



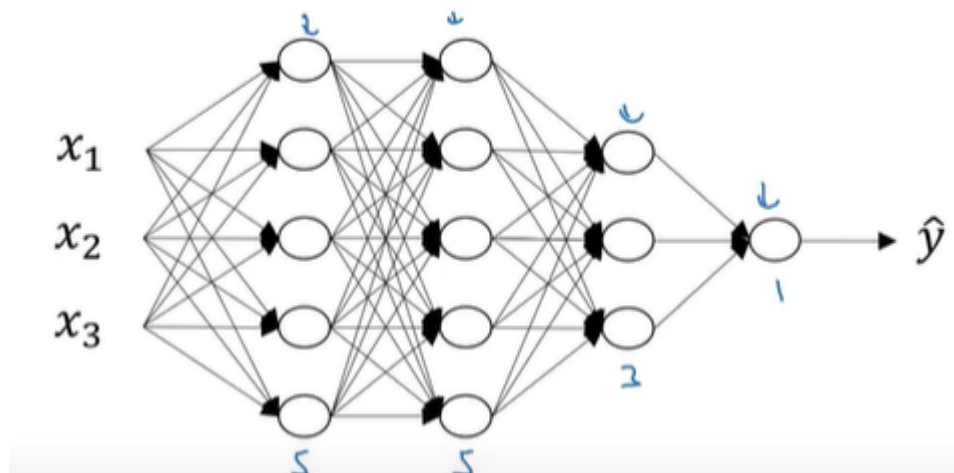
5주차

태그

완료

더 많은 층의 심층 신경망

- 신경망의 층을 셀 때는 은닉층과 출력층의 개수만 고려한다.
- 얼마나 깊은 신경망을 사용해야 하는지 미리 예측하기 어렵다.



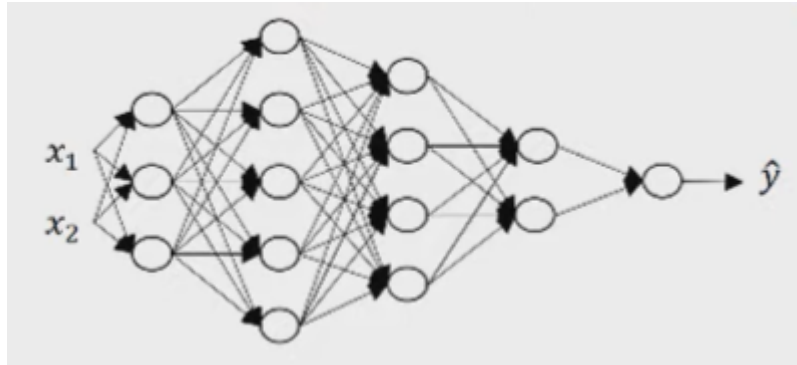
- $L=4(\text{\#layers})$
- $n^{[l]} = \text{\#units in layer } l$
 - $n^{[1]} = 5, n^{[2]} = 5, n^{[3]} = 3, n^{[4]} = 1, n^{[0]} = n_x = 3$
- $a^{[l]} = \text{activations in layer } l$
- $a^{[0]} = X, a^{[L]} = \hat{y}$

심층 신경망에서의 정방향 전파

- 일반적인 정방향 전파 수식- l 층에 대하여
 - $z^{[l]} = W^{[l]} * a^{[l-1]} + b^{[l]}$
 - $a^{[l]} = g^{[l]}(z^{[l]})$

