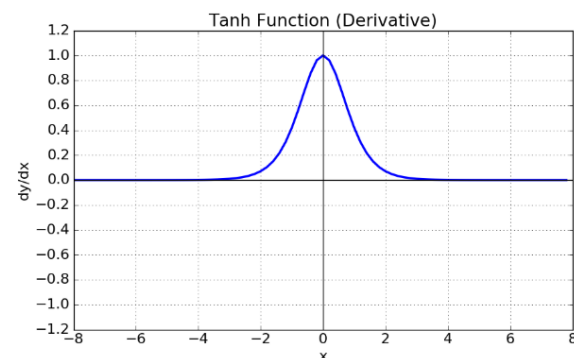
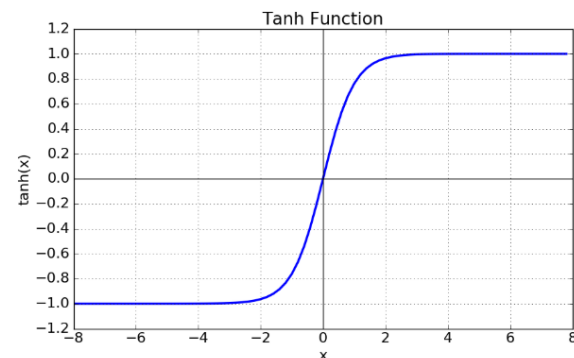
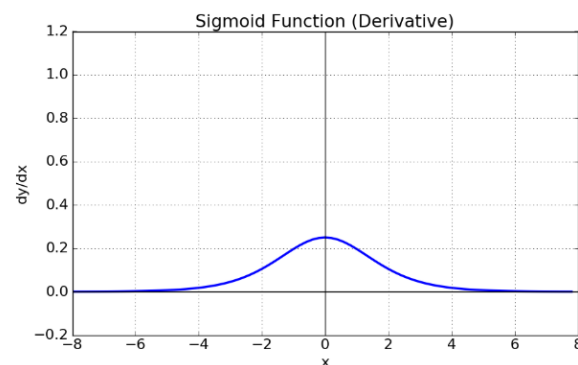
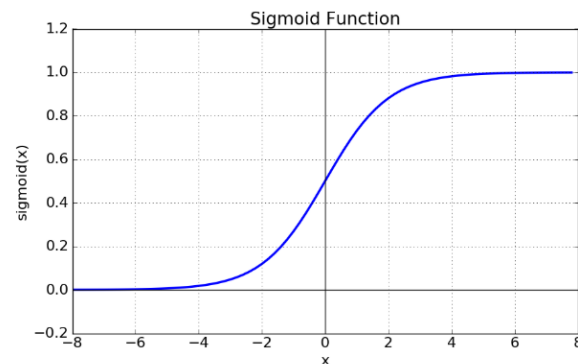


2주차 question

1. Tanh이 평균이 0이라서 좋다는데 평균이 0인게 왜 좋나
2. ReLU가 진짜 좋은 이유는? 도함수가 1,0이 나오는 것이 왜 좋나
3. Sigmoid를 출력층에 쓰면 좋은 이유
4. 비편향 추정이 필요없다는 뜻이 무엇인가?
5. 모델을 압축(L1 정규화)한다는 뜻은 무엇인가?
6. W가 희소해진다는 뜻(L1정규화), 왜 희소해짐
7. 은닉 유닛의 영향력을 작게 하면 왜 과대적합을 막음?
8. 신경망을 반복을 하면 할수록 왜 w가 커짐

