

[Week5]_이원주

신경망 학습 (딥러닝)

흐름

	input	output
순전파	$a^{[l-1]}$ (+) output 계산을 위해 $w^{[l-1]}, b^{[l-1]}$ 도 이용	$a^{[l]}$ (+) 역전파 때 쓰려고 $z^{[l]}$ 도 저장
역전파	$da^{[l]}$ (+) 저장해둔 $z^{[l]}$ 도 input으로 이용 (+) 마찬가지로 $w^{[l-1]}, b^{[l-1]}$ 도 이용	$da^{[l-1]}$ (+) loss 계산을 위해 $dw^{[l-1]}, db^{[l-1]}$ 도 계산

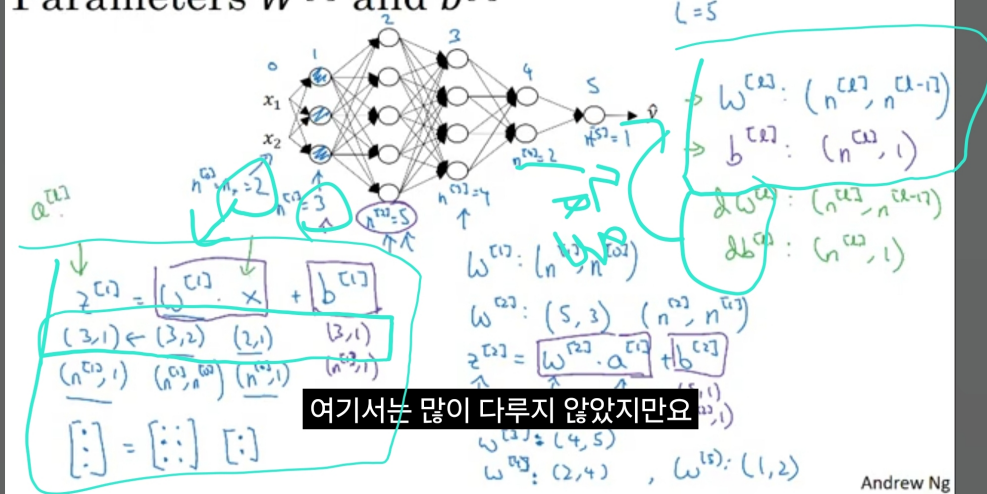
▼ 표 부연설명

- $a^{[l-1]}$: level $(l - 1)$ 의 $a = \sigma(z)$
- 정전파 → a 계산
역전파 →
 da 계산 (a 를 미분한 것)
- 순전파/역전파 구현에서
 - 캐시(cache)는 순전파에서의 변수를 → 해당하는 역전파 단계에 전달한다.
 - output인 a 말고도 z, w, b 를 전달하면 미분값을 계산할 때 유용하기 때문이다.

디버깅 : 행렬 shape을 확인하자

▼ 추후 정리

Parameters $W^{[l]}$ and $b^{[l]}$



그러나 z, a, x 의 차원은
조금 달라집니다

$$z^{[l]} = W^{[l]} \cdot x + b^{[l]}$$

$(n^{[l]}, 1)$ $(n^{[l]}, n^{[l-1]})$ $(n^{[l-1]}, 1)$ $(n^{[l-1]}, 1)$

$$[z^{[1]}, z^{[2]}, \dots, z^{[L]}]$$

$$\rightarrow Z = W \cdot X + b$$

$(n^{[L]}, m)$ $(n^{[L]}, n^{[L-1]})$ $(n^{[L-1]}, m)$ $(n^{[L]}, 1)$

아래가 벡터화된 거.

- w, b 는 모든 데이터에 대해 공통이므로 그대로, Z, X 만 바뀜.

$$[z^{[1]}, z^{[2]}, \dots, z^{[L]}]$$

$$z^{[1]} = \begin{bmatrix} -w_1^{[1]T} \\ -w_2^{[1]T} \\ \vdots \\ -w_n^{[1]T} \end{bmatrix} \begin{bmatrix} x^{(1)} & x^{(2)} & \dots & x^{(m)} \\ | & | & & | \end{bmatrix} + b^{[1]}$$

- b 는 그대로긴 한데 $(n^{[1]}, m)$ 으로 브로드캐스팅된 후 더해짐.

