[Week9]_문가을

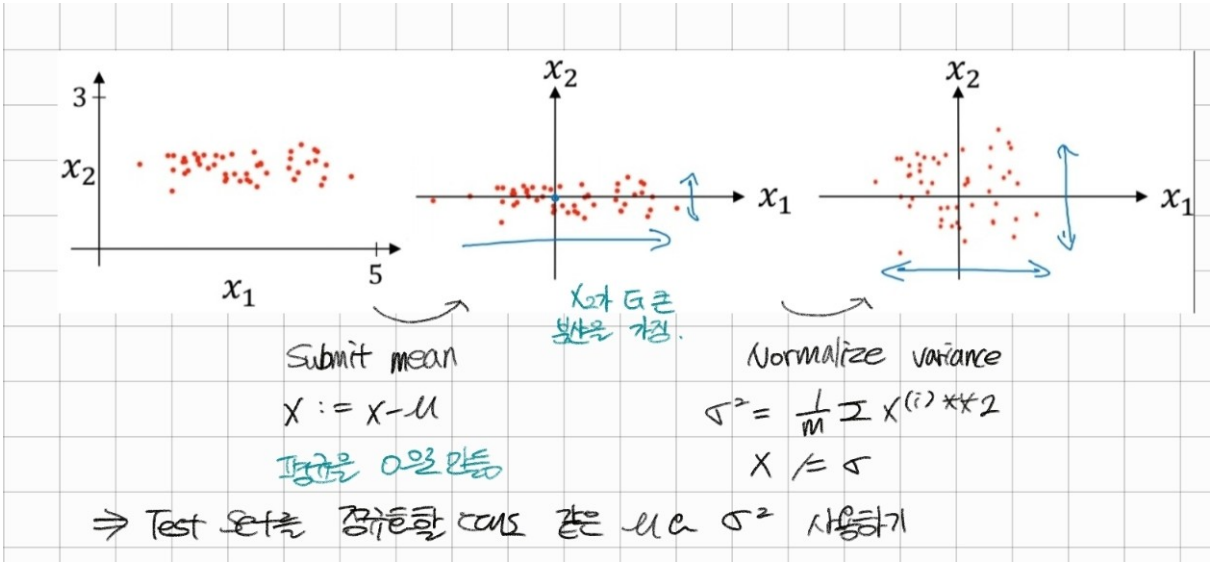
딥러닝 2단계 : 심층 신경망 성능 향상시키기

3. 최적화 문제 설정

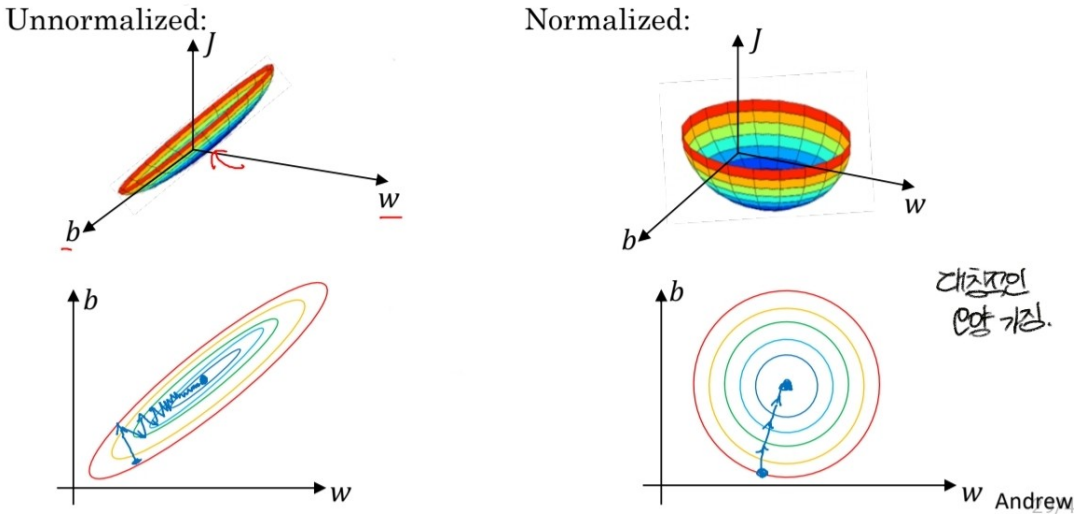
입력값의 정규화

신경망의 훈련을 빠르게 하기 위해 입력값을 정규화함.

