



# 5. 심층 신경망 네트워크

2024.04.09  
임채영

# 목차

- #01 더 많은 층의 심층 신경망
- #02 심층 신경망에서의 정방향전파
- #03 행렬의 차원을 알맞게 만들기
- #04 왜 심층신경망이 더 많은 특징을 잡아낼 수 있을까?
- #05 심층신경망 네트워크 구성하기
- #06 장방향전파와 역방향전파
- #07 변수 vs 하이퍼파라미터
- #08 인간의 뇌와 어떤 연관이 있을까?

퀴즈1

퀴즈2



# #01 더 많은 층의 심층 신경망

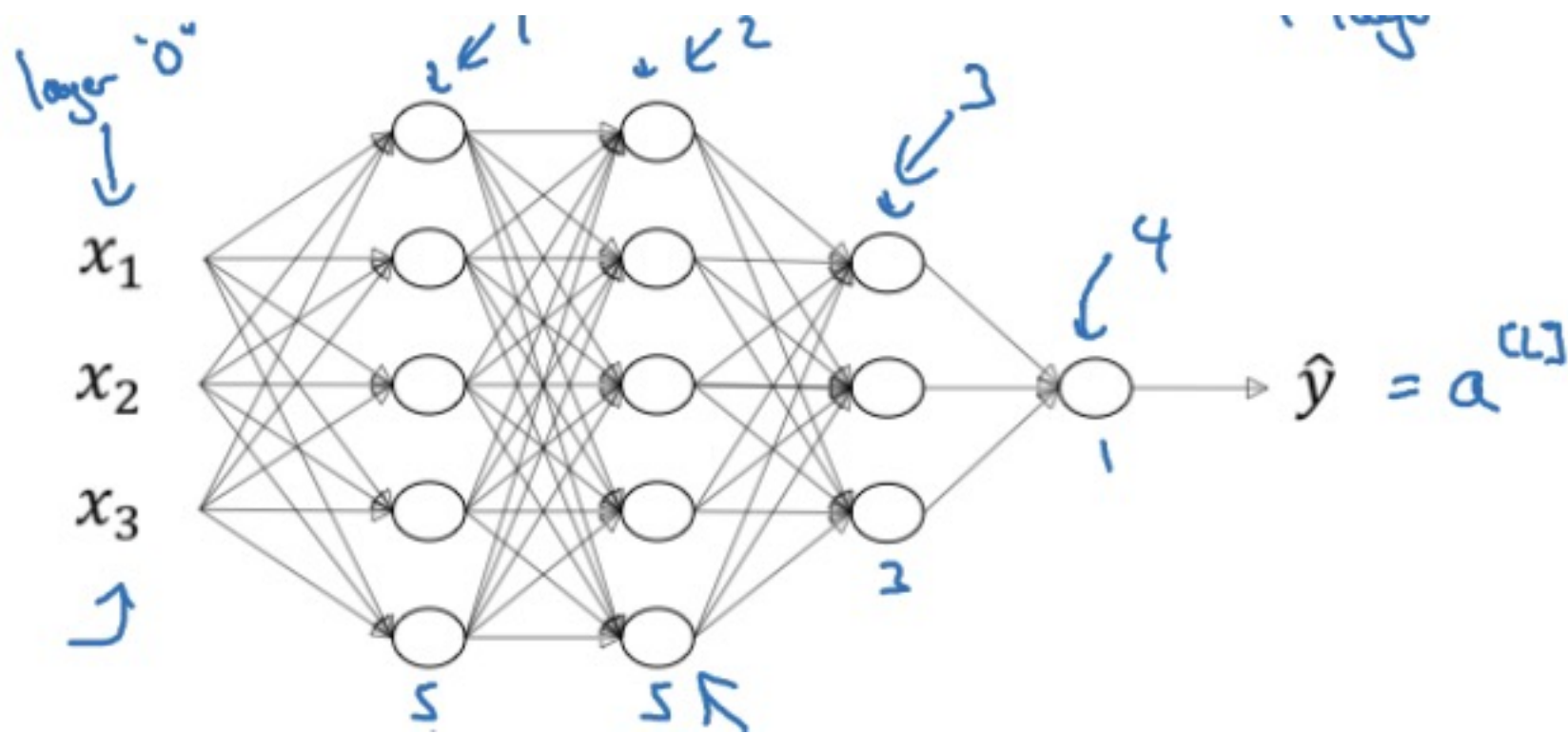
## 📌 심층신경망이란?

