

# 5주차

📅 날짜	@2024년 4월 9일
📖 과제	강의 요약 출석 퀴즈
☰ 세부내용	[딥러닝 1단계] 5. 심층 신경망 네트워크
📎 자료	[Week5] 출석퀴즈 C1_W4.pdf
📎 과제물	[Week5] 강의요약.pdf

## 심층 신경망 네트워크

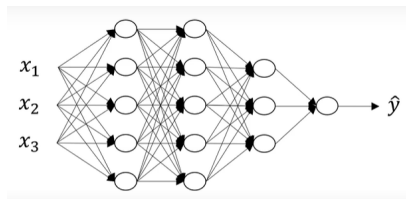
### 1 더 많은 층의 심층 신경망

What is a deep neural network

- 로지스틱 회귀 : 매우 얇은 모델, 1 layer NN
- 1 hidden layer model : 2 layer NN, 로지스틱 회귀보다 덜 얇음
- 2 hidden layer model : 3 layer NN
- 5 hidden layers model : 매우 깊은 모델
- 문제에 따라 얼마나 깊은 신경망을 사용해야 하는지는 알 수 없음
  - 은닉층의 개수가 또 다른 하이퍼파라미터가 됨
  - 은닉층의 개수를 다양하게 시도하고 검증 데이터에서 평가

Deep Neural Network Notation

- 4 layer NN ( 3 hidden layers )



- 각각의 은닉층 유닛 개수 : 5 / 5 / 3 + 1개의 출력 유닛
- $L$  : # of layers  $\rightarrow L = 4$
- $n^{[l]}$  : # of units in layer  $l$ 
  - $\rightarrow n^{[0]} = n_x = 3 \mid n^{[1]} = 5 \mid n^{[2]} = 5 \mid n^{[3]} = 3 \mid n^{[4]} = n^{[L]} = 1$
- $a^{[l]} = g^{[l]}(z^{[l]})$  : activation in layer  $l$ 
  - $\rightarrow X = a^{[0]}, \hat{y} = a^{[L]}$
- $w^{[l]}$  : weights for  $z^{[l]}$

### 2 정방향전파와 역방향전파

Forward Propagation for Layer  $l$

- Input :  $a^{[l-1]}$

- output :  $a^{[l]} = g^{[l]}(z^{[l]})$ , cache ( $z^{[l]} = W^{[l]}a^{[l-1]} + b^{[l]}$ )
- vectorization :  $A^{[l]} = g^{[l]}(Z^{[l]})$ ,  $Z^{[l]} = W^{[l]}A^{[l-1]} + b^{[l]}$
- $A^{[0]}$  : 전체 훈련 데이터셋을 진행할 때의 입력 특성, 정방향 전파의 시작

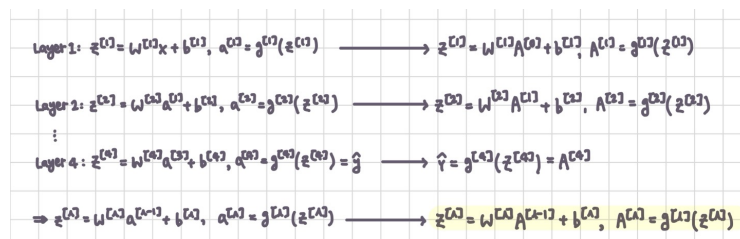
#### Backward Propagation for Layer $l$

- input :  $da^{[l]}$
- output :  $da^{[l-1]}, dW^{[l]}, db^{[l]}$
- $dz^{[l]} = da^{[l]} * g^{[l]'}(z^{[l]})$   
 $dw^{[l]} = dz^{[l]}a^{[l-1]}$   
 $db^{[l]} = dz^{[l]}$   
 $da^{[l-1]} = w^{[l]T} dz^{[l]}$
- vectorization  
 $dZ^{[l]} = dA^{[l]} * g^{[l]'}(Z^{[l]})$   
 $dW^{[l]} = \frac{1}{m} dZ^{[l]} A^{[l-1]T}$   
 $db^{[l]} = \frac{1}{m} \text{np.sum}(dZ^{[l]}, \text{axis}=1, \text{keepdims}=\text{True})$   
 $dA^{[l-1]} = W^{[l]T} dZ^{[l]}$

