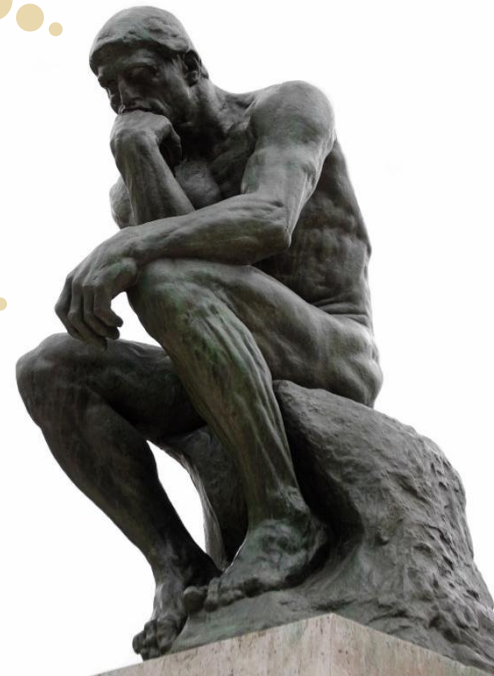
출퇴근이 편한가?

연봉은 적절한가?

조직문화가
수평적인가?

내가 잘 할 수 있는
업무인가?

밥은 맛있나?



LG CNS에 입사할 것인가?

30

-20

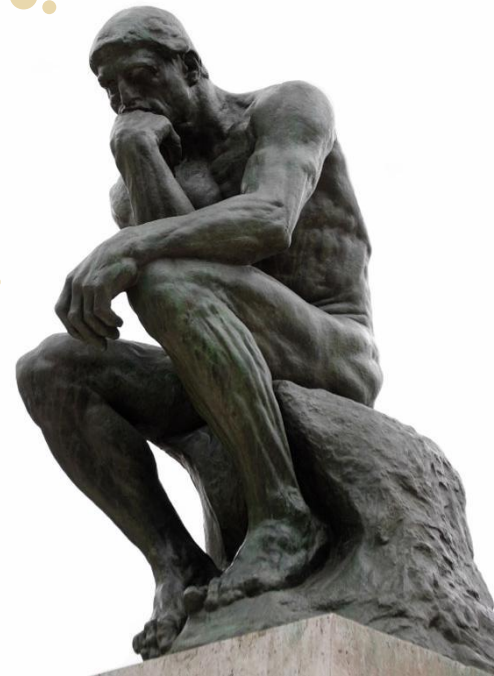
15

10

15

10

100



LG CNS에 입사할 것인가?

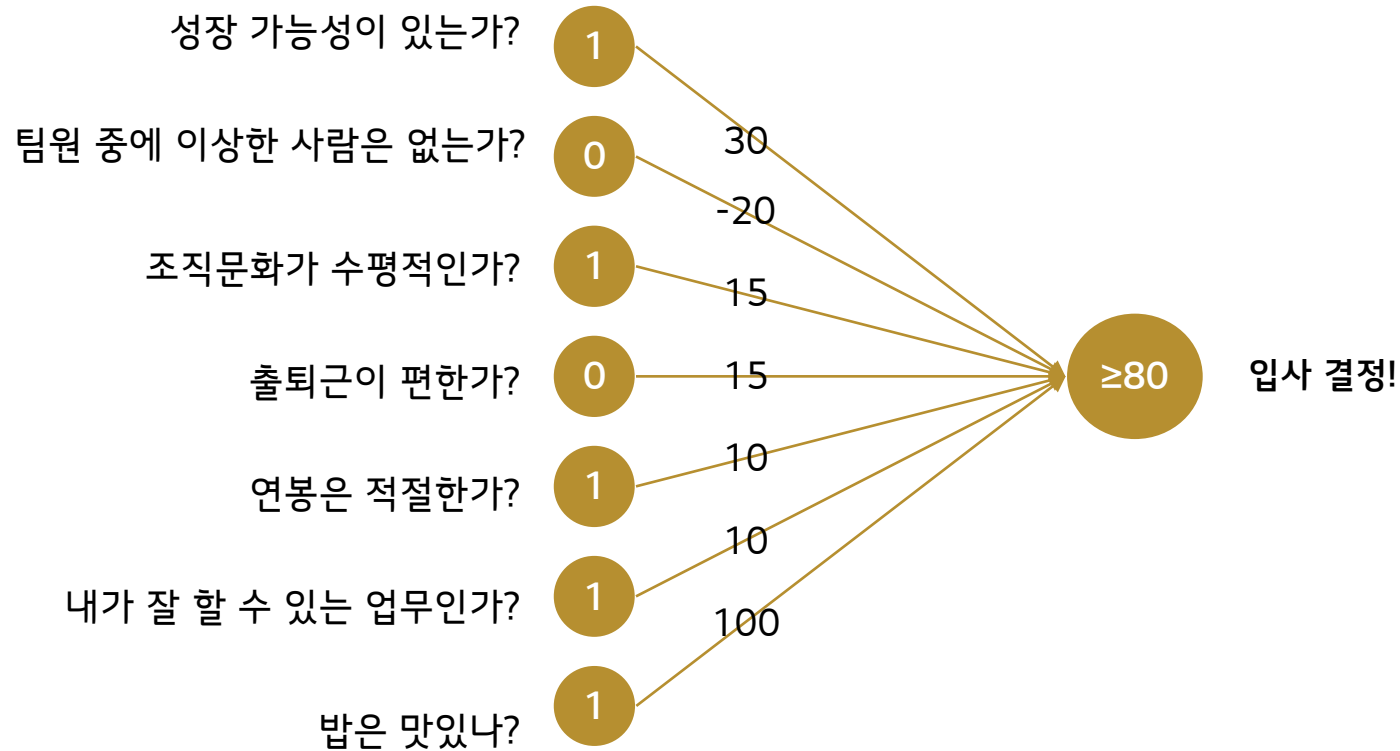
기준 : **80점** 이상 만족 시 입사 결정!