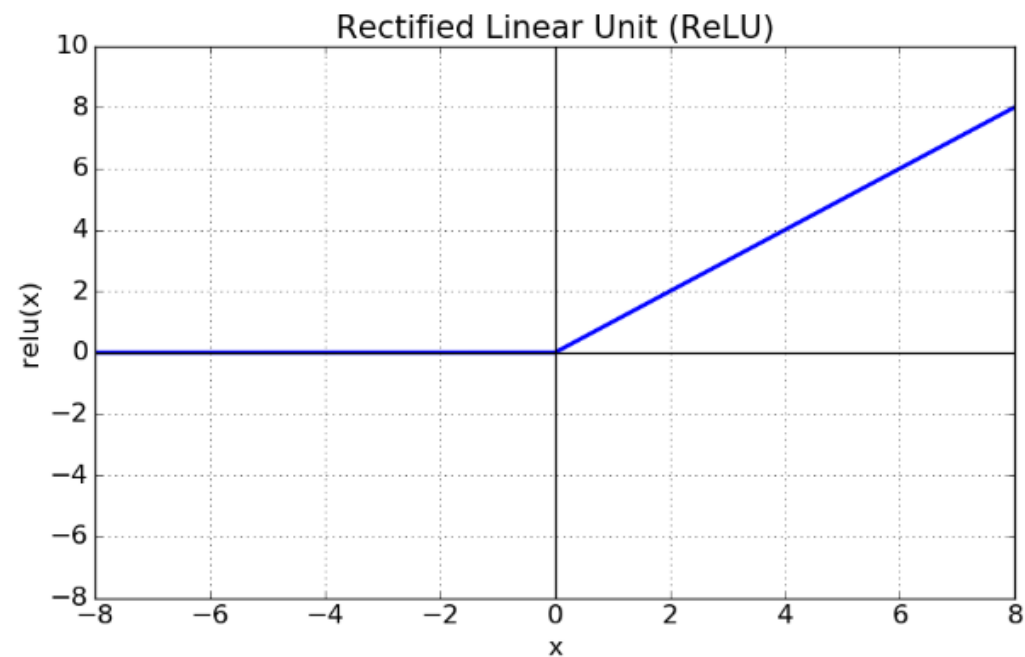
1. Tanh이 평균이 0이라서 좋다는데 평균이 0인게 왜 좋나 → 스터디 때 내린 결론은 틀린듯 밑에 글을 봐주셈

- 함수값 중심이 0이 아니다. 함수값 중심이 0이 아니라 학습이 느려질 수 있다. 그 이유를 알아보면, 만약 모든 x 값들이 같은 부호(ex. for all x is positive) 라고 가정하고 아래의 파라미터 w 에 대한 미분함수식을 살펴보자. $\frac{\partial L}{\partial w} = \frac{\partial L}{\partial a} \frac{\partial a}{\partial w}$ 그리고 $\frac{\partial a}{\partial w} = x$ 이기 때문에, $\frac{\partial L}{\partial w} = \frac{\partial L}{\partial a} x$ 이다. 위 식에서 모든 x 가 양수라면 결국 $\frac{\partial L}{\partial w}$ 는 $\frac{\partial L}{\partial a}$ 부호에 의해 결정된다. 따라서 한 노드에 대해 모든 파라미터 w 의 미분값은 모두 같은 부호를 갖게된다. 따라서 같은 방향으로 update되는데 이러한 과정은 학습을 zigzag 형태로 만들어 느리게 만드는 원인이 된다.



2. ReLU가 진짜 좋은 이유는? 도함수가 1,0이 나오는 것이 왜 좋나

- 0보다 작으면 도함수가 0이고
즉, 꺼버리고 0보다 크면 도함수가 1이다
즉 꺾다 꺾다 하는 것인데
0보다 작다는 것은 $w x + b$ 가
작은 것인데, 이는 영향력이
작으므로 꺼버리고, 1보다 크면
영향력이 크므로 키는 것이다
또한 양극단의 도함수를 0에 가깝게 만들지
않기 때문에 기울기 소실 문제가 없다.

