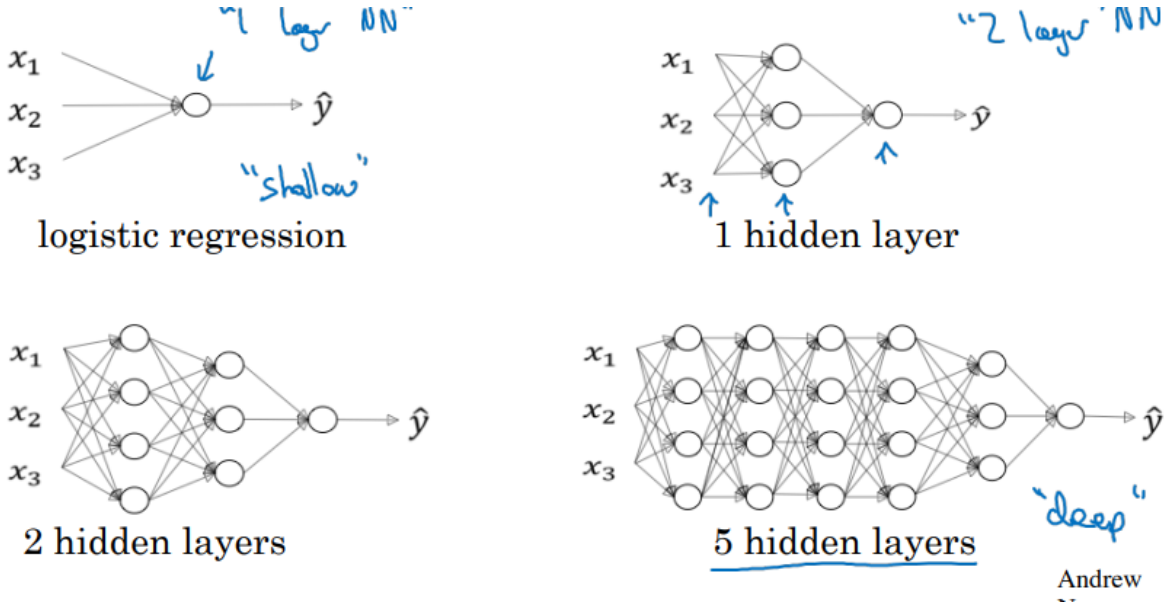
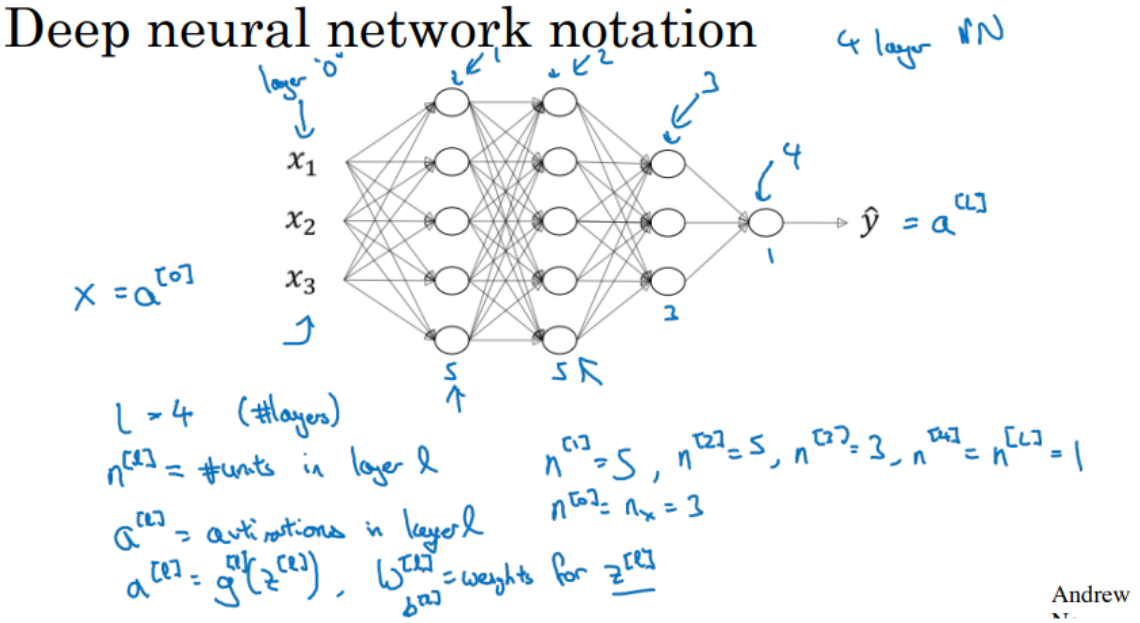
5장. 심층 신경망 네트워크