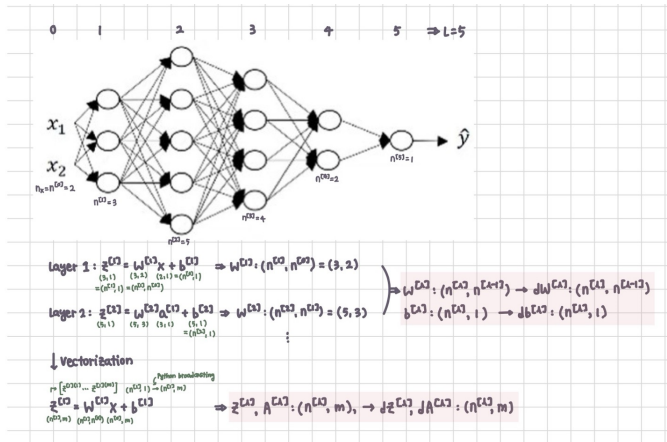
#### Summary

- $X \rightarrow \text{ReLU} \rightarrow \text{ReLU} \rightarrow \text{Sigmoid} \rightarrow \hat{y} \rightarrow L(\hat{y}, y)$
- $L(\hat{y}, y) \rightarrow dw^{[3]}, db^{[3]}$  계산하여  $da^{[2]}$  전달  $\rightarrow dw^{[2]}, db^{[2]}$  계산하여  $da^{[1]}$  전달  $\rightarrow dw^{[1]}, db^{[1]}$  계산  
 $\Rightarrow$  캐시에서  $z^{[1]}, z^{[2]}, z^{[3]}$ 를 옮김
- 정방향 반복은 입력  $X$ 로 초기화
- 역방향 반복은  $da^{[L]}$ 로 초기화  $\rightarrow$  벡터화  $dA^{[L]}$ 로 초기화

### 3 심층 신경망에서의 정방향 전파



### 4 행렬의 차원을 알맞게 만들기

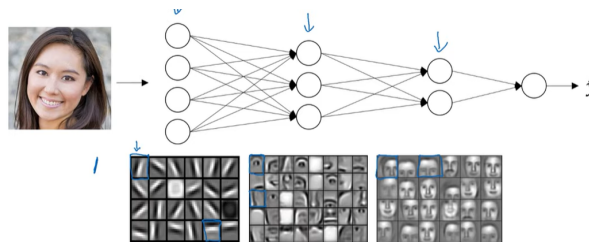


- 같은 층의 W와 dW의 차원, b와 db의 차원은 서로 같아야 함
- 같은 층의 z와 a의 차원은 서로 같아야 함

## 5 왜 심층 신경망이 더 많은 특징을 잡아낼 수 있을까요?

Intuition about deep representation

- 얼굴 인식이나 감지 같은 시스템 구축 시 심층 신경망에서 일어나는 일

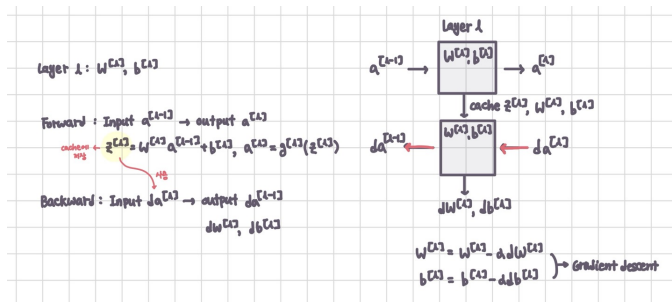


- 이미지를 입력으로 받음
- Layer 1 : 20개의 은닉 유닛이 사진을 보고 모서리가 어디에 있는지 파악  
→ 모서리 탐지기는 이미지의 더 좁은 영역을 볼 수 있음
- Layer 2 : 모서리를 그룹화해서 눈, 코 등의 얼굴 일부를 감지
- Layer 3 : 서로 다른 얼굴의 일부를 그룹화  
→ 얼굴 탐지기는 이미지의 더 넓은 영역을 볼 수 있음
- Layer 4(출력) : 최종적으로 서로 다른 얼굴을 감지할 수 있게 됨
- 간단한 것을 먼저 찾아 모으고 모아서 복잡한 것을 찾아나가는 과정
- 음성 인식의 경우 : 파형 특징 파악(low level) → 음소 파악 → 음성의 단어 인식 → 구/문장 인식  
⇒ 낮은 단계에서는 간단한 특징을, 깊은 층에서는 더 복잡한 특징을 탐지해 나감

