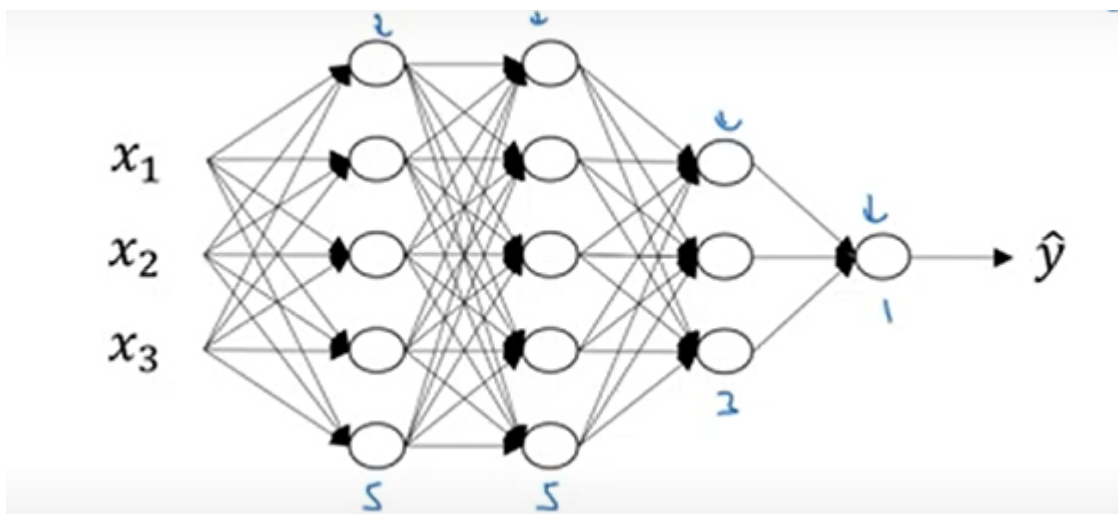
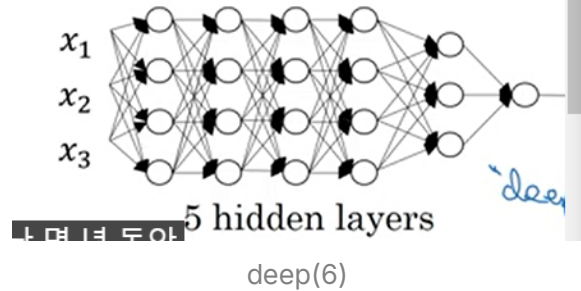
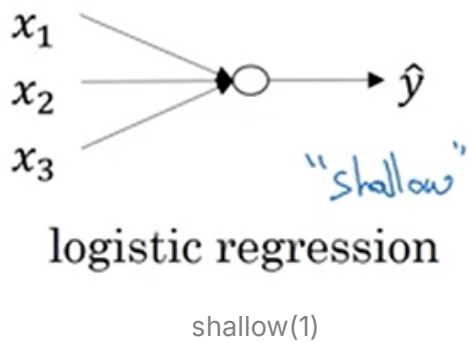


# 유런 5주차 변지은

## 더 많은 층의 심층 신경망



- 표기법
  - $L$  : 네트워크 층의 수 (4)
  - $n[l]$  :  $l$ 층에 있는 유닛 개수 (5,5,2,1)
  - $a[l]$  :  $l$ 층에서의 활성화값
  - $a[0]$  : 입력 특징 ( $X$ )
  - $a[L]$  : 예측된 출력값 ( $\hat{y}$ )

## 정방향 전파와 역방향 전파

Verwijl:

$$z^{[l]} = W^{[l]} \cdot A^{[l-1]} + b^{[l]}$$

$$A^{[l]} = g^{[l]}(z^{[l]})$$

정방향 전파

- $l$  번째 층에서 정방향 전파는 이전 층의 활성화 값인  $a^{[l-1]}$ 을 입력으로 받고, 다음 층으로  $a^{[l]}$  값을 출력으로 나오게 한다. 이때 선형결합된 값인  $z^{[l]}$  와 변수  $W^{[l]}, b^{[l]}$  값도 캐시로 저장해둔다.