정규화 방법



비용함수 비교



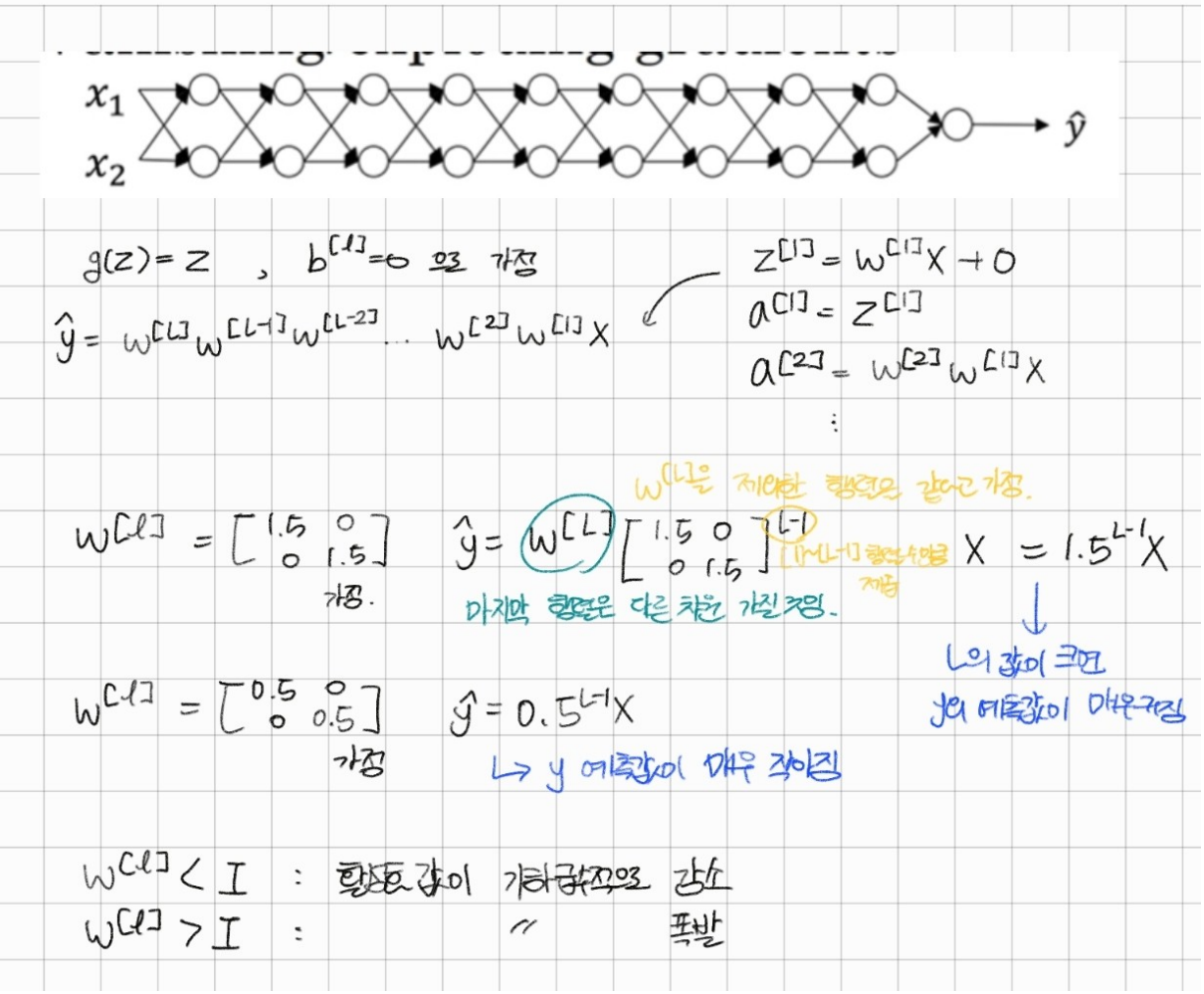
- 정규화하지 않은 데이터셋으로 훈련

→ 특성 값의 범위가 달라 매개 변수 값들이 매우 다를 것임. (그림에선 w와 b로 표현함.)

→ 경사하강법에서 매우 작은 학습률을 사용하게 됨. 앞뒤로 왔다갔다 하기 위해 많은 단계가 필요하기 때문.

