



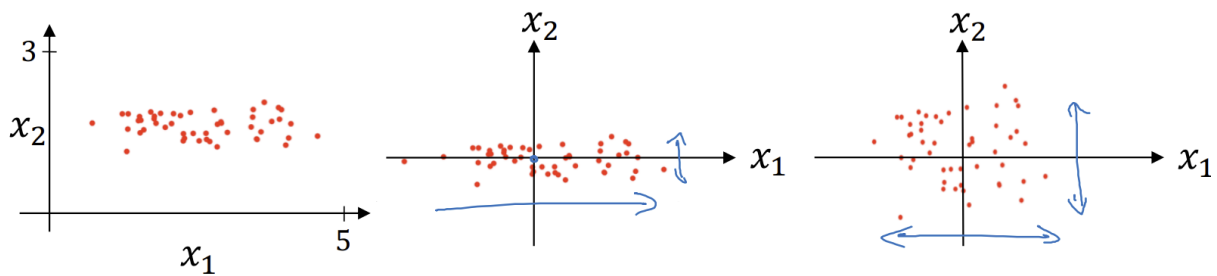
9주차 과제

📌 3-1. 입력값의 정규화

: 신경망의 훈련을 빠르게 만드는 기법 중 하나 $Z = \frac{X - \mu}{\sigma}$

기존의 데이터 값에서 평균을 빼고 그를 표준편차로 나눔

→ 그렇게 만든 데이터에선 평균이 0, 분산이 1이 됨.



학습 데이터와 테스트 데이터의 평균, 분산이 다르다면?

→ 대부분의 경우 그러함, 테스트 세트를 정규화할 때 테스트 세트의 평균 분산을 사용하는 게 아닌 학습 데이터의 평균 분산을 이용하는 것이 맞다.

[정규화하는 이유]

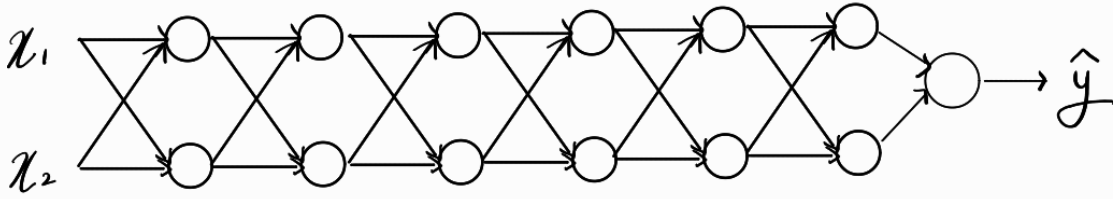
정규화 하지 않은 경우와 정규화한 경우의 비용함수 형태를 보면 각각 비대칭적, 대칭적 구조를 이루는 것을 확인 가능.

→ 비용함수가 최소가 되는 지점을 찾기 위해서 이동

→ 대칭적 구조를 이루면 왔다리 갔다리 할 필요 없이 금방 최솟값 지점을 찾기가 가능!

📌 3-2. 경사소실/경사폭발

: 신경망 훈련 시 기울기가 아주 작아지거나 매우 커지는 문제



activation function $g(z) = z$ (선형 함수)

$$\Rightarrow \hat{y} = w^{[1]} \cdot w^{[2]} \cdot \dots \cdot w^{[l-1]} \cdot w^{[l]} \cdot x$$

$$\Rightarrow \begin{cases} w > 1 \Rightarrow \hat{y} = w^l \cdot x \rightarrow \infty \Rightarrow \text{경사 폭발} \\ w < 1 \Rightarrow \hat{y} = w^l \cdot x \rightarrow 0 \Rightarrow \text{경사 소실} \end{cases}$$

📌 3-3. 심층 신경망의 가중치 초기화

: 위에서 살펴본 경사소실, 경사폭발 문제를 약간 해결할 수 있는 방안,
출력값이 너무 크게 되지 않게끔 n 이 커짐에 따라 w 들의 값을 작게 만들어야함
→ 합리적 방법 : w 의 분산을 $1/n$ 으로 만드는 것!

어떤 활성화 함수를 사용하는 지에 따라 w 의 분산을 다르게 만듦 (n : # of inputs)

- ReLU $\rightarrow \sigma^2(w_i) = \frac{2}{n^{[l-1]}}$
- tanh $\rightarrow \sigma^2(w_i) = \frac{1}{n^{[l-1]}}$ or $\frac{2}{n^{[l-1]} + n^{[l]}}$

📌 3-4. 기울기의 수치 근사

경사 검사: 역전파 맞게 구현했는지 확인용

$f'(x) = \frac{f(\theta + \epsilon) - f(\theta - \epsilon)}{2\epsilon}$ 와 같이 도함수를 구하면 한 쪽 방향으로 구한 도함수에 비해 미분값 오차가 더 작다! 더 정확하다! (ϵ = 아주아주 작은 수)

~~크렁지만 시간이 두 배는 더 걸리겠지...~~

📌 3-5. 경사 검사

기존의 W, b 를 모두 벡터 θ 로 변환 $\rightarrow J(W, b) = J(\theta)$, 위의 3-4에서 한 바와 같이 미

$$\rightarrow d\theta_{approx}^{[i]} = \frac{J(\theta_1, \theta_2, \dots, \theta_i + \epsilon, \dots) - J(\theta_1, \theta_2, \dots, \theta_i - \epsilon, \dots)}{2\epsilon} \approx d\theta^{[i]}$$

그래서 결론 적으로 이 두 벡터가 근사한지를 확인!

아래의 식 값이 10^{-7} 보다 작다면 더 거의 근사하다고 할 수 있음!

$$\rightarrow \frac{\|d\theta_{approx} - d\theta\|_2}{\|d\theta_{approx}\|_2 + \|d\theta\|_2}$$

3-6. 경사 검사 시 주의할 점

- 학습 시엔 사용 X, 디버깅 시에만 사용
- 경사 검사(grad check)에 실패시 어느 지점에서 실패했는 지 확인
- 정규화를 잊지 말기

비용함수 정의 =

$$J(w^{[1]}, b^{[1]}, \dots, w^{[L]}, b^{[L]}) = \frac{1}{m} \sum_{i=1}^m L(\hat{y}^{(i)}, y^{(i)}) + \frac{\lambda}{2m} \sum_{l=1}^L \|w^{[l]}\|_F^2$$

→ 여기서 뒤에 붙는

$\frac{\lambda}{2m} \sum_{l=1}^L \|w^{[l]}\|_F^2$ 도 잊지 말고 포함해서 계산해야함!

- 경사 검사는 드롭아웃에서는 작동하지 X
모든 반복마다 은닉 유닛의 서로 다른 부분집합을 무작위로 삭제하기 때문.
→ 드롭아웃을 끄고 검사 후 드롭아웃을 켜다!
- 가끔! 랜덤 초기화를 했을 때, W, b가 0에 가까울 때 경사하강법의 구현이 맞은 경우
→ 경사하강법 진행 시 W, b가 점차 증가! 점점 더 튼튼!
→ 랜덤 초기화에서 경사 검사를 진행