

경사하강법

전설

Loss function = cost function

sum of squared errors



$$J(w) = \frac{1}{2} \sum_{i=1}^n (y^{(i)} - \phi(z^{(i)}))^2$$

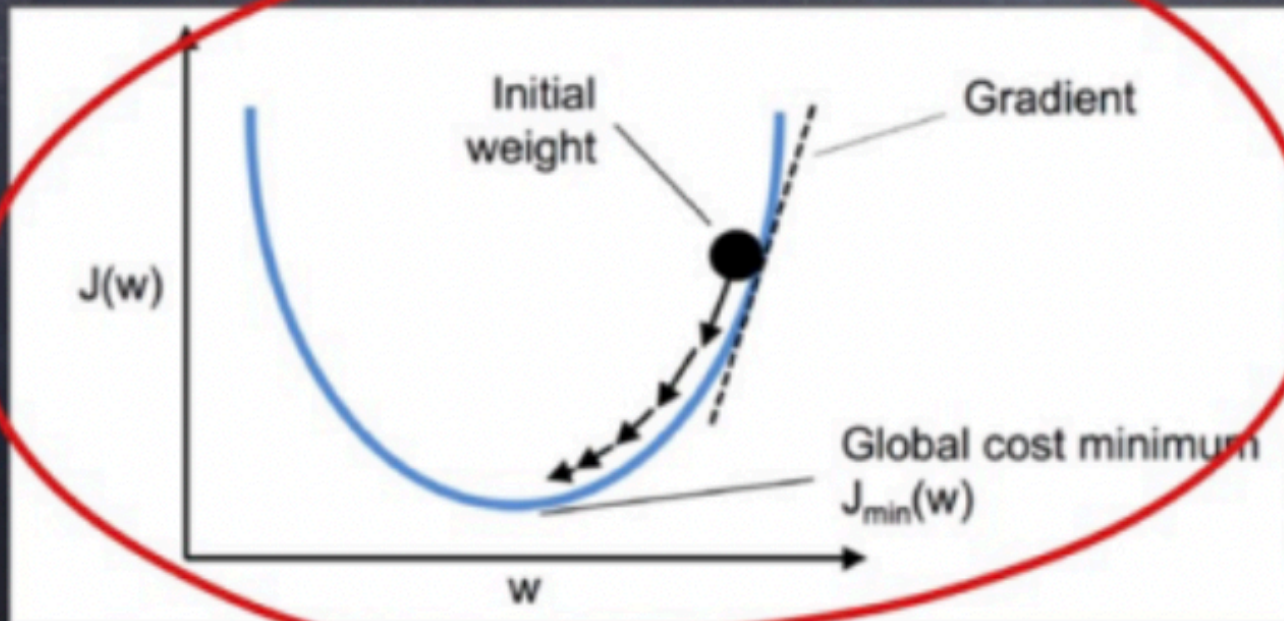
Global optimum

가설 함수 $H(w, b)$

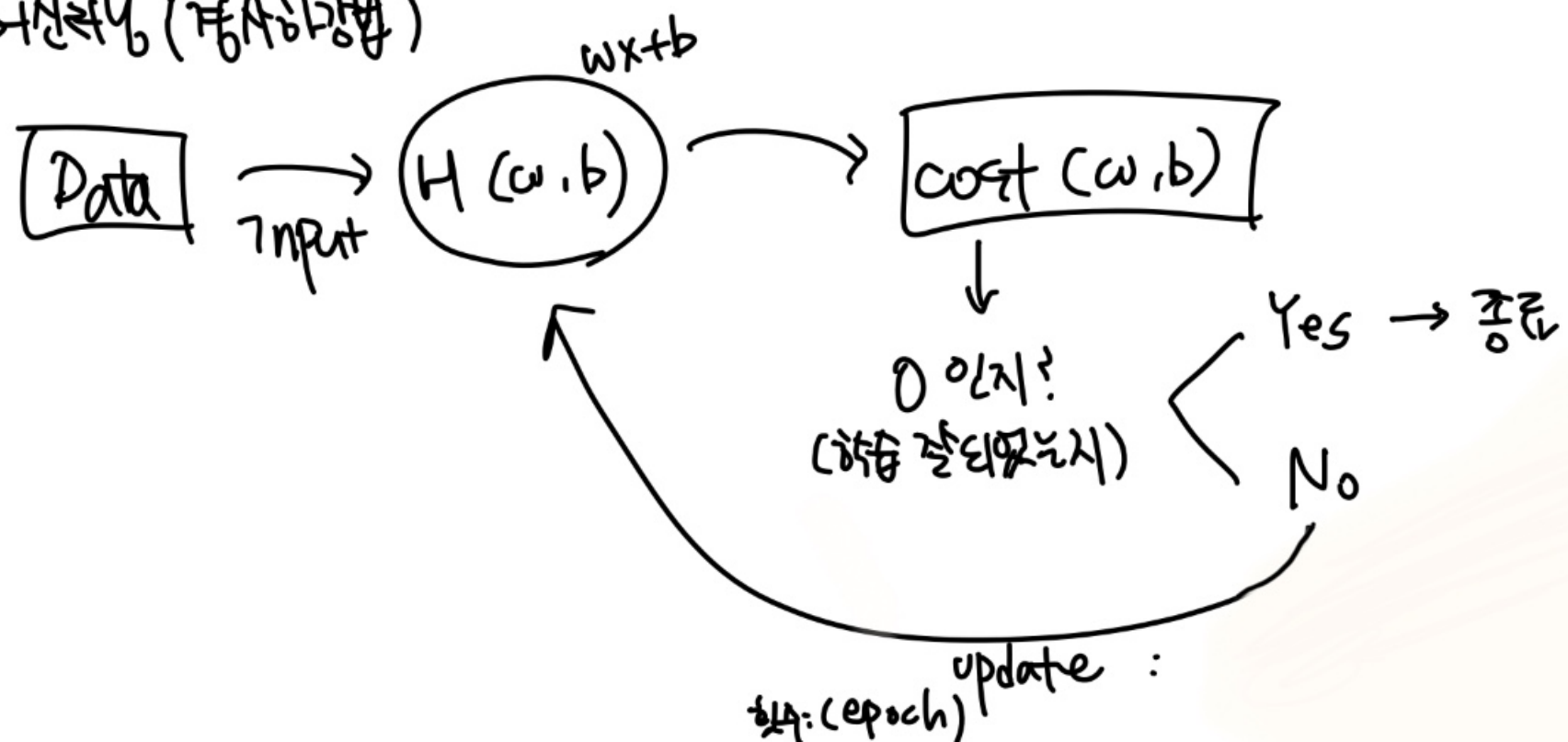
$$H = w \cdot x + b$$



$$w := w + \Delta w$$

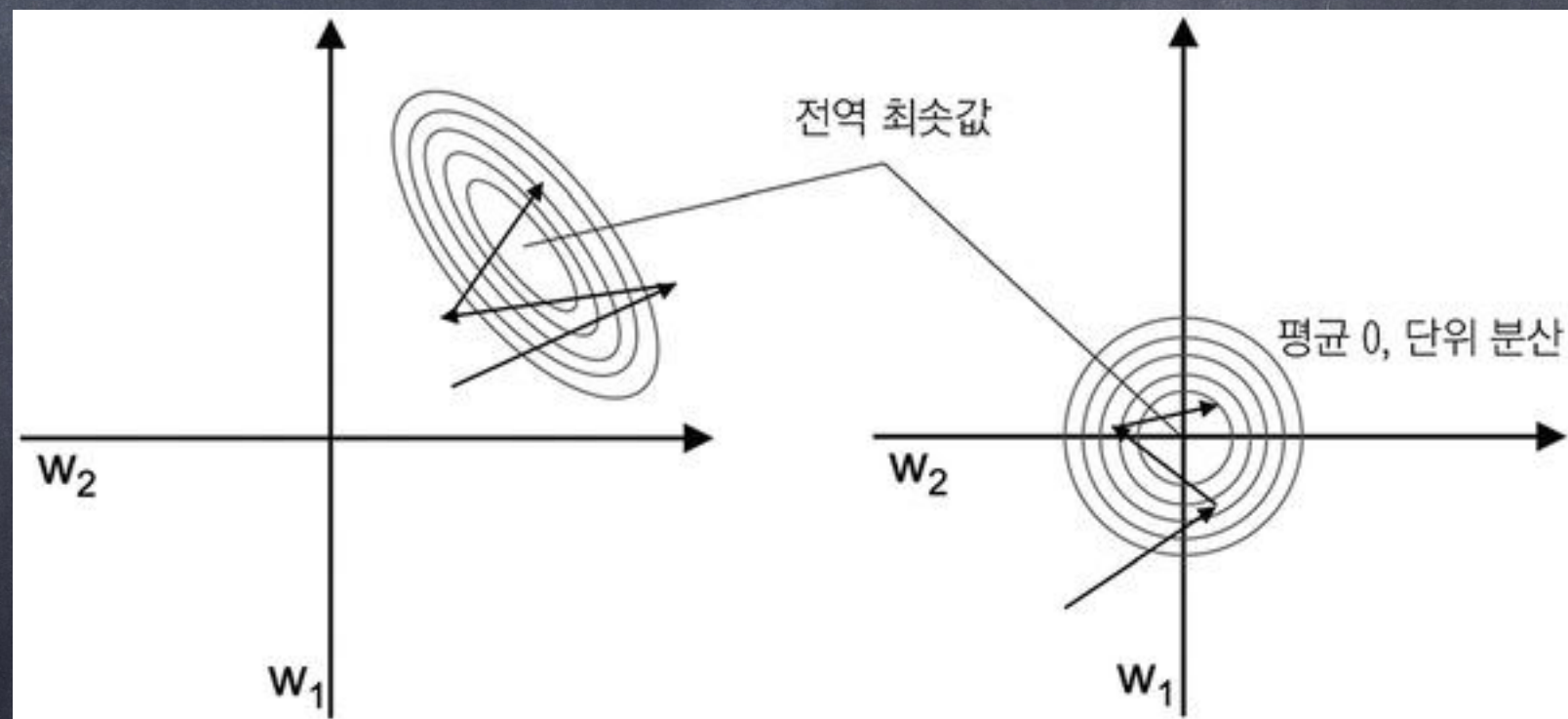