1. 더 많은 층의 심층 신경망

- 로지스틱 회귀 = **얕은 모델(Shallow)** = 한 층의 신경망
- 얕음 & 깊음 : 정도의 문제



- 1개의 은닉층이 있는 모델 = 총 2개의 층이 있는 신경망
  - 입력층은 층으로 세지 X
  - 은닉층의 개수 : 또 다른 하이퍼 파라미터 취급.
- 표기법



- **L: 네트워크의 층의 수** → L=4
- **n^[l]: L층의 단위 개수** → n^[1] = 5, n^[2]=5, n^[3]=3, n^[4]=n^[L]=1, n^[0]=n_x=3
- **a^[l] : 층 l에서의 활성값**
  - 정방향 전파에서 a^[l]: **z^[1]에 대해 활성화 함수 g를 계산한 값**
- **w^[l]:** z^[1]의 값을 계산하기 위한 가중치
- **입력 특징** : X = 층 0의 활성값 = a^[0]
- **마지막 층의 활성값** = a^[L] = Y의 예측값 = 신경망의 예측된 출력값

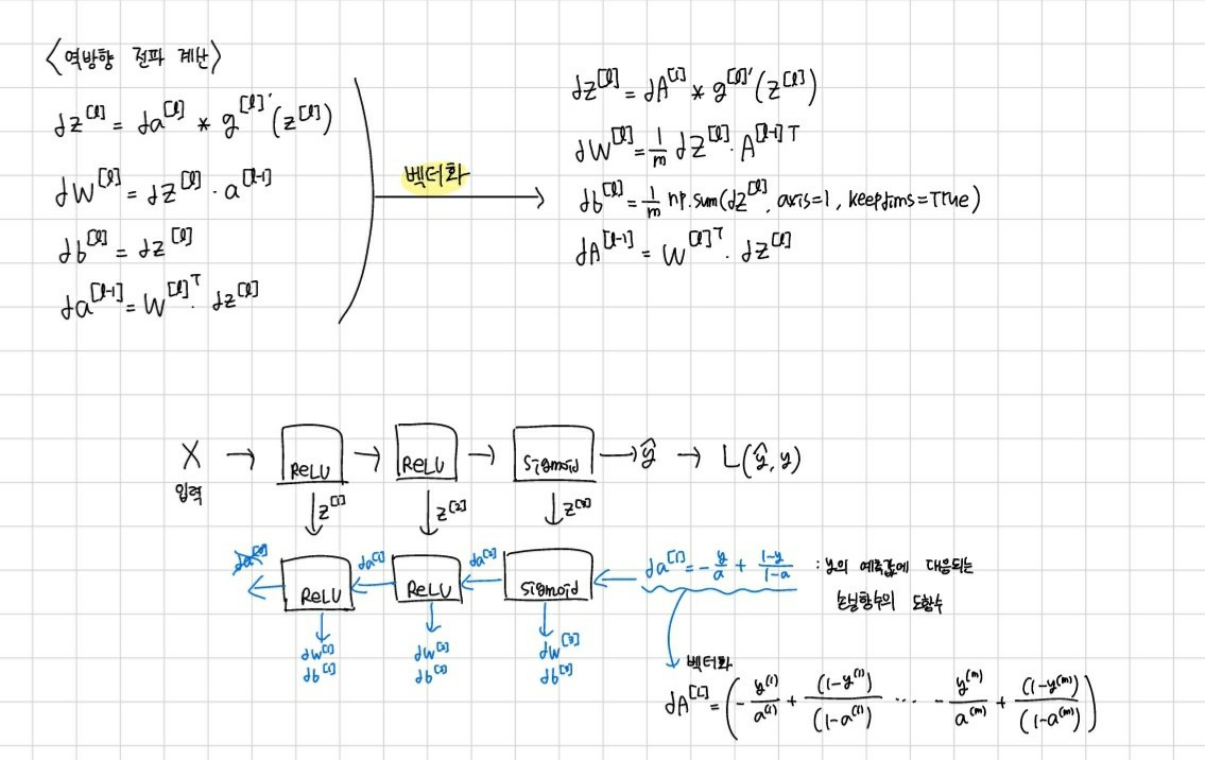
2. 정방향 전파와 역방향 전파

1) 정방향 전파

- Input: a^[l-1]
- Output: a^[l], cache(z^[l])
- 함수 구현: z^[l] = W^[l]*a^[l-1] + b^[l]
  - a^[l] = g^[l](z^[l])
  - → 벡터화 구현: Z^[l] = W^[l]*A^[l-1]+b^[l]
    A^[l] = g^[l](Z^[l])

