

SEMICOLON

기울기 소실

데이터 사이언스와 고급 머신러닝 딥러닝



그래디언트 소실

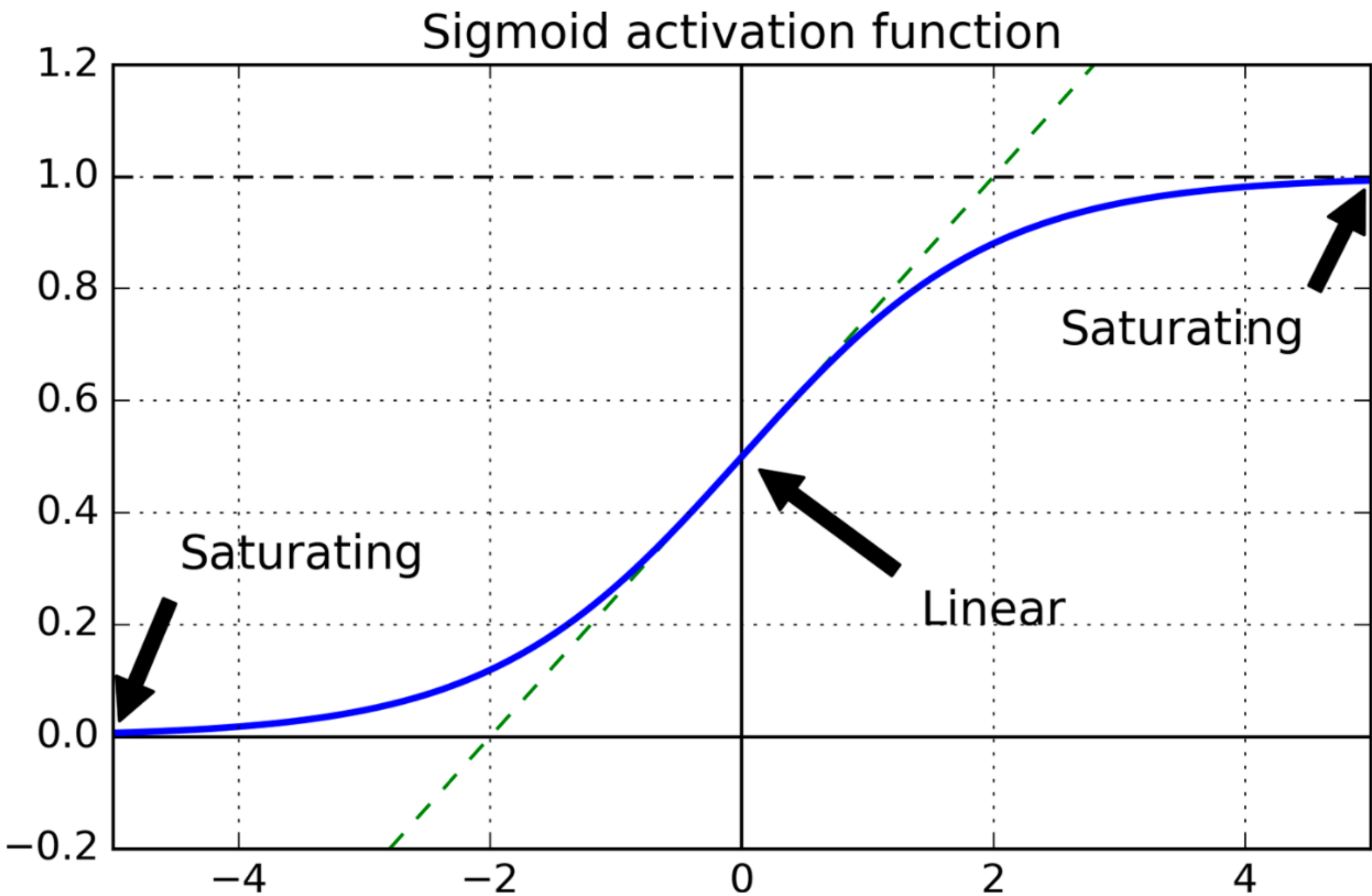
“

알고리즘이 하위 층으로 진행될수록 그래디언트가 점점 작아져서 가중치가 업데이트 되지 않는 문제

”