로지스틱 회귀 : 매우 얇은 모델, 한 층의 신경망

심층 신경망 : 더 깊은 모델, 여러 개의 은닉층+출력층을 가진 신경망

## 📌 심층 신경망의 초기화

심층 신경망은 많은 수의 파라미터를 가지고 있기 때문에 올바른 초기화가 매우 중요  
초기화란 각 뉴런의 가중치와 편향을 어떤 값으로 설정하는지를 말함



- $L$  : 네트워크 층의 개수
  - $L=4$
- $n^{[l]}$  :  $l$ 층에 있는 유닛 개수
  - $n^{[1]}=5, n^{[2]}=5, n^{[3]}=3, n^{[4]}=n^{[L]}=1$
  - $n^{[0]}=n_{\{x\}}=3$
- $a^{[l]}$  :  $l$ 층에서의 활성화값
- $w^{[l]}$  :  $z^{[l]}$ 의 값을 계산하기 위한 가중치
- $b^{[l]}$
- $a^{[0]}=x$  : 입력 특징
  - $g^{[l]}(z^{[l]})$
- $a^{[L]}=\hat{y}$  : 예측된 출력값

# #01 더 많은 층의 심층 신경망

📌  $w$ 의 값을 모두 0으로 초기화할 경우  
두 은닉 유닛이 같은 함수를 계산하는 것으로 시작하기 때문에  
두 은닉 유닛이 출력 유닛에 항상 같은 영향을 주게 되고,  
첫번째 반복 이후에 같은 상태가 계속해서 반복됨 → 은닉 유닛이 실제로는 1개라고 할 수 있음

📌 해결방법: 변수를 랜덤으로 초기화  
 $w^1 = \text{np.random.randn()} * 0.01$ 으로 설정  
 $w^2$ 도 마찬가지

📌 왜 0.01을 곱하는가  
가중치의 초기값을 매우 작은 값으로 정하는게 좋음  
만약 tanh 함수나 시그모이드 활성화함수를 사용할때 가중치가 너무 큰 값을 가지는 경우 활성화값을  
계산하면 몇몇 값이 굉장히 크거나 굉장히 작은 상태가 될수있음 → 경사 하강법 기울기가 작아서 학습  
속도가 느려짐

# #02 심층 신경망에서의 정방향전파

## 단일 학습 데이터 정방향 전파

$$z[l] = w[l]a[l-1] + b[l]$$
$$a[l] = g_l(z[l])$$

## 전체 학습 세트 벡터화

$$Z[l] = w[l]A[l-1] + b[l]$$
$$A[l] = g_l(Z[l])$$

층 1부터 L까지 각각에 대한 활성화를 계산하는 반복문을 사용하는 것은 괜찮음  
(명시적 반복문 외에 구현 방법이 없음)

# #03 행렬의 차원을 알맞게 만들기

  $z[l] = W[l]x + b[l]$

차원은 각각  $(n[1], 1)(n[1], n[0])(n[0], 1)(n[1], 1)$

$W[l]: (n[l], n[l-1])$

$b[l]: (n[l], 1)$

W와 b 벡터가 올바른 차원을 가지는지 이중으로 확인  
역전파 구현 경우: W는 dW, b는 db와 같은 차원을 가짐

$dW[l]: (n[l], n[l-1])$

$db[l]: (n[l], 1)$

# #03 행렬의 차원을 알맞게 만들기

## 벡터화된 구현

$$Z[l] = W[l]X + b[l]$$

차원은 각각  $(n[1], m)$   $(n[1], n[0])$   $(n[0], m)$   $(n[1], 1)$   
 $b[l]$ 은  $(n[1], m)$ 으로 broadcasting