- a^[0] : 한 번에 하나씩 할 경우의 학습 데이터에 대한 입력 특성
- A^[0] : 전체 학습 세트를 진행할 때의 입력 특성

2) 역방향 전파

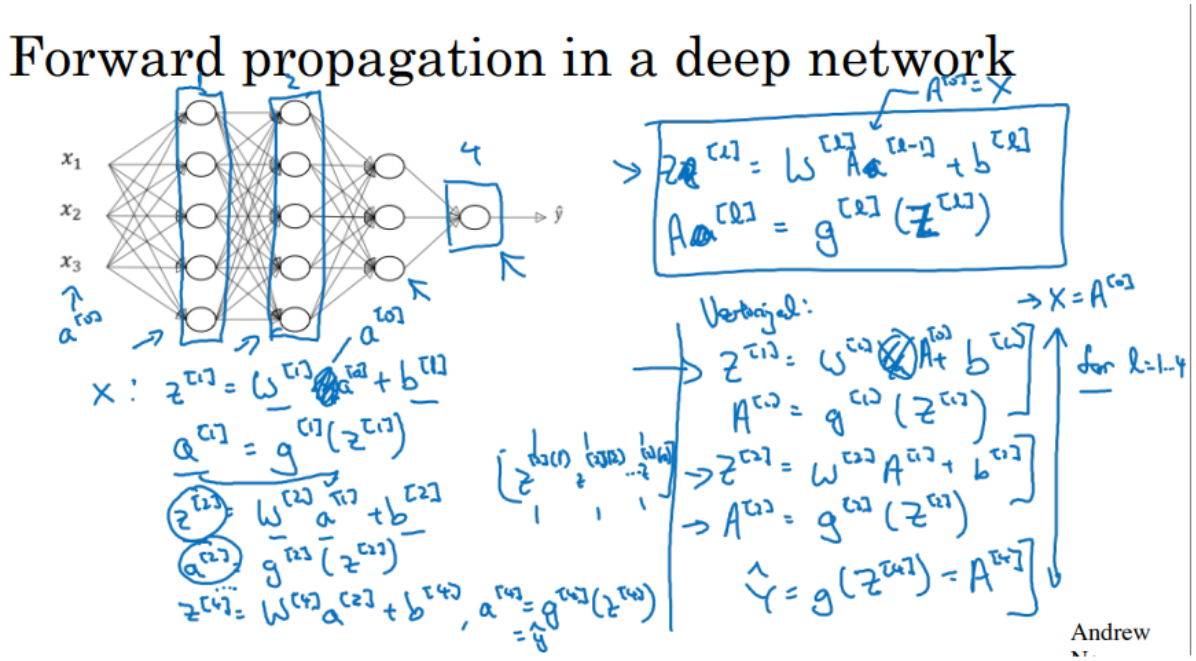


3. 심층 신경망에서의 정방향전파

1) 단일 학습 데이터 x에 대한 정방향 전파

- 첫 번째 층에 대한 **z^[1] = W^[1]*x + b^[1] = W^[1]*a^[0] + b^[1]**
- 첫 번째 층에 대한 **활성화 계산 : a^[1] = g(z^[1])**
- 활성화 함수 g는 층마다 다름.
- 두 번째 층에 대한 **z^[2]= W^[2]*a^[1] + b^[2]**