경사 소실(vanishing gradients) / 경사 폭발(exploding gradients)

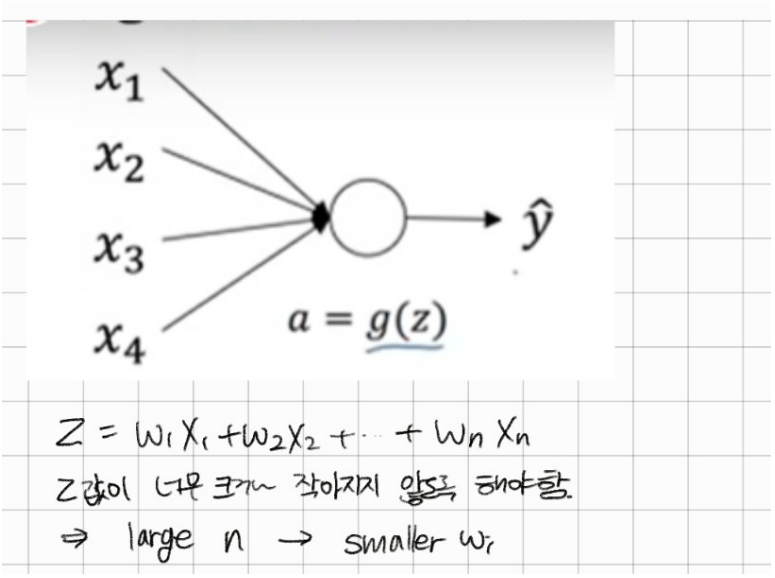
미분값이 매우 작아지거나 매우 커지는 문제.

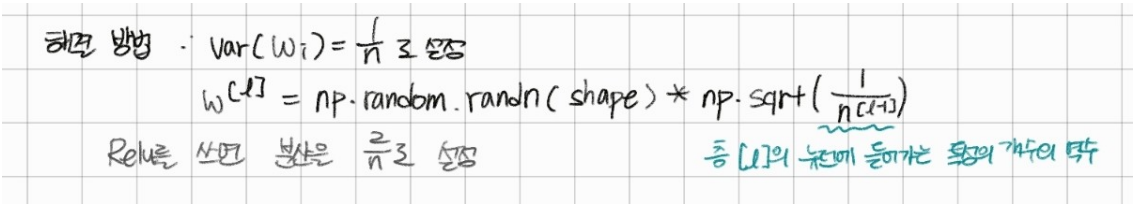


같은 원리로 미분값이 소실/ 폭발하는 원리를 설명할 수 있음.

해결하기 위해서 가중치를 어떻게 초기화하는 지가 중요해짐.

심층 신경망의 가중치 초기화





ReLU 활성화 함수를 사용하는 경우 w_i 의 분산을 $\frac{2}{n^{[l-1]}}$ 으로 설정합니다.

tanh 활성화 함수를 사용하는 경우 w_i 의 분산을 $\frac{1}{n^{[l-1]}}$ 또는 $\frac{2}{n^{[l-1]} + n^{[l]}}$ 으로 설정합니다.

→ 입력 특성 혹은 활성값의 평균이 대략 0이고 표준편차가 1이 됨.

→ z 값이 비슷한 크기를 갖게 됨.

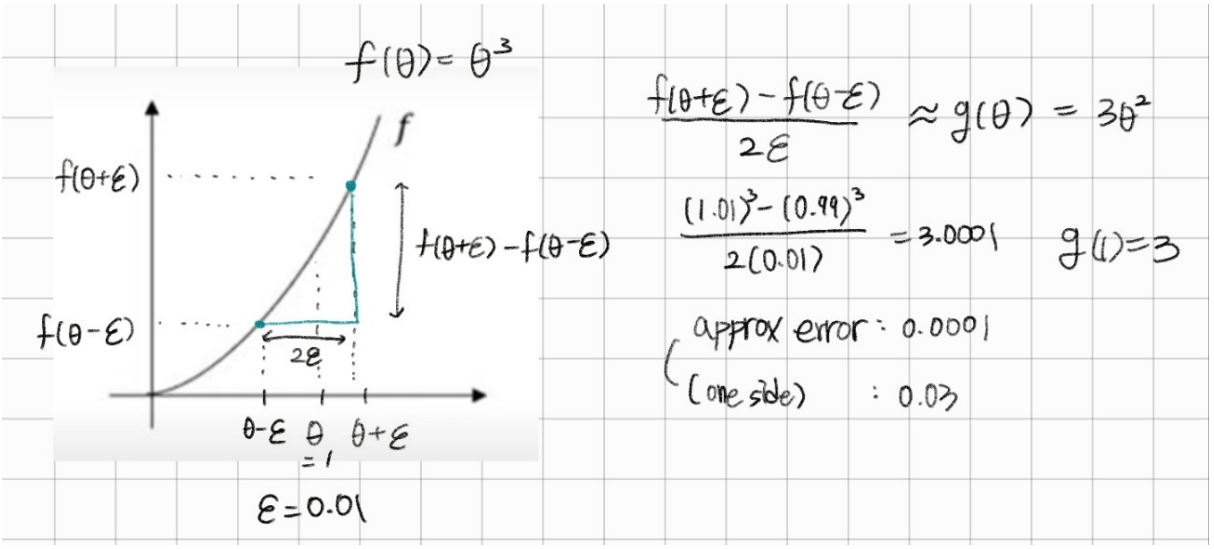
→ 경사 소실과 폭발 문제에 도움을 줌.