$$Z[l], A[l]:(n[l], m)$$
$$dZ[l], dA[l]:(n[l], m)$$

$l=0$  일 때  $A[0] = X = (n[0], m)$   
 $z[l], a[l]:(n[l], 1)$

$Z/A$ 와  $dZ, dA$ 는 같은 차원을 가짐

# #04 왜 심층신경망이 더 많은 특징을 잡아낼 수 있을까?

📌 네트워크가 더 깊어질수록, 더 많은 특징을 잡아낼 수 있음

낮은 층에서는 간단한 특징을 학습

깊은 층에서는 탐지된 간단한 것들을 함께 모아 더 복잡한 것들을 탐지

## 📌 얼굴인식

입력값: 얼굴사진

20개의 은닉층 : 작은 네모상자

첫번째 층 : 특성 탐지기, 모서리 탐지기 >> 사진을 보고 모서리가 어디에 있는지 파악

두번째 층 : 모서리를 형성하기 위해 픽셀을 그룹화 >> 감지된 모서리와 그룹화된 모서리를 받아서 얼굴의 일부 형성. 즉, 많은 모서리를 한데 모아서 얼굴의 일부 감지 가능

세번째 층: 최종적으로 서로 다른 얼굴의 일부를 모아서 서로 다른 종류의 얼굴 감지 가능 >>

모서리처럼 간단한 것을 찾고 같이 모아서 더 복잡한 것을 찾고 또 모아서 더 복잡한 것을 찾는다

## 📌 음성인식시스템

입력 : 음성

첫번째 층 : 낮은 단계의 음성 파형 특징을 탐지

두번째 층 : 소리의 기본 단위 탐지를 학습 (음소)