- 활성화: 가중치 행렬 * 층1의 출력 + 층2의 편향 벡터
  a^[2] = g^[2](z^[2])
- 마지막 층: z^[4]= W^[4]*a^[3] + b^[4]
- 활성화: **a^[4] = g^[4](z^[4])= yhat**

→ z^[l] = W^[l]*a^[l-1] + b^[l]
→ a^[l] = g^[l](z^[l])



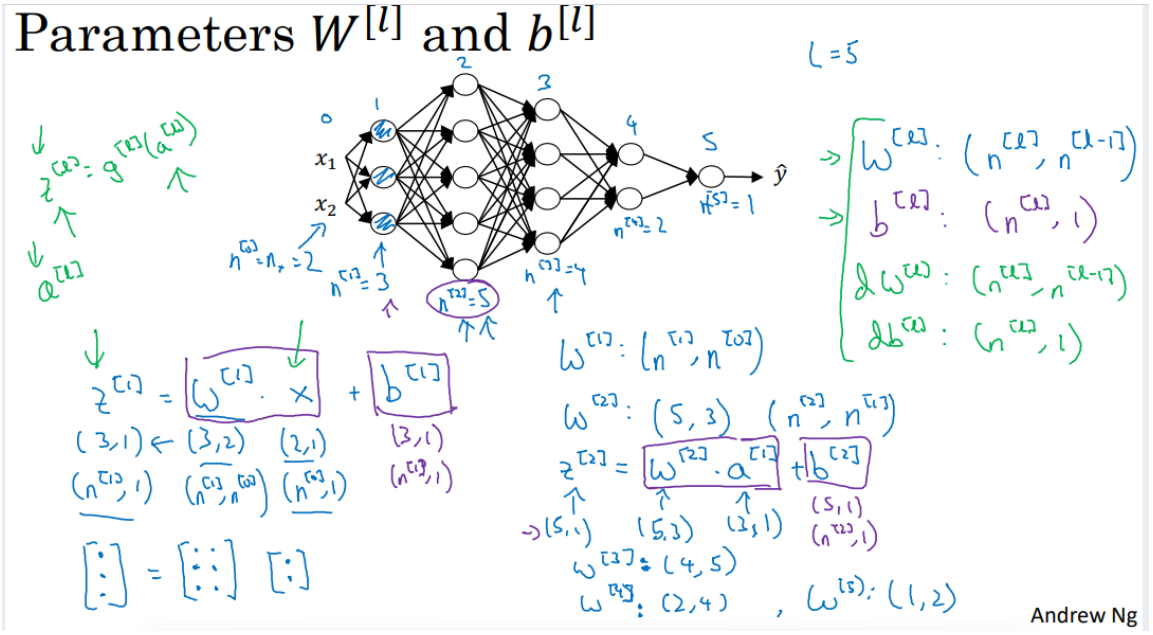
2) 전체 학습 세트에 대한 벡터화

- Z^[1] = w^[1]X + b^[1] , X=A^[0]
- A^[1] = g^[1](Z^[1])
- X, Z, A: 열벡터 형태
- Y hat = g(Z^[4]) = A^[4]