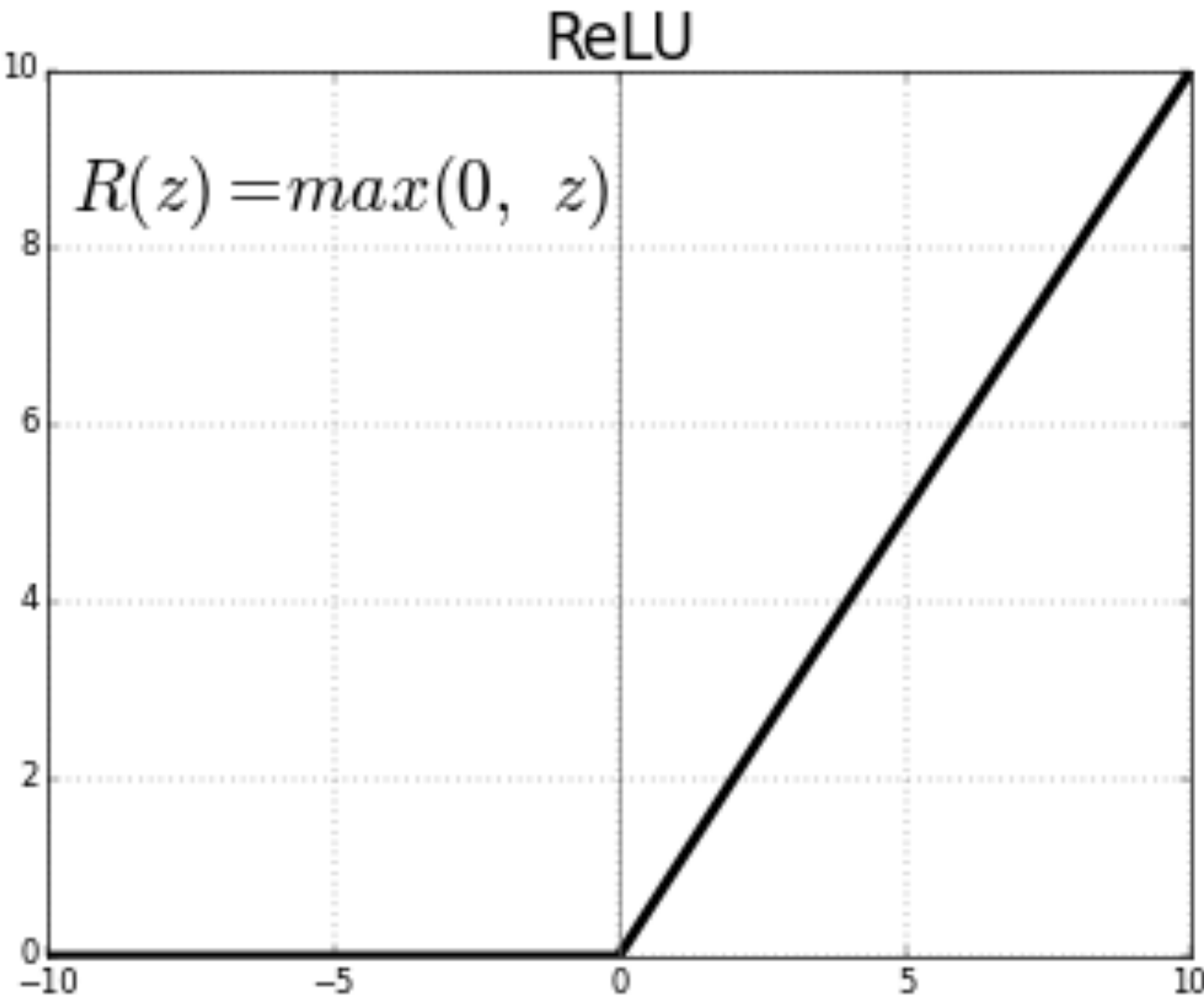
- 역전파 알고리즘에서 출력층에서 입력층 방향으로 이동하며 그래디언트를 계산하는데 미분이 0에 가까워짐