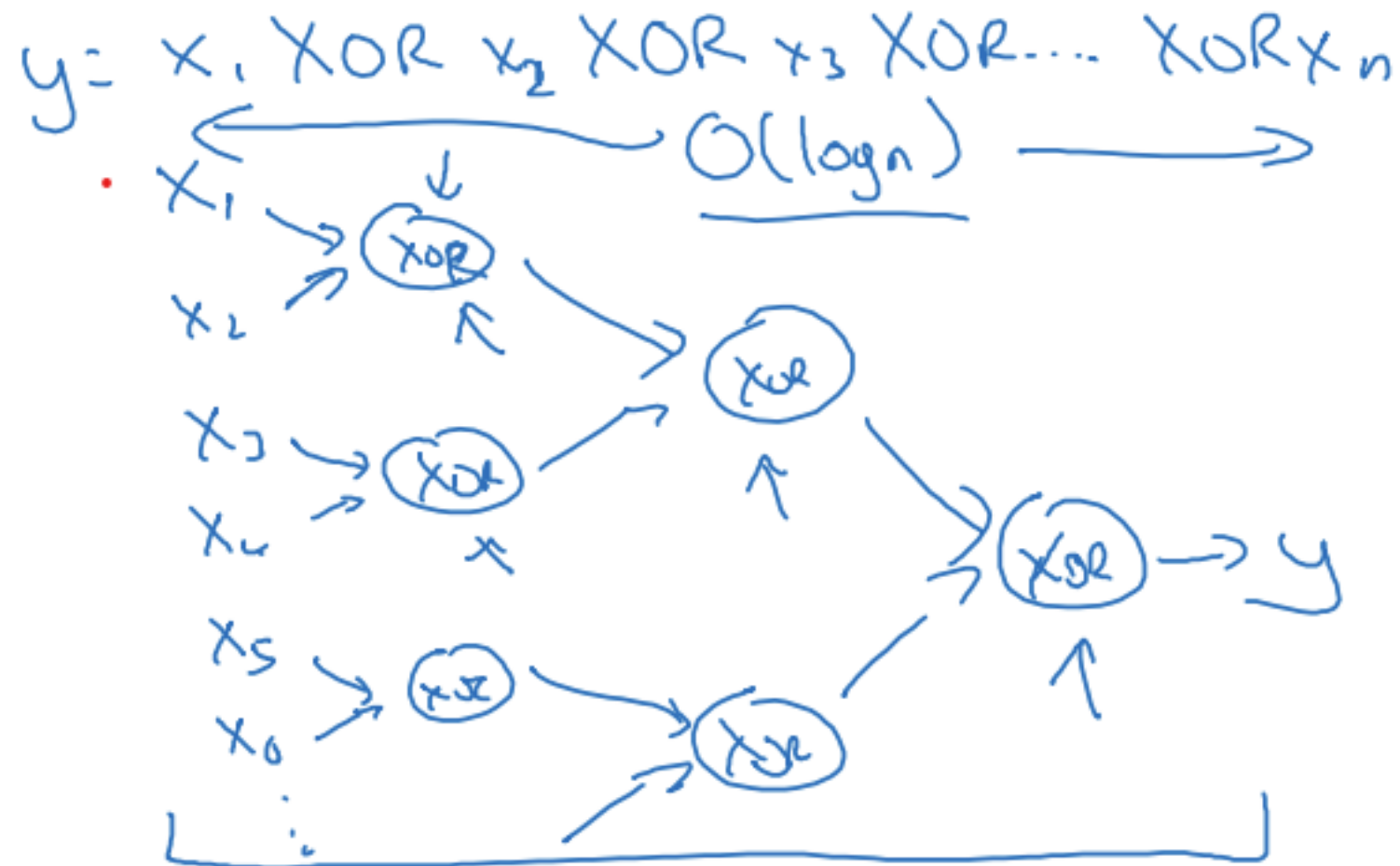
세번째 층 : 단어

네번째 층 : 문장/구

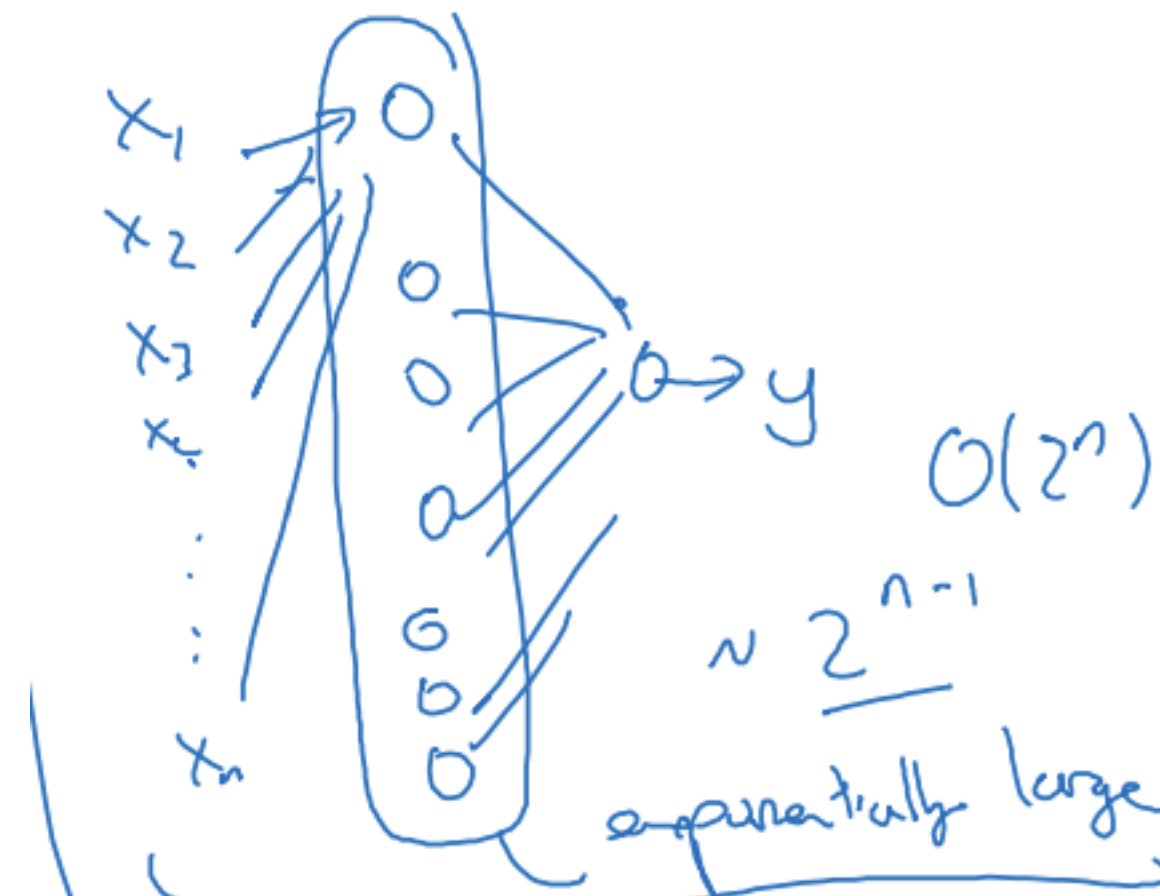


# #04 왜 심층신경망이 더 많은 특징을 잡아낼 수 있을까?

- 순환 이론에서 따르면, 상대적으로 은닉층의 개수가 작지만 깊은 심층 신경망에서 계산할 수 있는 함수가 있음
- 얇은 네트워크로 같은 함수를 계산하려고 하면, 즉 충분한 은닉층이 없다면 기하급수적으로 많은 은닉 유닛이 계산에 필요