$$z^{[u]}, a^{[u]} : (n^{[u]}, 1)$$

$$\rightarrow z^{[u]}, A^{[u]} : (n^{[u]}, m)$$

$$\rightarrow dz^{[u]}, dA^{[u]} : (n^{[u]}, m)$$

변수(파라미터) vs. 하이퍼 파라미터(매개변수)



[요약]

- 변수 : 신경망에서 학습 대상
 - w , b
- 하이퍼 파라미터 : 여러 번 시도해서 학습 알고리즘에게 우리가 정해줘야 하는 것
 - learning rate α
 - 반복 *iteration* 횟수
 - 은닉층 *hidden layer* 개수 L
 - 은닉유닛 *hidden units* 개수 $n^{[l]}$
 - 활성화 함수 선택 $\sigma^{[l]}$
 - 모멘텀 항 *momentum term*
 - 미니배치 크기 *mini batch size*

하이퍼 파라미터(매개변수) 초기화

- 방법
 - 매우 경험적임
 - 하나로 정하고 → 여러 번 시도 → 비용함수 J 값 보면서 조정
- 조언
 - 몇 년씩 하는 장기 프로젝트라면 → 몇 달마다 점검하기.
 - 왜냐면 지금 최적인 경로 정했어도, 나중에 최적인 값이 바뀔 수도 있음.
CPU, GPU, 네트워크, 데이터가 달라지니까.

learning rate



[요약]

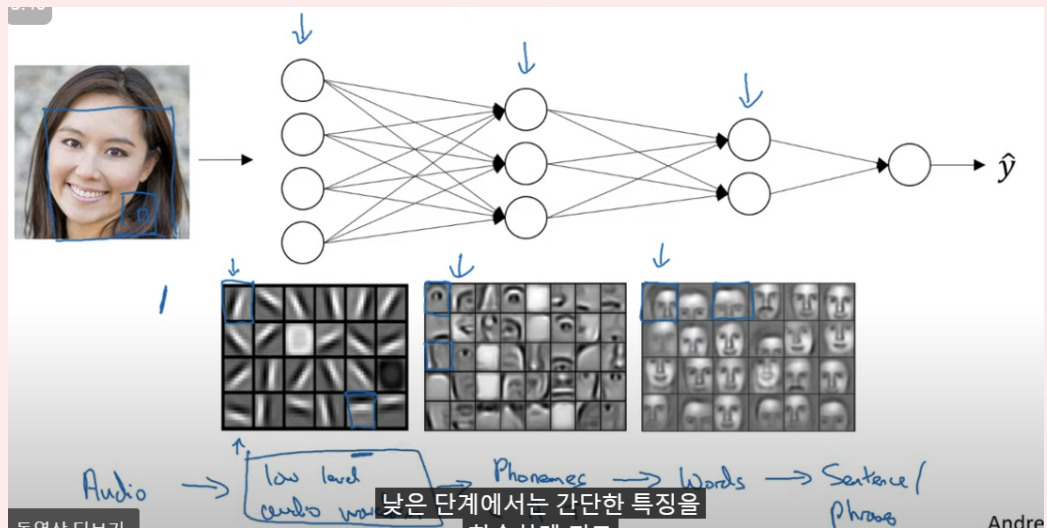
- learning rate 정하기
 - 일단 **0.01**로 시작 (?)

심층 신경망

얕은 신경망 vs. 심층 신경망

- 정의 :
 - 층이 얕은지 vs 깊은지 (은닉층 多)
- 비교 :
 - 심층 신경망이 더 잘 작동함.
- 이유 :
 - 같은 성능을 내려면, 얕은 신경망의 경우 훨씬 많은 노드가 필요함. (기하급수적)
 -

▼ 추후 정리



- ← 낮은 층 : 간단한 특징을 탐지
- 깊은 층 → : 그 간단한 특징을 모아서 복잡한 특징을 탐지
- ex)

이미지 (얼굴)	모서리	얼굴 일부 (코, 입술, 눈)	얼굴로 사람 식별	
----------	-----	---------------------	-----------	--

	오디오	말소리 (말하고 있는지 백 색소음인지)	음소 (한 글자)	단어	문장
--	-----	-----------------------------	--------------	----	----