→ Input  $da^{[l]}$

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

$$dz^{[l]} = da^{[l]} * g^{[l]'}(z^{[l]})$$

$$dW^{[l]} = dz^{[l]} \cdot a^{[l-1]}$$

$$db^{[l]} = dz^{[l]}$$

$$da^{[l-1]} = W^{[l]T} \cdot dz^{[l]}$$

$$dz^{[l]} = W^{[l+1]} \cdot dz^{[l+1]}$$

따라서 역방향 함수를 구현하는 방법이었습니다

$$dz^{[l]} = dA^{[l]} * g^{[l]'}(z^{[l]})$$

$$dW^{[l]} = \frac{1}{n} dz^{[l]} \cdot A^{[l-1]T}$$

$$db^{[l]} = \frac{1}{n} np.sum(dz^{[l]}, axis=1, keepdims=True)$$

$$dA^{[l-1]} = W^{[l]T} \cdot dz^{[l]}$$

Andrew Ng

역방향 전파

- $l$  번째 층에서 역방향 전파는  $da^{[l]}$  을 입력으로 받고,  $da^{[l]}$  를 출력한다. 이때 업데이트를 위한  $dW^{[l]}$  와  $db^{[l]}$  도 함께 출력한다. 이들을 계산하기 위해서 전방향 함수 때 저장해두었던 캐시를 쓰게 된다.