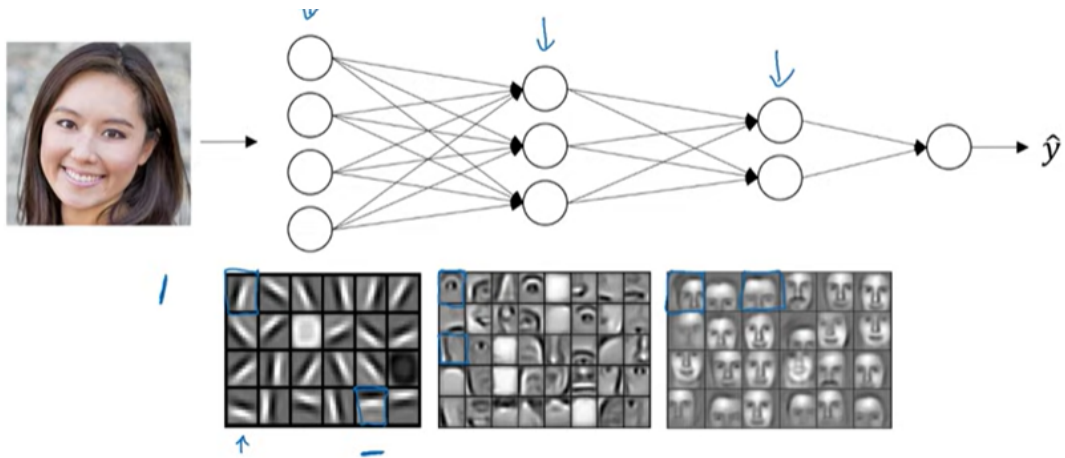
- vectorized- l 층의 m 개의 데이터셋에 대하여
 - $Z^{[l]} = W^{[l]} * A^{[l-1]} + b^{[l]}$
 - $A^{[l]} = g^{[l]}(Z^{[l]})$
 - $l = 1, 2, \dots, L$ (L 은 #layer)만큼 for-loop를 써서 반복한다.

행렬의 차원을 알맞게 만들기



- $L=5$ 인 심층 신경망이다.
- $z^{[l]} = (n^{[l]}, 1)$
- $Z^{[l]} = (n^{[l]}, m)$: $z^{[l]}$ 열벡터로 이루어짐
- $W^{[l]} = (n^{[l]}, n^{[l-1]}) = dW^{[l]}$
- $a^{[l-1]} = (n^{[l-1]}, 1)$
- $A^{[l-1]} = (n^{[l-1]}, m)$: $a^{[l-1]}$ 열벡터로 이루어짐
- $b^{[l]} = (n^{[l]}, 1) = db^{[l]}$: m 개의 데이터셋이 있으면 차원은 $(n^{[l]}, m)$ 이 됨

왜 심층 신경망이 더 많은 특징을 잡아낼 수 있을까요?

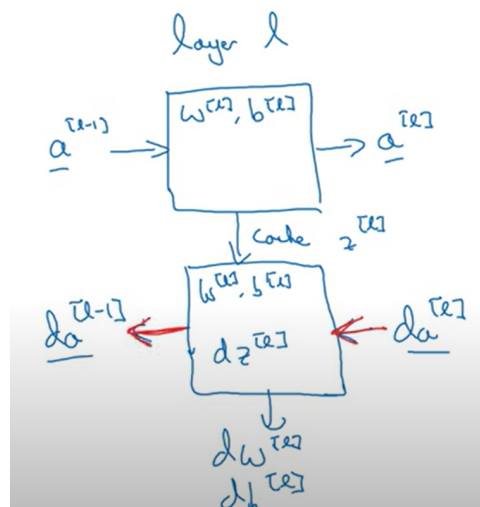


- 첫 번째 은닉층에서는 수직 방향의 모서리와 수평 방향의 모서리를 찾는다.
- 두 번째 은닉층에서는 첫 번째 은닉층의 결과를 이용해 눈, 코, 입과 같은 얼굴의 특징을 찾는다.
- 세 번째 은닉층에서는 두 번째 은닉층의 결과를 이용해 얼굴을 인식한다.

1 심층 신경망이 깊어질수록 낮은 레이어에서는 간단한 특징을, 높은 레이어에서는 복잡하고 다양한 특징을 많이 추출할 수 있다.