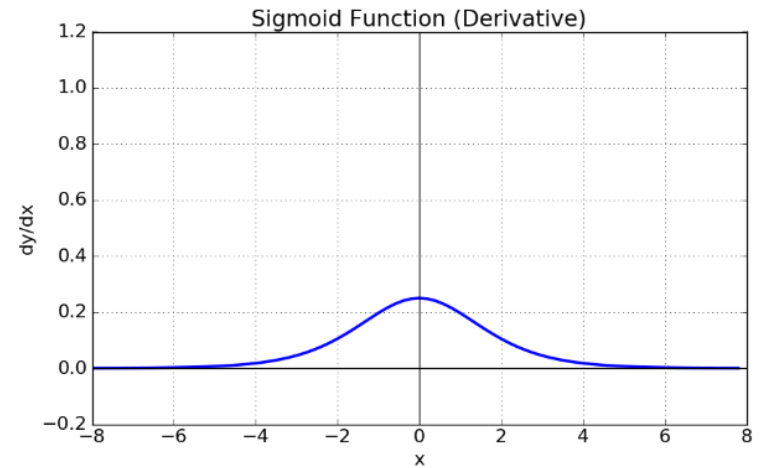
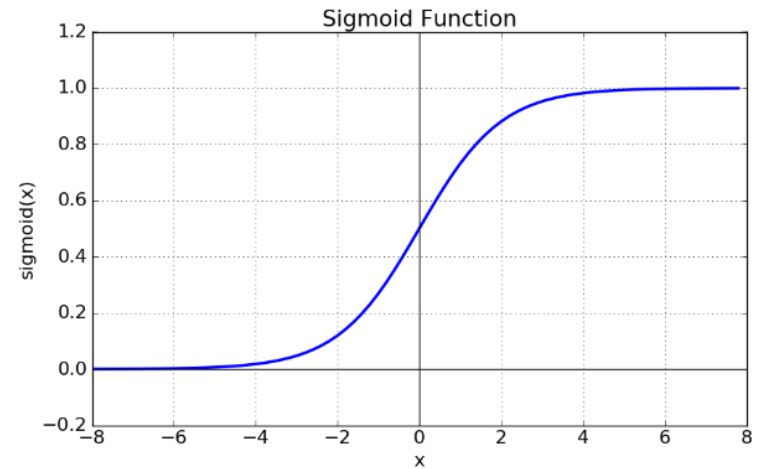


ReLU함수의 특징을 살펴보자.

- $x > 0$ 이면 기울기가 1인 직선이고, $x < 0$ 이면 함수값이 0이된다.
- sigmoid, tanh 함수와 비교시 학습이 훨씬 빨라진다.
- 연산 비용이 크지않고, 구현이 매우 간단하다.
- $x < 0$ 인 값들에 대해서는 기울기가 0이기 때문에 뉴런이 죽을 수 있는 단점이 존재한다.

3. Sigmoid를 출력층에 쓰면 좋은 이유

- 0과 1사이를 가지기 때문에
Binary한 문제를 정의하기 쉽다



3. 비편향 추정이 필요없다는 뜻이 무엇인가?

- 스터디에서 결론이 안남
- 모르겠음
- 그냥 test set 씹시다

3. 모델을 압축(L1 정규화)한다는 뜻은 무엇인가?

4. W가 희소해진다는 뜻(L1정규화), 왜 희소해짐

- 필요 없는 피쳐의 영향력을 줄인다는 뜻이다!
- W가 희소해진다는 것은 행렬에 0이 많다는 뜻, 그것이 피쳐의 영향력을 줄이는 것임

- 7. 은닉 유닛의 영향력을 작게 하면 왜 과대적합을 막음?
- 8. 신경망을 반복을 하면 할수록 왜 w 가 커짐

- 은닉 유닛의 영향력이 각각 커지면, 이상치나 왜곡된 피처에 영향을 많이 받아 과적합이 일어날 수 있는데, 그러한 피처에 관한 은닉 유닛의 영향력을 줄이면 과대적합을 막는다
- W 는 피처가 영향력이 높으면 커짐, 그래서 신경망 반복을 하면 영향력 있는 피처의 w 가 커짐

Q1) 정규화 방법

로그 씌우는 것도 있던데 이것도 정규화 방법??
표준정규분포로 정규화 하는거랑 차이점

- 정규화(regulation)와 nomilzation과 혼동한 듯

- Dw 와 w 의 차원이 왜 같나
- 행렬은 미분해도 차원이 같다함