시그모이드



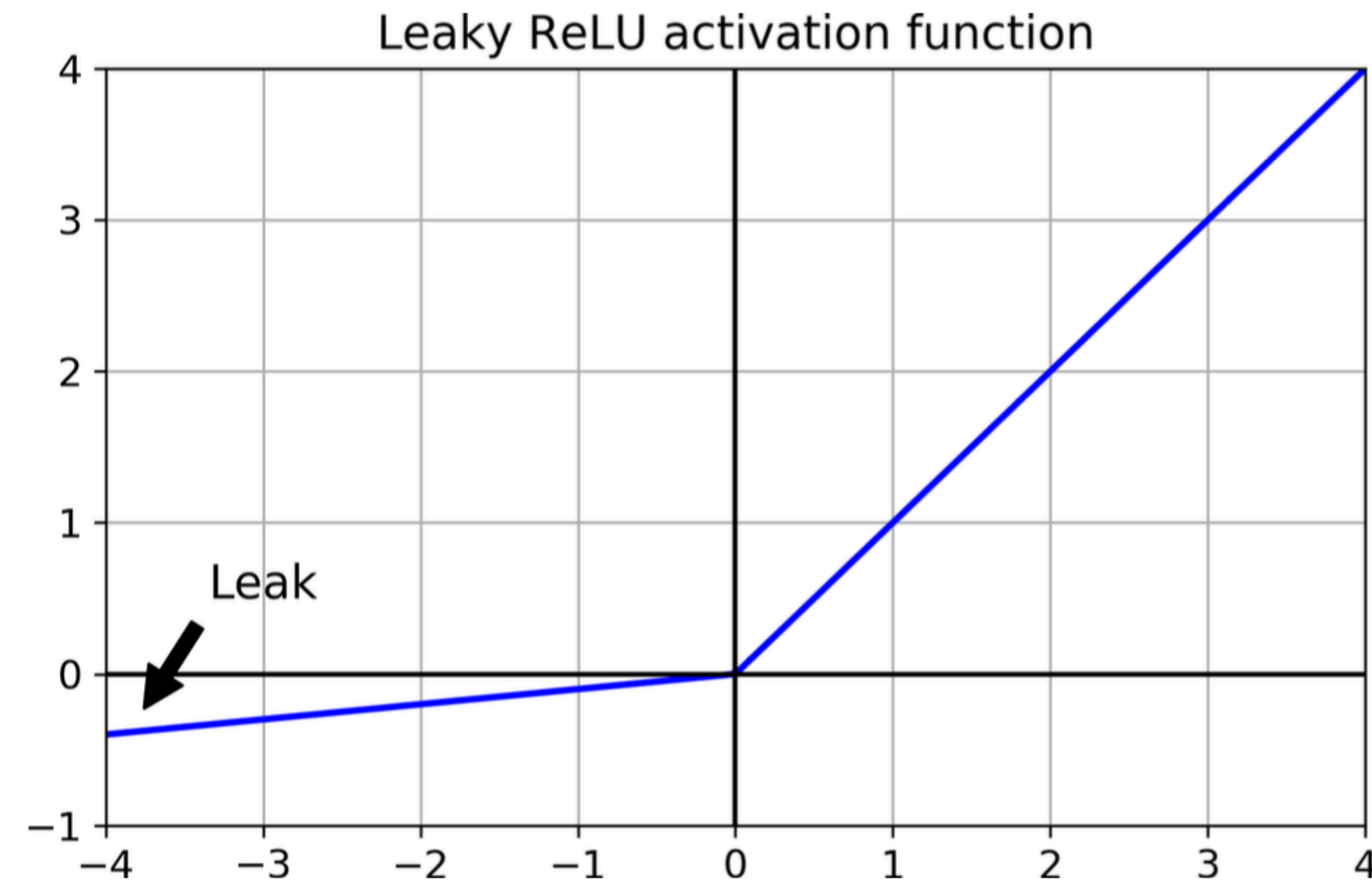
평균이 0이고 표준 편차가 1인 정규분포를 이용해서 가중치 초기화를 할 경우
층을 지날 때마다 분산이 커져 활성화 함수가 0이나 1로 수렴하는데 기울기가 0에 매우 가까워집니다
이 경우에는 다른 가중치 초기화 방법을 선택합니다

ReLU



훈련하는 동안 일부 뉴런이 0 이외의 값을 출력하지 않는다는 의미에서 죽은 ReLU 문제가 발생합니다
가중치 합이 음수이면 ReLU 함수의 그래디언트가 0이 되므로 경사 하강법이 작동하지 않습니다