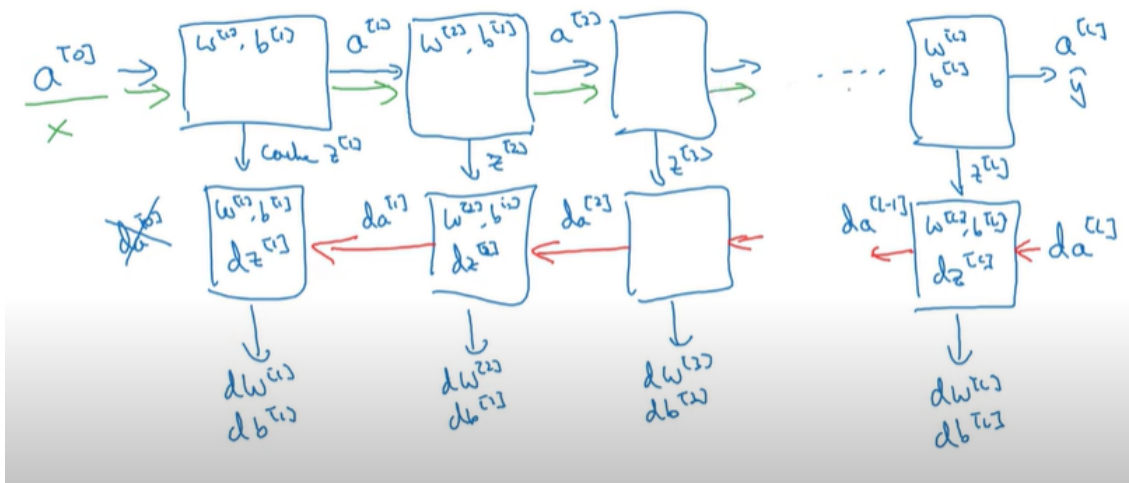
2 순환 이론에 따르면, 은닉층이 얼마 없다면 은닉층의 노드 개수가 기하급수적으로 많아져야 한다. $\rightarrow 2^{(n-1)}$ 개의 노드가 필요하다.

심층 신경망 네트워크 구성하기



- 정방향 전파의 레이어 l 에서 $a^{[l-1]}$ 을 입력으로 받아 $a^{[l]}$ 을 출력으로 내놓는다. 이 과정에서 $z^{[l]}, W^{[l]}, b^{[l]}$ 값을 캐시값으로 저장한다.

- 역방향 전파는 $da^{[l]}$ 을 입력으로 받아 $da^{[l-1]}$ 을 출력으로 내놓는다. 이때 $dz^{[l]}, dW^{[l]}, db^{[l]}$ 도 계산한다.
- 전체적 흐름은 아래 그림과 같다.
- 이후 $dW^{[l]}$ 와 $db^{[l]}$ 으로 w 와 b 를 업데이트한다.



정방향전파와 역방향전파

- 정방향전파 for layer l
 - input $a^{[l-1]}$
 - output $a^{[l]}$
 - 한 번에 한 개의 데이터셋: $z^{[l]} = W^{[l]} * a^{[l-1]} + b^{[l]}$
 - 한 번에 m개의 데이터셋(vectorized): $Z^{[l]} = W^{[l]} * A^{[l-1]} + b^{[l]}$
 - $z^{[l]}, W^{[l]}, b^{[l]}$ 의 값을 캐시로 저장해둔다.
 - 가중치 $W^{[l]}$ 은 행 벡터로 이루어져 있으며, 역방향 전파에 의해 값이 업데이트된다.

$$Z^{[l]} = \begin{bmatrix} | & | & | \\ z^{[l]}(1) & z^{[l]}(2) & \dots & z^{[l]}(m) \\ | & | & | \end{bmatrix} \in (n-x, m)$$

- $n-x$: l 층의 노드 개수
- m : training-set 개수

$$W^{[l]} = \begin{bmatrix} - & \omega^{[1]} & - \\ - & \omega^{[2]} & - \\ & \vdots & \\ - & \omega^{[l]} & - \end{bmatrix} \in (n-x, n-x^{[l-1]})$$

- $n-x$: l 층의 노드 개수
- $n-x^{[l-1]}$: $(l-1)$ 층의 노드 개수

$$A^{[l-1]} = \begin{bmatrix} | & | \\ a^{[l-1]}(1) & \dots & a^{[l-1]}(m) \\ | & | \end{bmatrix} \in (n-x^{[l-1]}, m)$$

- $n-x^{[l-1]}$: $(l-1)$ 층의 노드 개수
- m : training-set 개수

$$b^{[l]} = \begin{bmatrix} | & | \\ b^{[l]}(1) & \dots & b^{[l]}(m) \\ | & | \end{bmatrix} \in (n-x, m)$$