여러개의 은닉층이 있는 네트워크  
 $O(\log n)$



매우 큰 은닉층 필요  
 $O(2^n)$

# #05 심층신경망 네트워크 구성하기

Layer  $l$  :  $W[l], b[l]$

## 정방향 전파

Input  $a[l-1]$

Output  $a[l]$

$$z[l] = W[l]a[l-1] + b[l]$$

$$a[l] = g[l](z[l])$$

나중에 사용할 경우를 대비해 cache에  $z[l]$  값을 저장

## 역방향 전파

Input  $da[l]$ /cache  $z[l]$  사용

Output  $da[l-1]$

# #06 정방향전파와 역방향전파

## 정방향 전파

Input  $a[l-1]$

Output  $a[l]$ , cache ( $z[l]$ )

구현

$z[l] = W[l]a[l-1] + b[l]$

$a[l] = g[l](z[l])$

벡터화된 구현

$Z[l] = W[l]A[l-1] + b[l]$

$A[l] = g[l](Z[l])$

## 역방향 전파

Input  $da[l]$

Output  $da[l-1]$ ,  $dW[l]$ ,  $db[l]$

$$\begin{aligned} dz^{[l]} &= da^{[l]} * g'^{[l]}(z^{[l]}) \\ dW^{[l]} &= dz^{[l]} \cdot \underline{a^{[l-1]}} \\ db^{[l]} &= dz^{[l]} \\ \underline{da^{[l-1]}} &= W^{[l]T} \cdot dz^{[l]} \\ \underline{dz^{[l]}} &= \underline{W^{[l+1]T} dz^{[l+1]} * g'^{[l+1]}(z^{[l+1]})} \end{aligned}$$

$$\begin{aligned} dz^{[l]} &= dA^{[l]} * g'^{[l]}(z^{[l]}) \\ dW^{[l]} &= \frac{1}{n} dz^{[l]} \cdot A^{[l-1]T} \\ db^{[l]} &= \frac{1}{n} \text{np.sum}(dz^{[l]}, \text{axis}=1, \text{keepdims}=True) \\ dA^{[l-1]} &= W^{[l]T} \cdot dz^{[l]} \end{aligned}$$

# #07 변수 vs 하이퍼파라미터

 **변수** : 신경망에서 학습 가능한  $w, b$

 **하이퍼 파라미터** : 변수  $w, b$ 를 통제하는 매개변수

- 학습률 (learning rate,  $\alpha$ )
- 반복횟수 (numbers of iteration)
- 은닉층의 갯수 (numbers of hidden layer,  $L$ )
- 은닉유닛의 갯수 (numbers of hidden units)
- 활성화 함수의 선택 (choice of activation function)

# #08 인간의 뇌와 어떤 연관이 있을까?

 **신경망과 인간의 뇌 간의 연관은 그렇게 크지 않다.**

그렇다면 사람들이 왜 딥러닝과 인간의 뇌를 비유하는 것일까?

- 신경망의 복잡한 과정은 직관적으로 전달 어렵
- 단순화해서 뇌세포의 프로세스로 비유하게 되면, 사람들에게 조금 더 직관적/효과적으로 전달 가능

그러나 최근에는 이런 비유가 점점 무너져 가고 있음

- 오늘날 신경 과학자들조차도 하나의 뉴런이 무엇을 하는지 거의 모름
- 하나의 뉴런은 훨씬 더 복잡하고 알기 어려움
- 뉴런이 신경망처럼 역전파를 통해서 학습 하는지도 의문이기 때문

# #퀴즈1

다음 중 하이퍼 파라미터로 맞는 것을 모두 골라주세요.