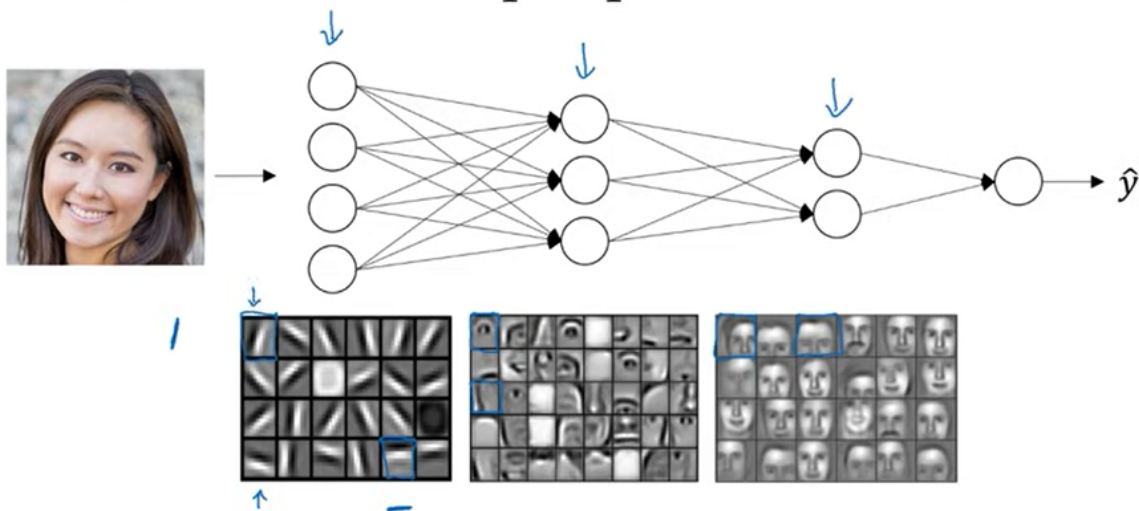
## 행렬의 차원을 알맞게 만들기

$$z^{[1]} = w^{[1]} * x + b^{[1]}$$

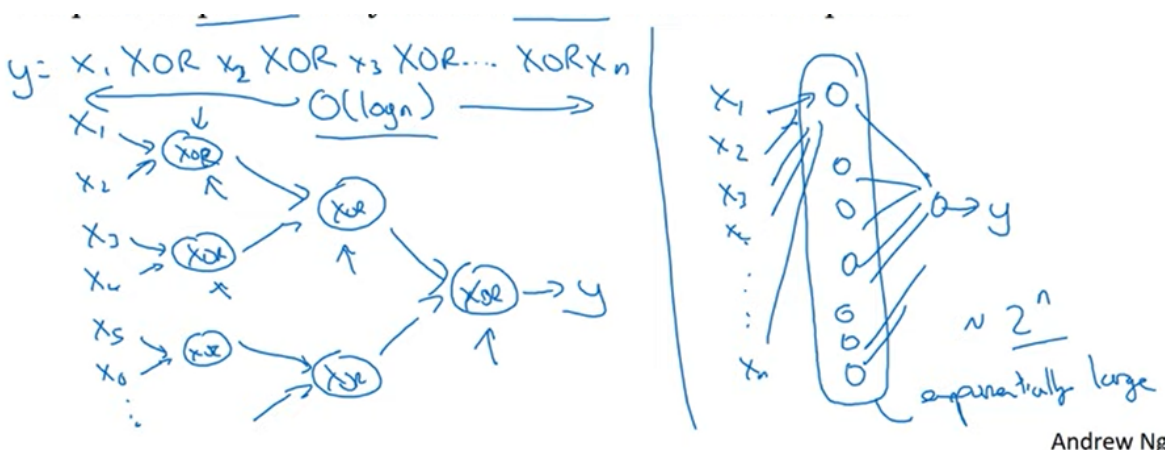
- $w^{[1]}$  : 각 행은 다음 layer 의 각 노드에 들어가는 weight 들이다. 예를들어  $w^{[1]}$ 의 1행은 layer 2 의 첫번째 노드에 들어오는 weight 들이다.
- $dw^{[1]}$  :  $w^{[1]}$ 의 차원과 같다.

## 왜 심층 신경망이 더 많은 특징을 잡아 낼수 있을까?

### Intuition about deep representation



- 직관 1: 네트워크가 더 깊어 질수록, 더 많은 특징을 잡아낼 수가 있다. 낮은 층에서는 간단한 특징을 찾아내고, 깊은 층에서는 탐지된 간단한 것들을 함께 모아 복잡한 특징을 찾아낼 수 있다.



Andrew Ng

정확하게는  $2^{(n-1)}$ 의 은닉 유닛이  
필요할 것입니다

- 직관 2: 순환 이론에서 따르면, 상대적으로 은닉층의 개수가 작지만 깊은 심층 신경망에서 계산할 수 있는 함수가 있다. 그러나 얇은 네트워크로 같은 함수를 계산하려고 하면, 즉 충분한 은닉층이 없다면 기하급수적으로 많은 은닉 유닛이 계산에 필요하게 될 것이다. 위의 사진이 그 예시이다.
  - 순환 이론: 로직 게이트의 서로 다른 게이트에서 어떤 종류의 함수를 계산할 수 있을지에 관한 것이다.

## 심층 신경망 네트워크 구성하기

- 캐시에 선형 결합된 값인  $z^{[l]}$  와 변수  $W^{[l]}, b^{[l]}$  값을 저장한다
- 캐시에 들어있는 값은 역방향 전파를 할 때 사용된다.

## 변수 vs 하이퍼파라미터

- 변수란 신경망에서 학습 가능한  $W$  와  $b$  를 뜻한다.
- 하이퍼파라미터
  - 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)
  - 모멘텀항(momentum term)
  - 미니배치 크기(mini batch size)
  - 정규화 매개변수
- 매개변수인 하이퍼파라미터를 결정함으로서 최종 모델의 변수를 통제할 수 있다.
- 하이퍼파라미터는 결정 된 것이 없으며, 여러 번의 시도를 통해 적합한 하이퍼파라미터를 찾아야 한다.

## 인간의 뇌와 어떤 점이 닮았을까?

사실 그렇게 관련 없는데 매체에 알릴 때에 매우 매력적임.

뉴런의 역할은 불가사의적임. 아직 사람의 뉴런이 어떤 역할을 하는지 거의 모른다.  
딥러닝 분야와 뇌의 비유는 거의 무너져 가는 중이다.