- $n-x$: l 층의 노드 개수
- m : training-set 개수

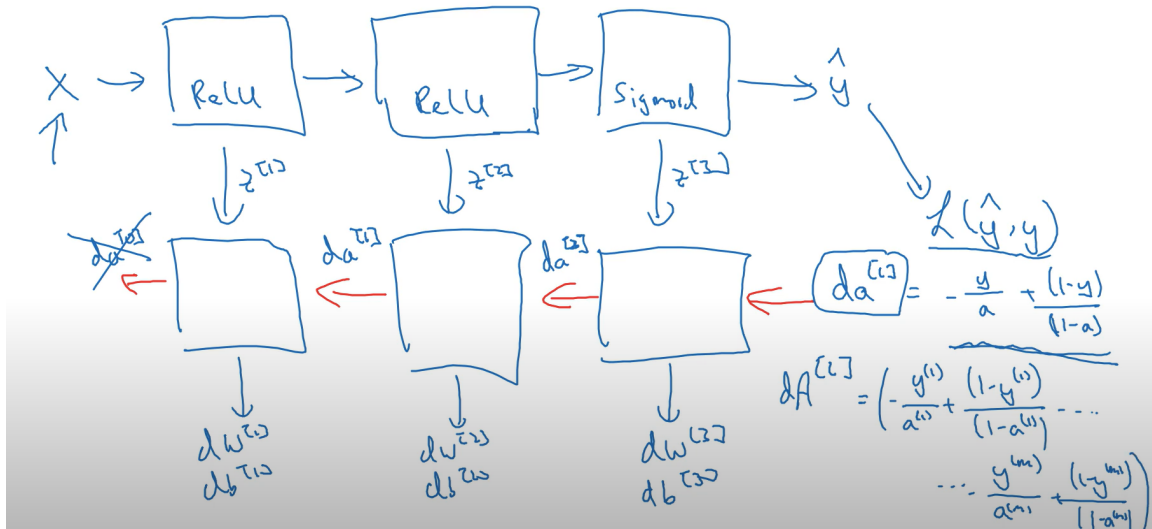
• 역방향전파 for layer l

- input $da^{[l]}, z^{[l]}$
- 역방향 전파의 계산에서 전방향 전파 때 저장해두었던 캐시($z^{[l]}, W^{[l]}, b^{[l]}$)를 사용한다.

→ Output $da^{[l-1]}, dW^{[l]}, db^{[l]}$

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

$$\begin{aligned} dz^{[l]} &= dA^{[l]} * g^{[l]}(z^{[l]}) \\ dW^{[l]} &= \frac{1}{m} dz^{[l]} \cdot A^{[l-1]T} \\ db^{[l]} &= \frac{1}{m} np.sum(dz^{[l]}, axis=1, keepdims=True) \\ dA^{[l-1]} &= W^{[l]T} \cdot dz^{[l]} \end{aligned}$$



- 정방향 전파는 입력값 X 로 값을 초기화한다.
- 역방향 전파는 $da^{[l]}$ 로 값을 초기화한다. $dA^{[l]}$ 은 벡터화한 것으로, $da^{[1]} \dots da^{[m]}$ 을 모두 더한 값으로 구할 수 있다.

변수 vs 하이퍼파라미터

- parameters: w, b
- hyperparameter: learning rate α , #iterations, #hidden layers, #hidden units(node), choice of activation function, momentum term(concept of velocity to the parameter update), mini-batch size
- hyperparameter로 최종 모델의 변수 parameter를 통제할 수 있다.
- applied deep learning is a very **empirical** process \rightarrow time consuming, hard, variable trials are required.
- 하이퍼 파라미터는 결정된 값이 없으며 여러 번의 시도를 통해 적절한 하이퍼파라미터를 찾아야 한다.

인간의 뇌와 어떤 연관이 있을까요?

- 뇌와 딥러닝을 연관지어 설명하는 것은 적절하지 않을 수 있다. 그렇지만 직관적으로 이해하거나 상상력을 자극하기에는 좋은 비유다.
- 하나의 뉴런과 단일 로지스틱 신경망이 대응된다. 입력을 받은 후 활성화 함수를 거친 값이 특정 값 이상이면 신호가 활성화된다는 점이 유사하다.

QUIZ

✕ 1. 다음 중 하이퍼 파라미터로 맞는 것을 모두 골라주세요. *

0/1

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

<input checked="" type="checkbox"/>	1	✓
<input checked="" type="checkbox"/>	2	✓
<input checked="" type="checkbox"/>	3	✕
<input checked="" type="checkbox"/>	4	✓
<input type="checkbox"/>	5	
<input checked="" type="checkbox"/>	6	✓
<input type="checkbox"/>	7	