조건	Weight (w)	Data (x)	Weighted Sum ($\sum wx$)
성장 가능성이 있는가?	30	O	$30*1$
팀원 중에 이상한 사람은 없는가?	-20	X	$-20*0$
조직문화가 수평적인가?	15	O	$15*1$
출퇴근이 편한가?	15	X	$15*0$
연봉은 적절한가?	10	O	$10*1$
내가 잘 할 수 있는 업무인가?	10	O	$10*1$
밥은 맛있나?	100	O	$100*1$

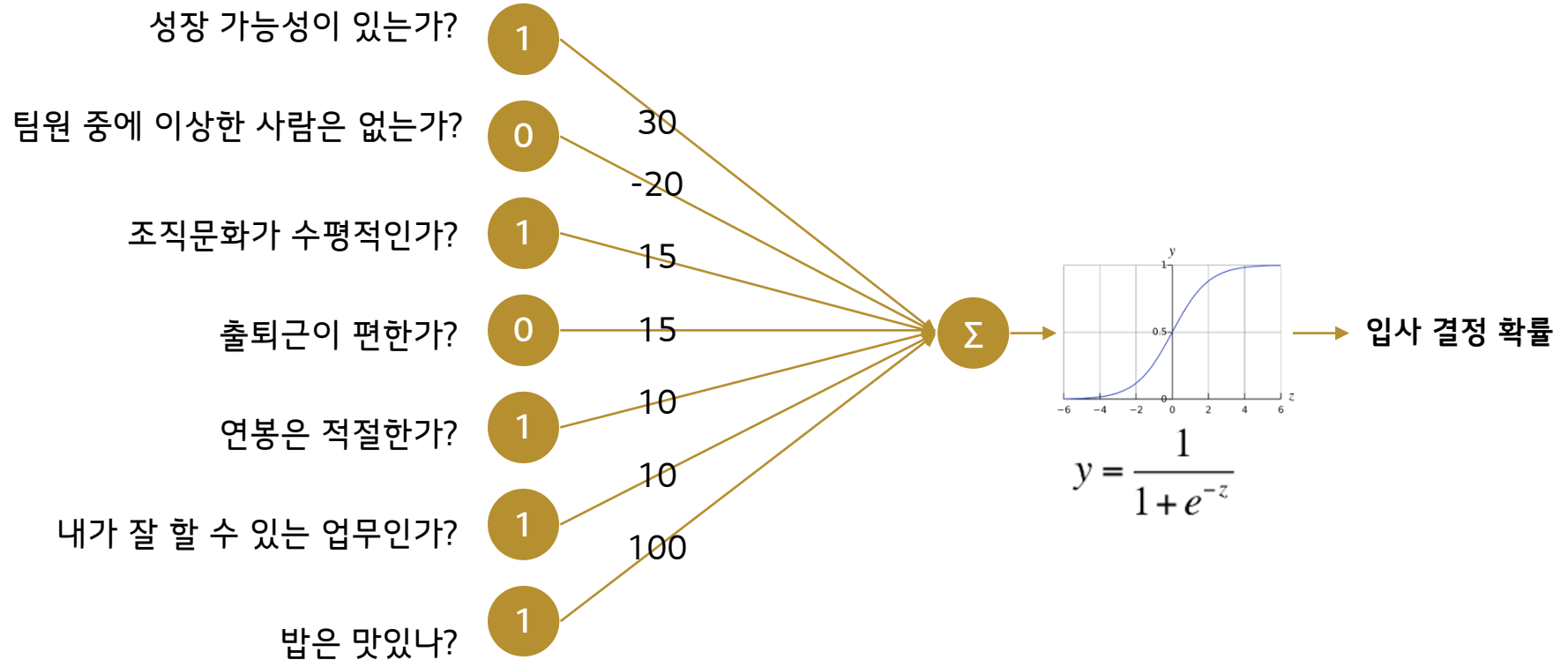
$= 165 \geq 80$

LG CNS 입사 결정!

LG CNS에 입사할 것인가?



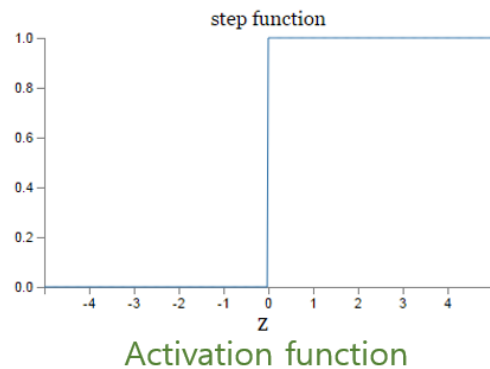
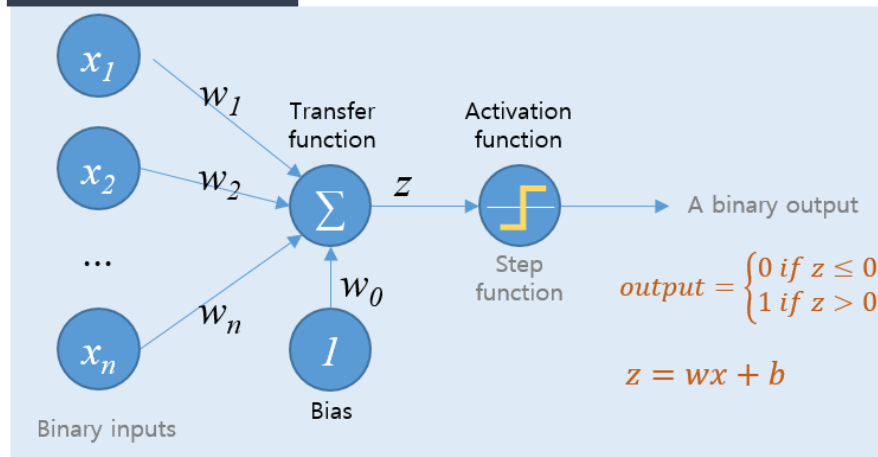
LG CNS에 입사할 것인가?