머신러닝 (통계학방식)



standardization

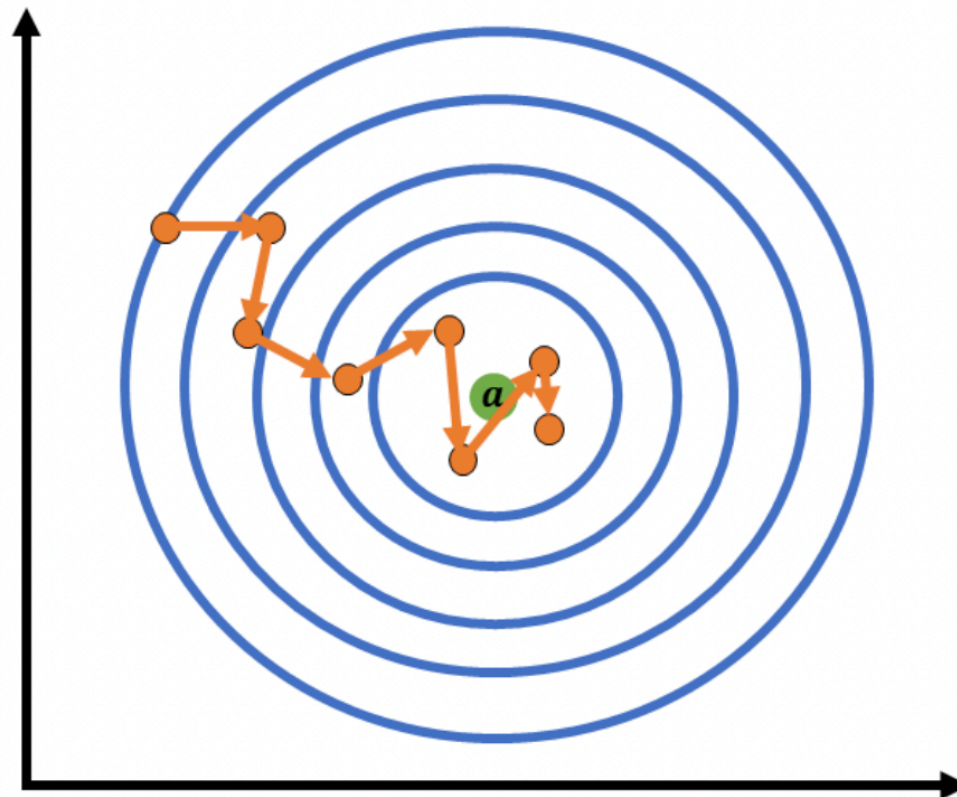
$$x'_j = \frac{x_j - \mu_j}{\sigma_j}$$



Stochastic gradient descent

2. 확률적 경사 하강법(Stochastic Gradient Descent, SGD)

- 전체 훈련 데이터셋을 대상으로 학습하는 것은 한정된 리소스를 가지고 있는 우리의 분석 환경에서 매우 비효율적이며, 파라미터 업데이트 수가 적다는 것은 랜덤 하게 뽑힌 시작 위치의 가중치 수도 적으므로, Local minimum 현상이 발생할 확률도 높다는 것이다.
- 그래서 나온 방법이 학습 데이터셋에서 무작위로 한 개의 샘플 데이터 셋을 추출하고, 그 샘플에 대해서만 기울기를 계산하는 것이다.
- **샘플 데이터 셋에 대해서만 경사(Gradient)를 계산**하므로, 매 반복에서 다뤄야 할 데이터 수가 매우 적어, 학습 속도가 매우 빠르다.
- 하나의 샘플만 대상으로 경사를 계산하므로, **메모리 소모량이 매우 낮으며**, 매우 큰 훈련 데이터 셋이라 할지라도 학습 가능하다.
- 그러나, 무작위로 추출된 샘플에 대해서 경사를 구하므로, 배치 경사 하강법보다 훨씬 불안정하게 움직인다.



Thank you