→ 반복문 필요: 1에서 L까지의 l에 대해, 각 층에 대한 활성화 계산

4. 행렬의 차원을 알맞게 만들기

- **정방향 전파 구현**



- **n^[0] = n_x, n^[1] = 3, n^[2] = 5, n^[3] = 4, n^[4] = 2, n^[5] = 1**
- 1단계: z^[1] = W^[1] * x + b^[1]
  - z: 첫 번째 은닉층에 대한 활성화 벡터, (3,1) 벡터
    → (n^[1], 1) 차원 벡터
  - x: (2,1) 행렬
    → (n^[0], 1) 차원 벡터
  - W: (3,2) 행렬
    → (n^[1], n^[0]) 차원 벡터
    → **W^[l]의 차원: (n^[l], n^[l-1])**
  - b: (3,1) 벡터
- 2단계: z^[2] = W^[2] * a^[1] + b^[2]
  - z^[2]: (5,1) 차원
  - a^[1]: (3,1) 차원
  - W^[2]: (5,3) 차원
  - b^[2]: (5,1) 벡터
    → **b^[l]의 차원: (n^[1], 1)**
- **dW^[l] 차원 = W^[l] 차원**
- **db^[l] 차원 = b^[l] 차원**
- 벡터화된 구현

- Z^[1] = W^[1] * X + b^[1]
- m이 훈련 집합의 크기라면:
  - Z^[1]의 차원 = (n^[1], m)
  - W^[1]의 차원 = (n^[1], n^[0])
  - X의 차원 = (n^[0], m) → 모든 학습 데이터가 수평으로 저장되기 때문
  - b^[1]의 차원 = (n^[1], 1) → 브로드캐스팅으로 (n^[1], m) 차원이 됨.

- **z^[l], a^[l]의 차원: (n^[l], 1)**
- **Z^[l], A^[l]의 차원: (n^[l], m)**
- **dZ^[l], dA^[l]의 차원: (n^[l], m)**

5. 왜 심층 신경망이 더 많은 특징을 잡아 낼 수 있을까요?

- 심층망 계산이란?
  1) 얼굴 인식, 감지 시스템 구축 시
  1. 얼굴 사진을 입력값으로 넣고
  2. 첫 번째 층: 특성 탐지기 or 모서리 탐지기 → 사진을 보고 모서리가 어디에 있는지 파악
  3. 두 번째 층: 픽셀을 그룹화하여 모서리 형성

  → 감지된 모서리와 그룹화된 모서리를 받아, 얼굴의 일부 형성 가능

  4. 세 번째 층: 서로 다른 얼굴의 일부를 최종적으로 모아서 얼굴 일부 감지 가능

  → 서로 다른 종류의 얼굴 감지할 수 있게 됨.

직관 1: 네트워크가 더 깊어 질 수록, 더 많은 특징을 잡아낼 수 있다. 낮은 층에서는 간단한 특징을 찾아내고, 깊은 층에서는 탐지된 간단한 것들을 함께 모아 복잡한 특징을 찾아냄.

2) 음성 인식 시스템 구축

1. 음성을 입력으로 줄 때
2. 첫 번째 층: 낮은 단계의 음성 파형 특징 탐지
3. 두 번째 층: 소리의 기본 단위 탐지를 학습(음소)
4. 세 번째 층: 음성의 단어 인식

→ 학습하여 단어를 구성하여 구나 문장 인식 가능해짐.