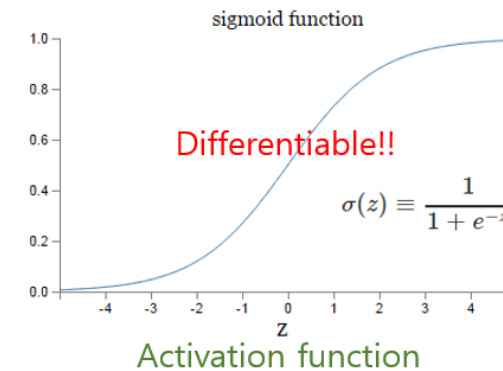
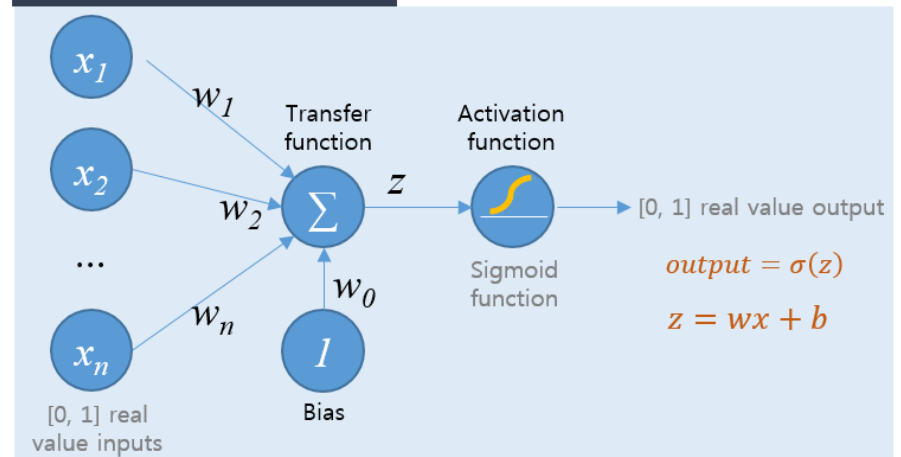
Feedforward Neural Network

시그모이드 뉴런 (Sigmoid Neuron)

A perceptron



A sigmoid neuron

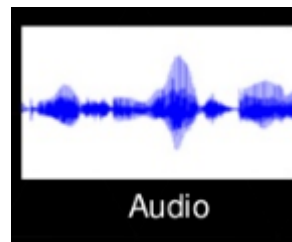


DEEP Feedforward Neural Network

딥러닝이 다루야 하는 데이터의 차원 또한 크고 복잡해진다

공부시간(x)	시험성적(y)
10	75
14	77
17	86
20	84
22	92
25	90

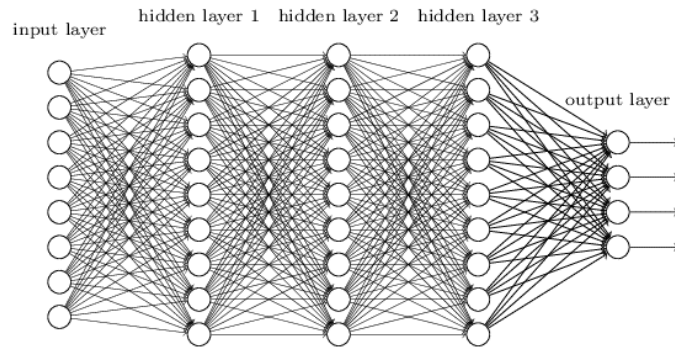
머신러닝이 다루는 정형 데이터
몇~몇십 차원



딥러닝이 다루는 비정형 데이터
수백~수십만 차원

DEEP Feedforward Neural Network

수많은 비정형 데이터를 처리하기 위해 더 복잡하고 깊은 인공신경망 기반의 네트워크 구조



일반적으로 2층 이상의 인공신경망 연산을 태우는 경우를 Deep learning 모델이라 함



구체적인 feature ▶▶▶

- Edge
- Color
- ...

▶▶▶

- 뾰족한 귀
- 둥그런 코
- ...