LeakyReLU



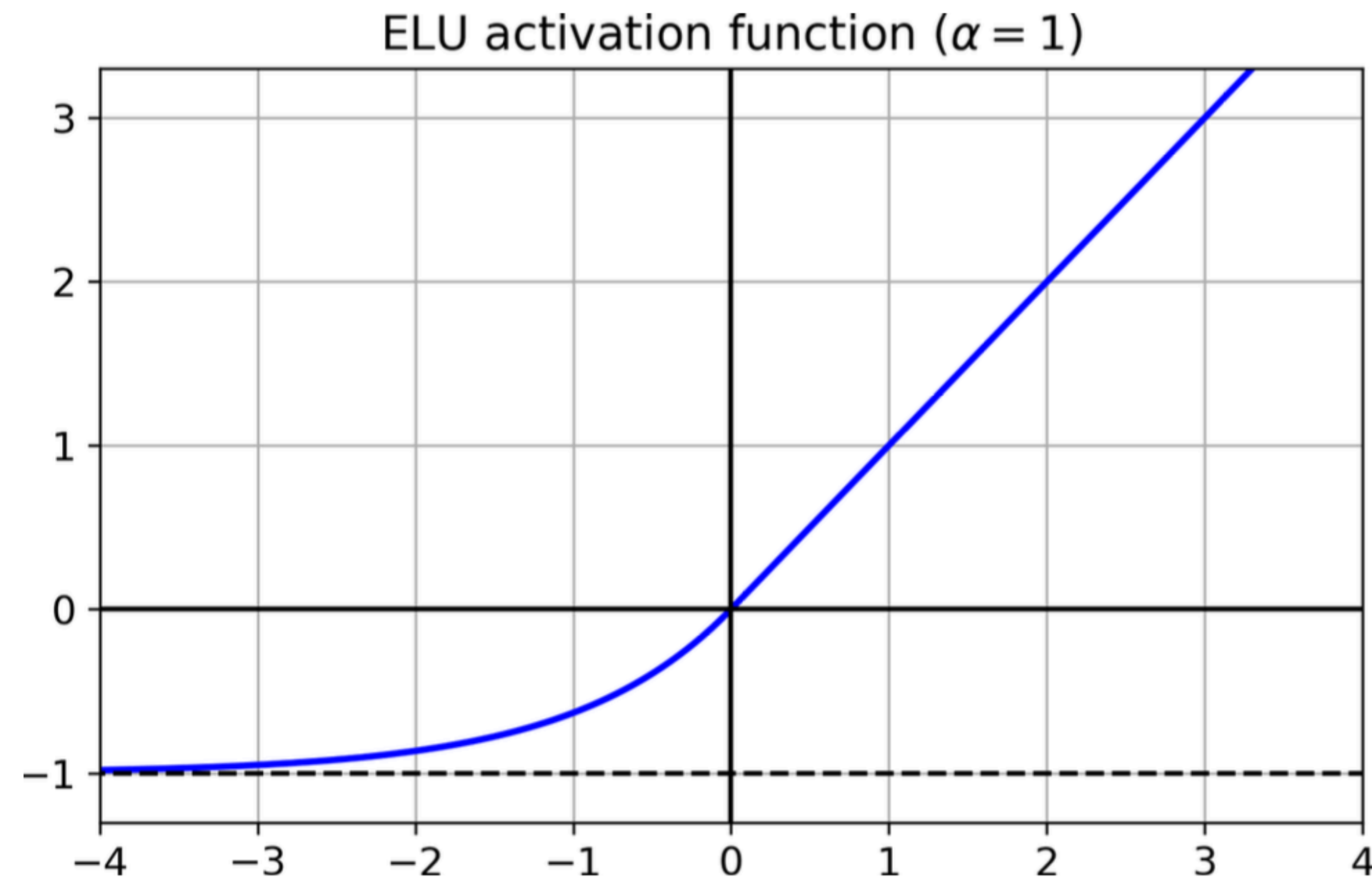
$$\text{LeakyReLU}_a(z) = \max(az, z)$$

하이퍼파라미터 a 가 이 함수가 새는 정도를 결정합니다

$z < 0$ 일 때 기울기가 LeakyReLU를 절대 죽지 않게 만듭니다

40. 기울기 소실 | 활성화 함수

ELU



$$ELU_a(z) = \begin{cases} \alpha(\exp(z) - 1) & z < 0 \\ z & z \geq 0 \end{cases}$$

ELU 활성화 함수는 ReLU 보다 훈련 시간이 줄고 신경망의 테스트 성능도 더 높았습니다
 $z < 0$ 이어도 그래디언트가 0이 아니므로 죽은 뉴런을 만들지 않습니다