- 얕은 네트워크로 같은 함수 계산 → 기하급수적으로 많은 은닉 유닛 필요

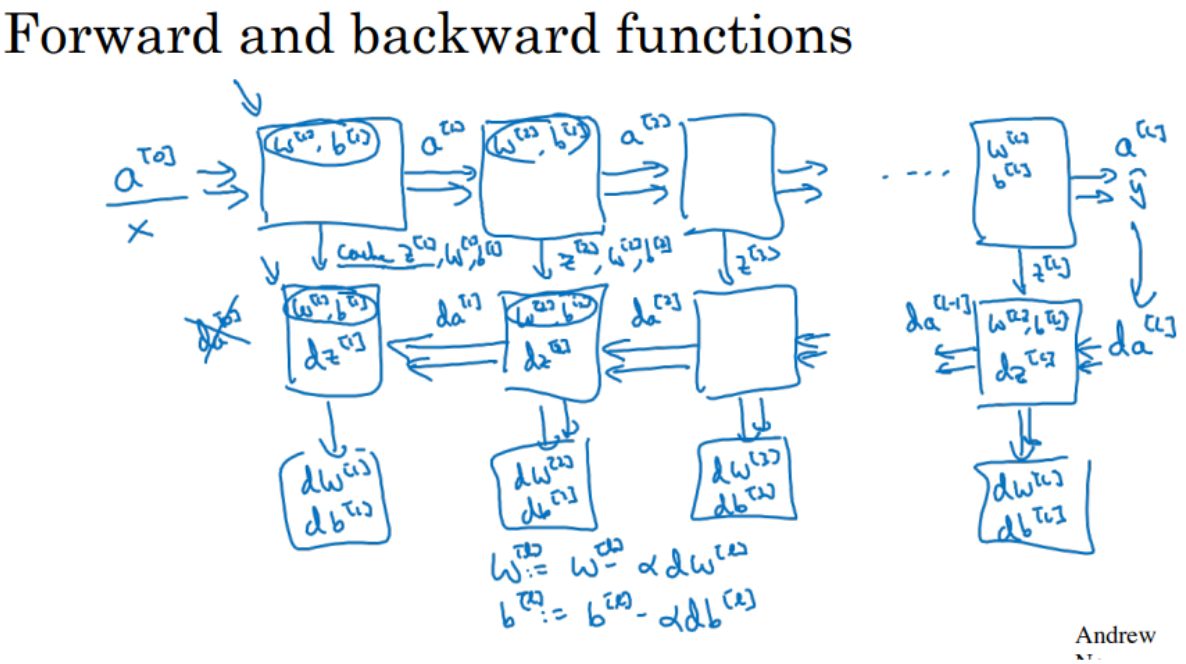
직관 2: 순환 이론에 따르면, 상대적으로 은닉층의 개수가 작지만 깊은 심층 신경망에서 계산할 수 있는 함수가 있다. 그러나 얕은 네트워크로 같은 함수를 계산하려고 하면, 즉 충분한 은닉층이 없다면 기하급수적으로 많은 은닉 유닛이 계산에 필요하게 될 것

6. 심층 신경망 네트워크 구성하기

- 층에 대한 계산: 층 L에 대해 매개변수 W^[l]과 b^[l]이 있다.
  1) 정방향 전파
  - Input: a^[l-1] → 이전 층의 활성값
  - Output: a^[l]
  - z^[l] = W^[l]*a^[l-1] + b^[l]
  - a^[l] = g^[l](z^[l])
  - cache z^[l]: z^[l]을 저장해두기
    → 역전파 단계에서 유용하게 쓰임