▶▶▶

추상화된 feature

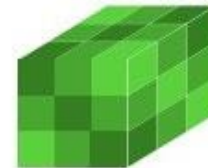
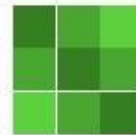
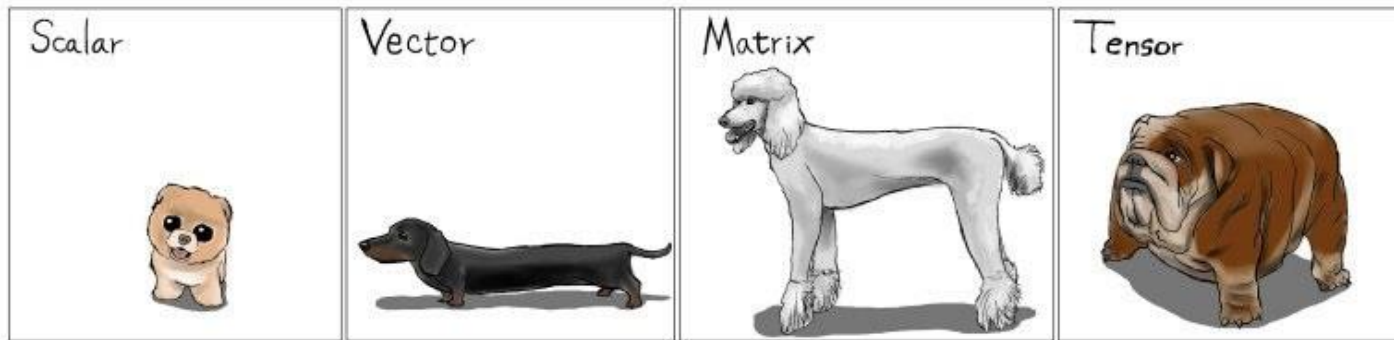
- 강아지의 특징이 보이는가?
- 고양이의 특징이 보이는가?

Deep learning is a class of machine learning algorithms that :

- use a cascade of multiple layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input.
- learn multiple levels of representations that correspond to different levels of abstraction; the levels form a hierarchy of concepts.

잠깐! 선형대수 기본상식

Tensor는 scalar, vector, matrix를 포함한 모든 dimension을 표현할 수 있다



1

$$\begin{pmatrix} 2.3 \\ 4.0 \end{pmatrix}$$

$$\begin{pmatrix} (0.1 & -2.0) \\ (5.8 & 10.2) \end{pmatrix}$$

$$\begin{pmatrix} [(1 \ 2) (3 \ 2)] \\ [(7 \ 1) (5 \ 4)] \end{pmatrix}$$

잠깐! 선형대수 기본상식

Shape : 각 차원의 element 수

1

$$\begin{bmatrix} 2.3 \\ 4.0 \end{bmatrix}$$

Shape []

$$\begin{bmatrix} 0.1 & -2.0 \\ 5.8 & 10.2 \end{bmatrix}$$

Shape [2]

$$\begin{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 3 & 2 \end{bmatrix} \\ \begin{bmatrix} 7 & 1 \end{bmatrix} \begin{bmatrix} 5 & 4 \end{bmatrix} \end{bmatrix}$$

Shape [2,2]

Shape [2,2,2]

[illegible]

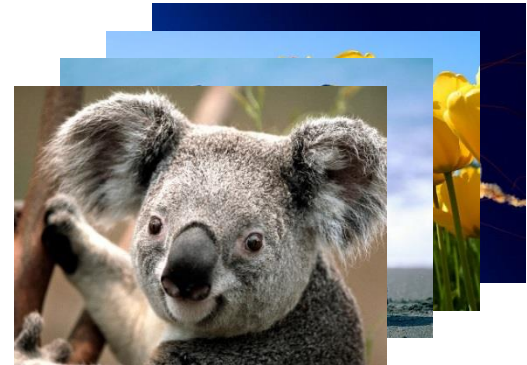
Shape [28,28]

150 px

100 px



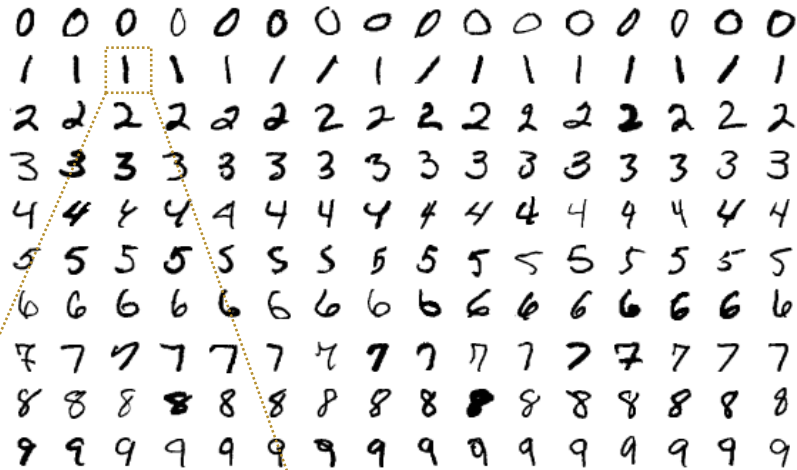
Shape [100,150,3]