Circuit theory and deep learning

- 회로 이론 : 논리 게이트의 서로 다른 게이트에서 어떤 종류의 함수를 계산할 수 있을지에 관한 것
- small DNN에서 계산할 수 있는 함수를 알면 NN에서 계산하려고 하면 더 많은 은닉 유닛이 필요함
- $X1 \text{ XOR } X1 \text{ XOR } X3 \text{ XOR } \dots \text{ XOR } Xn \rightarrow \text{XOR 트리로 계산하게 됨} \Rightarrow O(\log n)$
- 여러 은닉층을 갖지 못하는 경우 → 은닉 노드가 기하급수적으로 많이 필요  $\Rightarrow O(2^n)$   
⇒ 얇은 네트워크보다 깊은 네트워크에서 계산하기 쉬운 수학적 계산이 존재

## 6 심층 신경망 네트워크 구성하기



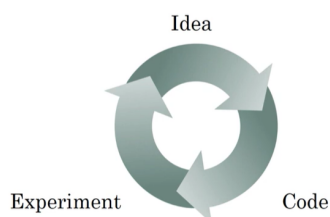
## 7 변수 vs 하이퍼파라미터

What are hyperparameters?

- Parameters :  $W^{[1]}, b^{[1]}, W^{[2]}, b^{[2]}, \dots$
- Hyperparameters
  - learning rate ( $\alpha$ ) : 매개변수가 어떻게 진전될지 결정
  - # of iterations : 경사하강법의 반복 횟수
  - # of hidden layers (L) : 은닉층의 수
  - # of hidden units ( $n^{[1]}, n^{[2]}, \dots$ ) : 은닉 유닛의 수
  - Choice of activation functions : ReLU, tanh, sigmoid ...
  - Momentum term, mini-batch size, 정규화 매개변수 ...

$\Rightarrow$  하이퍼파라미터가 최종 매개변수인  $W$ 와  $b$ 를 통제

Applied deep learning is a very empirical process

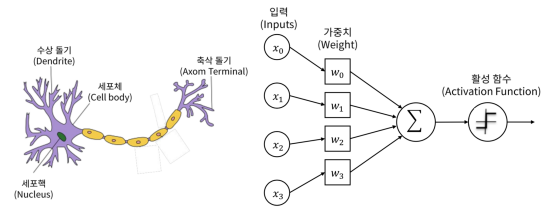


- Idea :  $\alpha = 0.01 \rightarrow$  code로 구현하고 결과를 확인  $\rightarrow$  다시  $\alpha$ 값 조정  $\rightarrow$  반복
  - 비용함수가 더 낮은 값으로 수렴하도록 하는 하이퍼파라미터를 찾는 과정을 거침
  - 경험적인 과정 : 많은 것을 시도하고 결과를 확인하는 것
  - 적절한 하이퍼파라미터를 찾아도 컴퓨터 환경에 따라서 또 달라질 수 있음
- $\Rightarrow$  다양한 하이퍼파라미터를 적용해보고 평가해보기

## 8 인간의 뇌와 어떤 연관이 있을까요?

- 딥러닝 구현 시 정방향 전파와 역방향 전파를 하게 됨
- $\rightarrow$  이 복잡한 식을 "It's like the brain" 이라고 간단하게 비유

## 신경 세포



생물학적인 신경 세포를 단순화하여 모델링한 Neuron. 여러 신호를 받아, 하나의 신호를 만들어 전달하는 역할을 한다.  
출력을 내기 전에 활성 함수(activation function)를 통해 비선형 특성을 가질 수 있다.

- 하지만 오늘날 신경과학도 아직 하나의 뉴런이 무엇을 하는지 거의 알지 못함
- 우리의 생각보다 하나의 뉴런은 훨씬 더 복잡하고 알기 어려움
  - ⇒ 뉴런과 딥러닝의 비유가 점점 무너져 가는 중