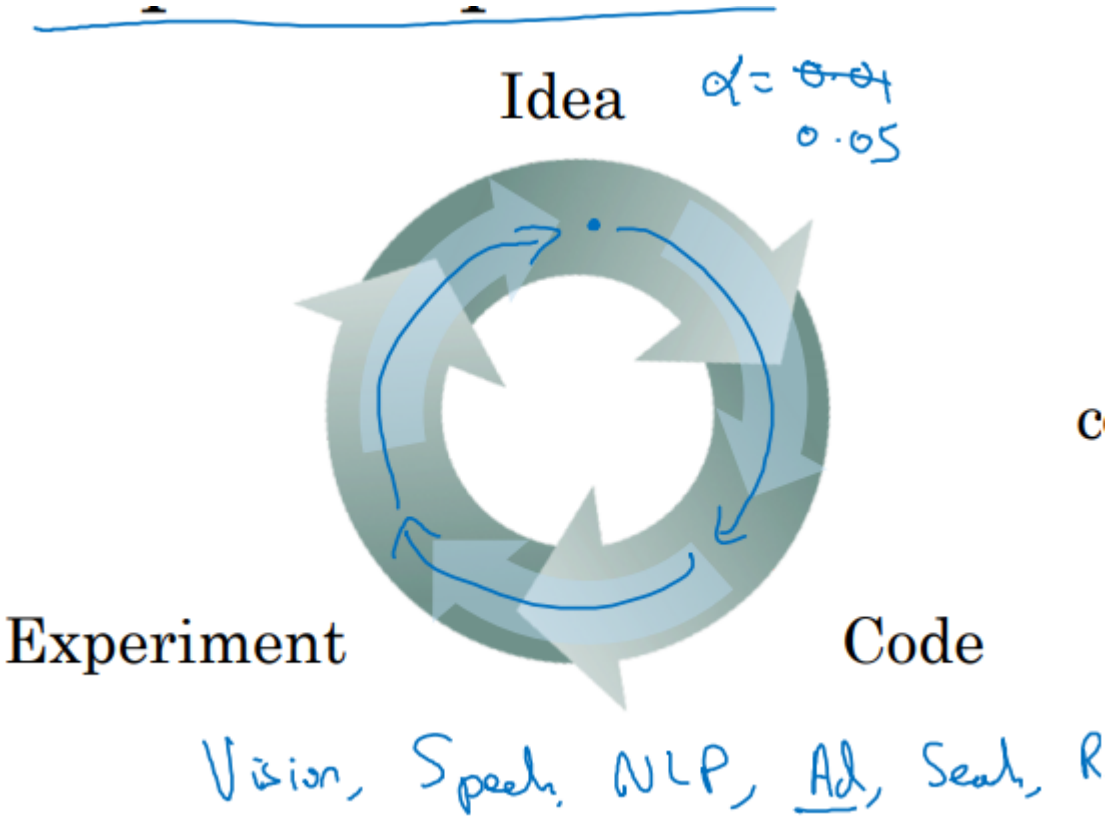
  2) 역전파 단계

  - Input: da^[l] ← 저장해놓은 ^[l] 이용
  - Output: da^[l-1], 경사하강법 학습을 위한 그래디언트
  - da^[0]: 입력 특성에 대응하는 도함수로, 지도 신경망의 가중치를 학습할 때에는 유용X



7. 변수 vs 하이퍼파라미터

- **Parameters: W^[1], b^[1], W^[2], b^[2],..**
- **Hyperparameters: learning rate alpha, 경사하강법의 반복 횟수, 은닉층의 개수 (L), 은닉 유닛의 개수(n^[1],..) , 활성화 함수의 선택(ReLU, TanH, Sigmoid..)**
  → W, b를 통제하는 매개변수
- 새로운 application 시작 시, 하이퍼파라미터의 가장 적합한 값을 미리 아는 것은 어렵.
  → 다양한 값을 시도해봐야 함. + process 거치기



- 하이퍼하라미터에 몇 가지 값을 시도하고, 더 좋은 값이 있는지 이중 확인해보는 것이 필요하다!

8. 인간의 뇌와 어떤 연관이 있을까요?

- 신경망과 인간의 뇌 간의 관계는 크지 않다. 다만, 신경망의 복잡한 과정을 단순화해서 뇌세포의 프로세스로 비유하게 되면, 사람들에게 조금 더 직관적이고, 효과적으로 전달 가능
- 신경 과학자들조차도 하나의 뉴런이 무엇을 하는지 거의 모릅니다. 우리가 신경과학에서 특징짓는 것보다 하나의 뉴런은 훨씬 더 복잡하고 알기 어렵습니다. 게다가 뉴런이 신경망 처럼 역전파를 통해서 학습 하는지도 의문입니다.
- 최근에는 이런 비유가 점점 무너져 가고 있습니다.