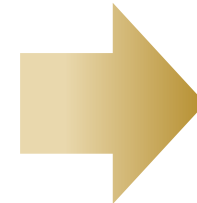
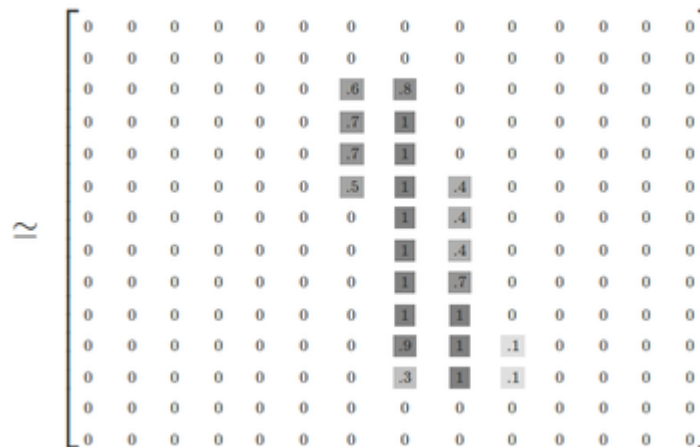
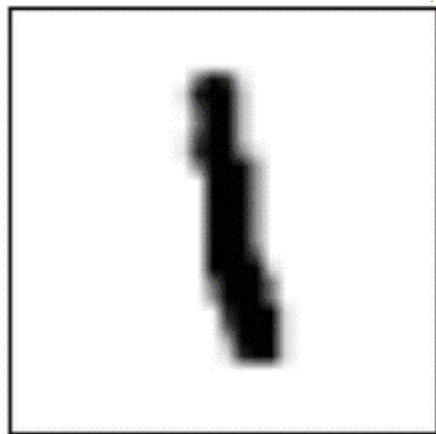
Shape [4,100,150,3]

필기체 숫자(MNIST*) 인식

MNIST dataset



- ✓ 숫자의 각 이미지는 28*28 흑백 이미지 (784픽셀)
- ✓ 정답 라벨(Ground Truth)은 길이 10의 열벡터



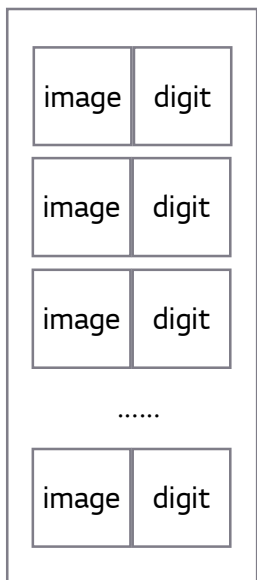
[0.]
[1.]
[0.]
[0.]
[0.]
[0.]
[0.]
[0.]
[0.]
[0.]

* Modified National Institute of Standards and Technology 데이터셋으로 필기체 숫자 이미지 모음이며 60,000개의 트레이닝 이미지와 10,000개의 테스트 이미지를 포함

필기체 숫자(MNIST) 인식

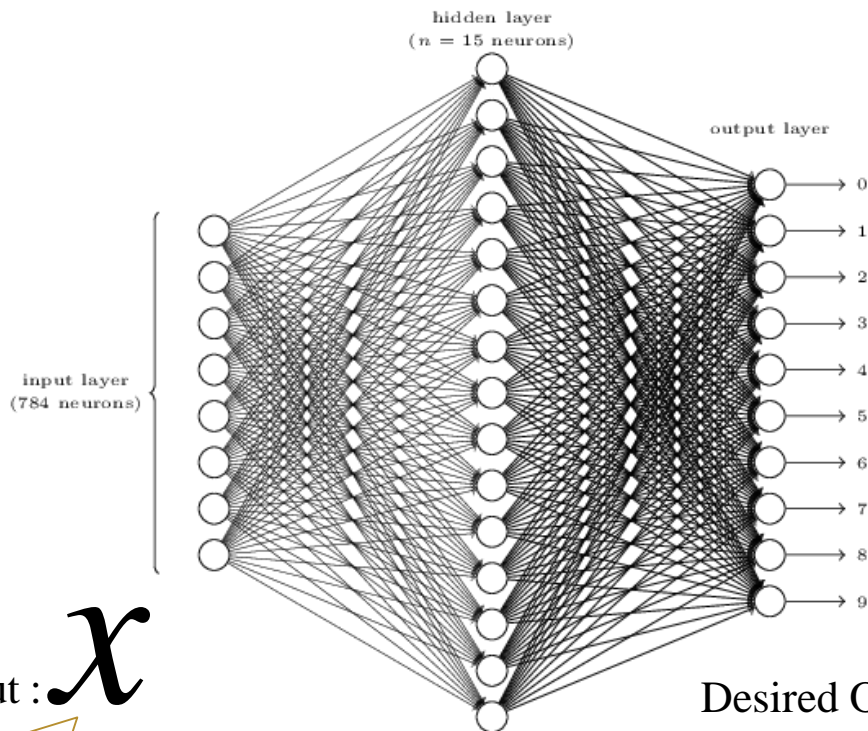
인공신경망 학습시키기 = weights와 biases 찾기

Dataset for training



Input : x

회색조 이미지에 대한 벡터로,
각 원소는 0~1사이의 값을 가짐
(255로 normalized)