1. iteration의 횟수
2. 신경망의 레이어 개수  $L$
3. 활성화 레이어  $a^{[l]}$
4. learning rate  $\alpha$
5. 가중치 행렬  $W^{[l]}$
6. 은닉층의 크기  $n^{[l]}$
7. 편향 벡터  $b^{[l]}$

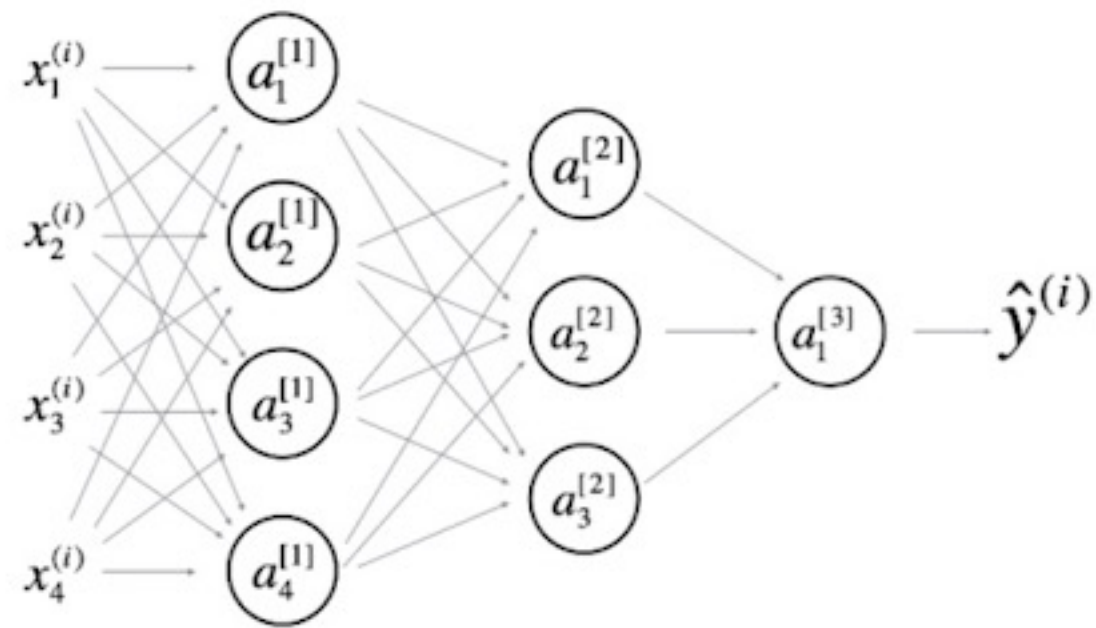
하이퍼파라미터는 머신 러닝 모델의 학습 과정을 제어하고 조절하는 매개변수

신경망 모델에서의 하이퍼파라미터로는 학습률(learning rate), 배치 크기(batch size), 에폭 수(epochs), 은닉층의 수(hidden layer count), 은닉층의 뉴런 수(hidden units), 드롭아웃 비율(dropout rate) 등이 있음

>> 1, 2, 4, 6

# #퀴즈2

다음과 같이 두개의 은닉층을 가진 신경망이 있을 때, 옳은 것을 모두 골라주세요.



1.  $W^{[1]}$ 의 shape는 (4,4)이다.
2.  $W^{[1]}$ 의 shape는 (3,4)이다.
3.  $W^{[2]}$ 의 shape는 (3,4)이다.
4.  $W^{[2]}$ 의 shape는 (3,1)이다.
5.  $W^{[3]}$ 의 shape는 (3,1)이다.
6.  $W^{[3]}$ 의 shape는 (1,3)이다.
7.  $b^{[1]}$ 의 shape는 (4,1)이다.
8.  $b^{[1]}$ 의 shape는 (3,1)이다.
9.  $b^{[2]}$ 의 shape는 (1,1)이다.
10.  $b^{[2]}$ 의 shape는 (3,1)이다.
11.  $b^{[3]}$ 의 shape는 (1,1)이다.
12.  $b^{[3]}$ 의 shape는 (3,1)이다.

입력 데이터의 형태(shape)는  
(데이터 샘플 수, 특성 수)로 결정

여기서 "데이터 샘플 수"는  
데이터의 전체 행 수를 나타내고,  
"특성 수"는 각 데이터 샘플의  
특성(입력 변수)의 수를 나타냄

>> 1, 3, 6, 7, 10, 11



THANK YOU