기울기의 수치 근사

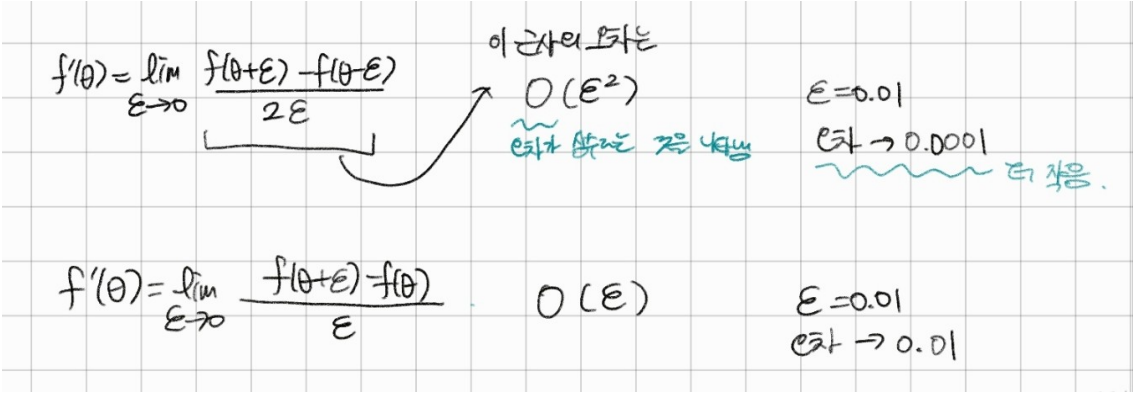
역전파를 맞게 구현했는지 확인하기 위해 경사 검사를 진행함.

→ 먼저 경사의 계산을 수치적으로 근사하는 방법이 필요함.

미분값 계산



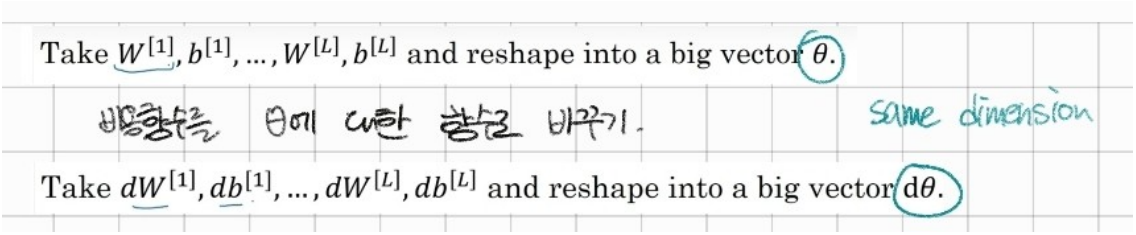
왜 일반적으로 하는 미분 계산과 다른 방법을 사용하는가?