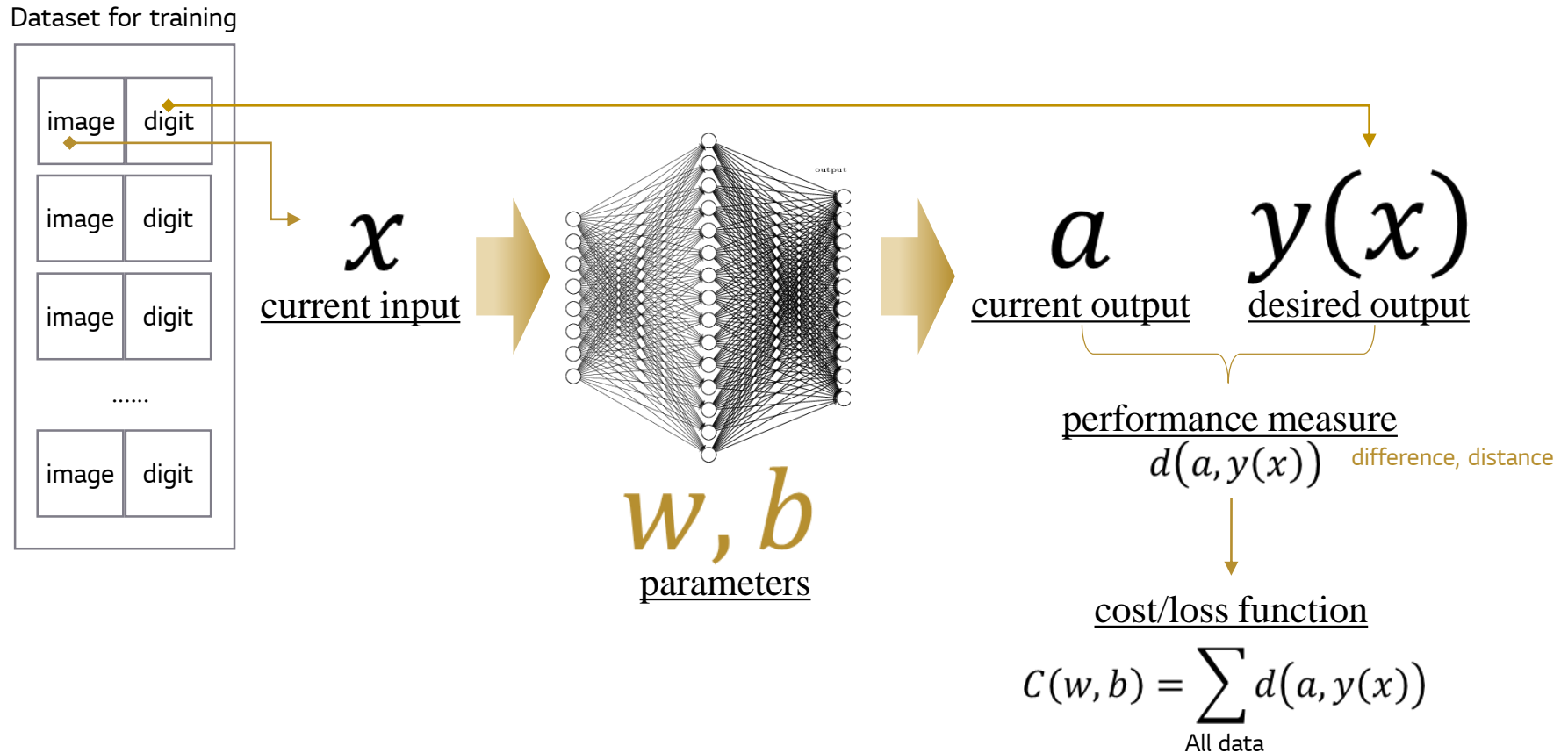
Desired Output : $y(x)$

w : all weights, b : all biases

10 dimension의 벡터로,
만일 정답이 '6'이라면
 $y(x) = (0,0,0,0,0,0,1,0,0,0)^T$

필기체 숫자(MNIST) 인식

인공신경망 학습시키기 = weights와 biases 찾기



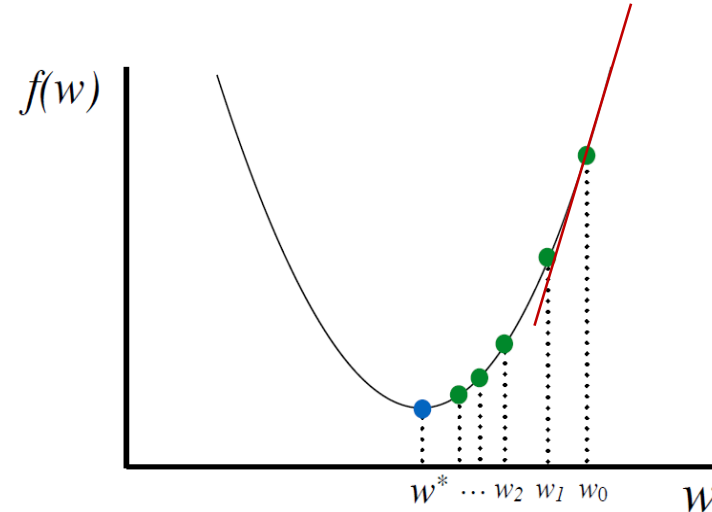
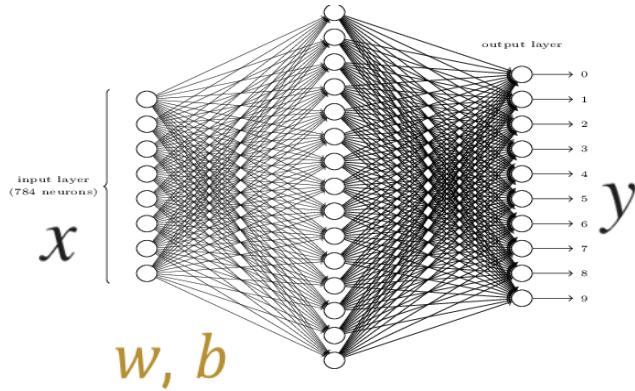
$d(a, y(x))$ 가 'a'와 'y(x)'의 차이를 나타내는 non-negative value라면,
학습의 목적은 $C(w, b) = 0$ or $\text{minimize } C(w, b)$ 이 되도록 하는 w, b 찾기이다.

경사하강법 : Gradient Descent

Learning이란 어떻게 일어나는가?

목적

모델의 예측값과 실제 정답값의 차이를 최소화하는 파라미터(w, b) 구하기



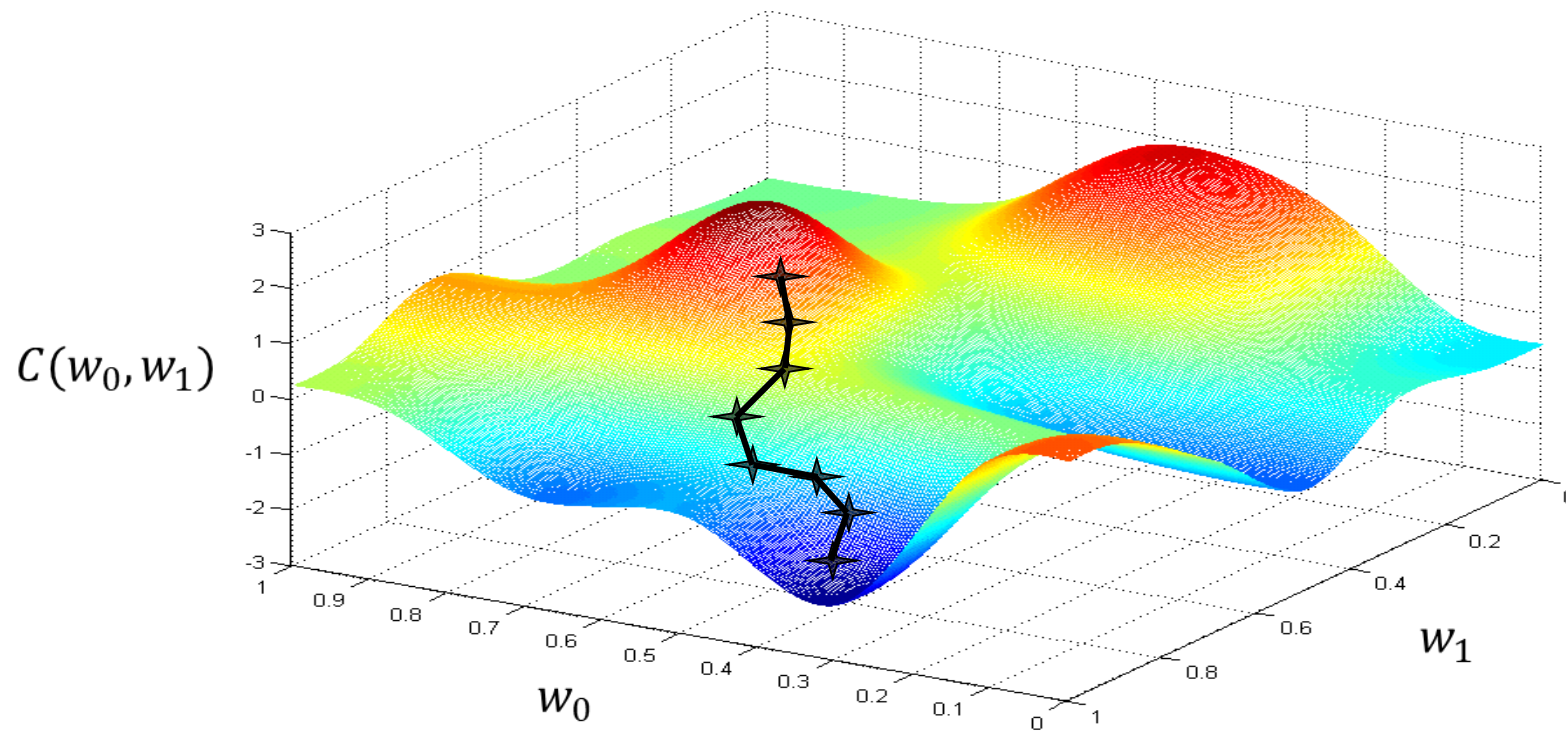
input data	x
Parameter	w, b
모델의 예측값	a
실제 정답값	y
Cost Function (Quadratic Cost function)	$C(w, b) = \frac{1}{2n} \sum_{i=1}^n (a_i - y_i)^2$



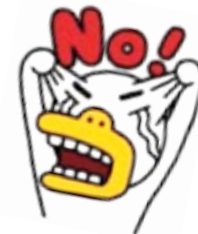
$$w_{j+1} \leftarrow w_j - \alpha \frac{\partial C(w)}{\partial w_j}$$

한 번에 최적의 parameter를 찾는 게 아니라
여러 번 좋은 방향으로 update 하며 '보정'해나가는 개념!

경사하강법 : Gradient Descent



하지만 이런 parameter들이 단순히 weight 한두어개가 아니라 어어엄청(보통 몇백만~몇십억) 많이 있을 것이기 때문에.... 시각화는 물론 머릿속에서 상상하는것도 몹시 어렵습니다.
우리는 parameter 수 만큼의 다차원 실수공간 기울기 계산(미분...!)을 통해 한번에 고려해야 합니다.