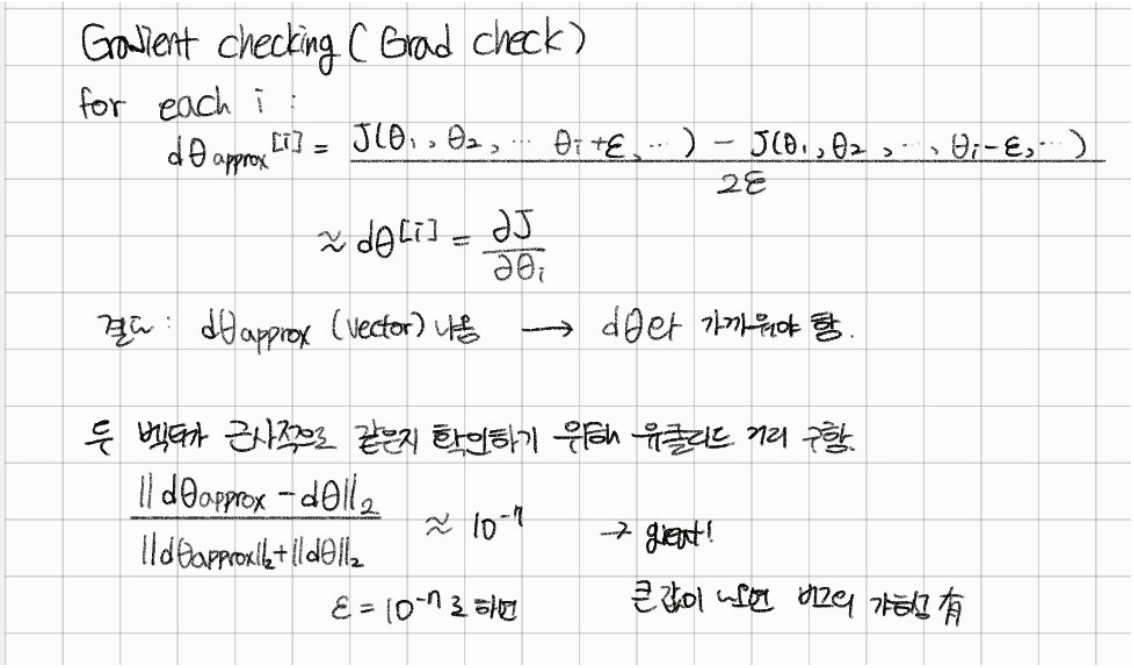
→ 오차가 작아짐.

경사 검사

경사 검사를 하는 방법



근사적인 미분값을 구하고 미분값과 비슷한지 확인하는 과정 거침.



경사 검사 시 주의할 점

1. 디버깅을 위해서만 경사 검사를 사용하기

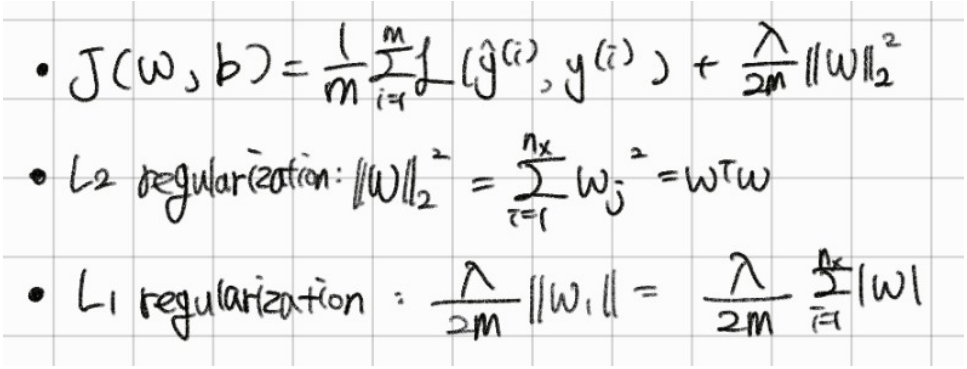
→ 근사 미분값을 구하는 데에 시간이 오래 걸리기 때문임.

2. 경사 검사의 알고리즘이 실패하면 (근사 미분값과 미분값의 차이가 크면)

→ 개별적인 컴포넌트를 확인해 버그를 확인하기.

→ 각각 i에 대해 dtheta(approx)[i]와 dtheta[i] 확인

3. 비용함수에 정규화 term이 있다는 것을 기억하기



4. 드롭아웃에서는 경사 검사가 작동하지 않는다.

→ 드롭아웃에서 비용함수를 계산하기 어렵기 때문임.

5. 무작위적 초기화에서 w와 b가 0에 가까울 때 경사 검사가 잘 되는 경우

→ 훈련을 조금 시켜 w와 b가 0에서 멀어지게 한 다음 경사 검사를 다시 해보기