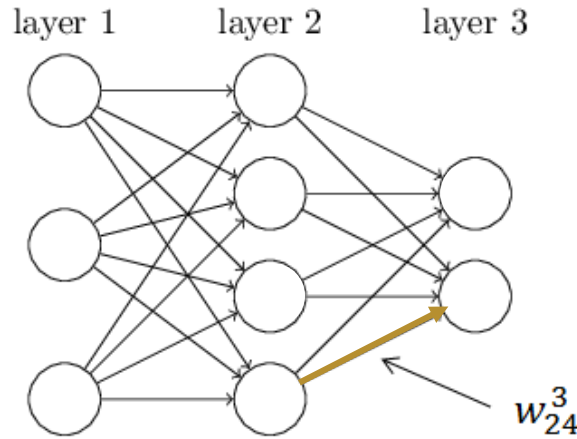


오류역전파 : Backpropagation

무수한 파라미터들.... Cost에 대한 기울기(Gradient)를 언제 다 구하지!?

특정 weight가
Cost에 미치는 영향은?

$$\frac{\partial C(w)}{\partial w_{jk}^l}$$



다음과 같이
Error δ 를 정의하자

$$\delta_j^l = \frac{\partial C}{\partial z_j^l}$$

l 번째 레이어의 j 번째 노드 output이
Cost에 미치는 영향

Gradient를 쉽고 빠르게 구하도록 도와주는 Backpropagation의 4대 방정식

1. Error at the output layer	2. Error relationship between two adjacent layers	3. Gradient of C in terms of bias	4. Gradient of C in terms of weight
$\delta^L = \nabla_a C \odot \sigma'(z^L)$	$\delta^l = \sigma'(z^l) \odot ((w^{l+1})^T \delta^{l+1})$	$\nabla_{b^l} C = \delta^l$	$\nabla_{w^l} C = \delta^l (a^{l-1})^T$

신경망의 맨 마지막 레이어 L 에서의
Error δ^L 을 구할 수 있다

$(l+1)$ 번째 레이어의 Error δ^{l+1} 를 알 수
있다면, l 번째 레이어의 Error δ^l 도 구할
수 있다

l 번째 레이어의 Error δ^l 를 안다면,
 l 번째 레이어의 bias와 weight에 대한 Gradient를
구할 수 있다

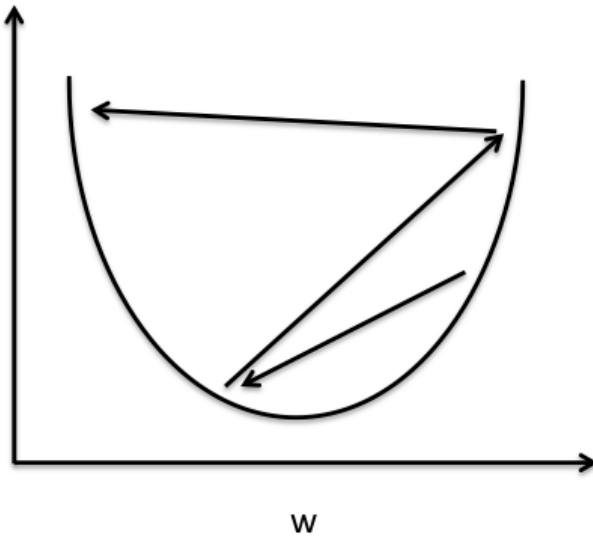


Hyperparameter : Learning rate

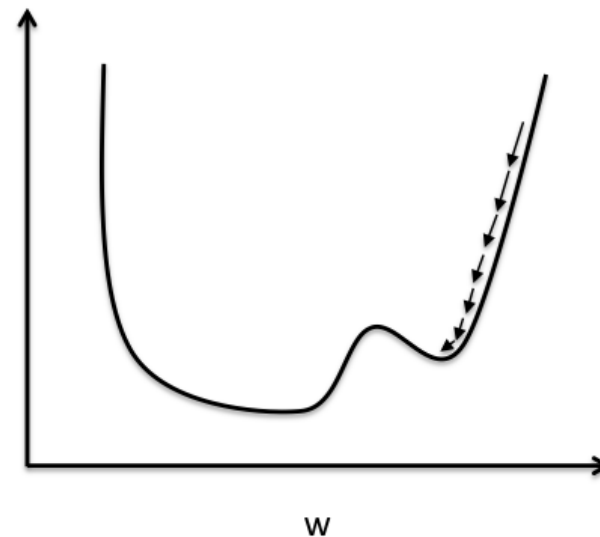
파라미터 업데이트의 정도를 조절하는 learning rate

$$w_{j+1} \leftarrow w_j - \alpha \frac{\partial C(w)}{\partial w_j}$$

Learning rate (>0)



Large learning rate: Overshooting.



Small learning rate: Many iterations until convergence and trapping in local minima.

용어정리 : Hyperparameter

학습하는 데 필요하지만, 학습은 되지 않는 하이퍼 파라미터

Hyperparameters

- 모델의 capacity(or complexity)를 결정
- Training동안 학습되지 않음 → 사람(신경망 설계자)이 설정해줘야 함
- 예 : learning rate, weight decay, hidden size, layer 수, mini-batch size, feature 수 등..

$$w_{j+1} \leftarrow w_j - \alpha \frac{\partial C(w)}{\partial w_j}$$

Parameters

- 모델에 의해 Training동안 학습됨 → 사람이 설정하는 것이 아니라, 데이터를 통해 결정됨
- 예 : 인공신경망의 weights & biases 등..

용어 정리

Epoch, Iteration, Mini-batch size

1 epoch	모든 Training data가 한 번씩 forward pass와 backward pass를 진행
1 iteration	한 번의 forward pass와 backward pass
Mini-batch size	한 iteration을 진행할 Training data 예제의 수

예제

Training data 50,000건이 있고 mini-batch size가 1,000이라면
1 epoch을 수행하는 데 50 iteration이 진행된다.