

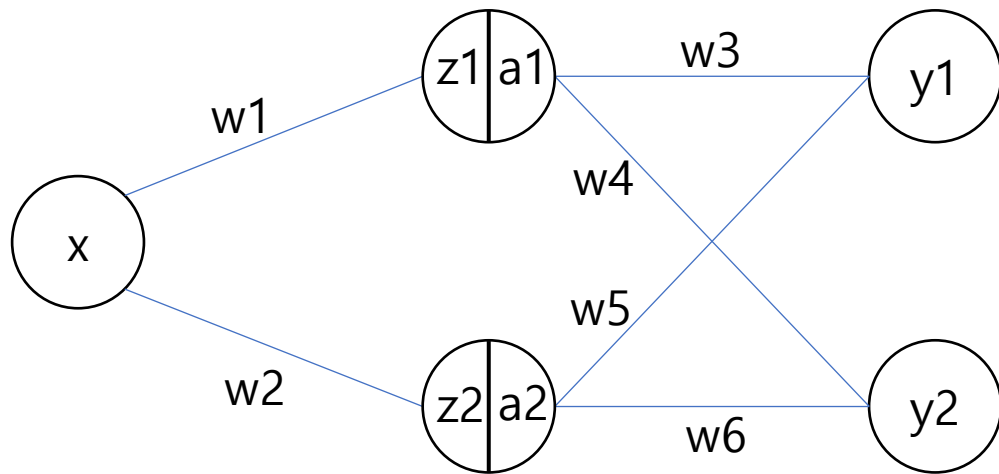
딥알못 탈출하기 #3

고려대학교 지능시스템 연구실

유민형 제작

Stochastic Gradient Descent Algorithm

- 지난번과 같은 신경망과 데이터가 있을 때, Gradient Descent Algorithm으로 weight를 3번 업데이트 하는 경우를 생각해보자.
- 이번에는 데이터가 여러 개다. Loss는 MSE, Lr=0.1



$$f(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{else} \end{cases}$$

$$z1 = w1 * x + b1$$

$$z2 = w2 * x + b2$$

$$a1 = f(z1)$$

$$a2 = f(z2)$$

$$y1 = w3 * a1 + w5 * a2$$

$$y2 = w4 * a1 + w6 * a2$$

Weight 초깃값

$$w1 = 0.1$$

$$w2 = -0.2$$

$$w3 = 0.4$$

$$w4 = -0.3$$

$$w5 = -0.15$$

$$w6 = 0.12$$

$$b1 = 0.05$$

$$b2 = -0.03$$

데이터쌍 (x, y1, y2)
 : (0.5, 1, -1)
 (0.3, 1.4, -0.7)
 (-0.2, -0.8, 1.2)

Stochastic Gradient Descent Algorithm

- (문제1) Numpy로 다음 3가지 경우의 Gradient Descent를 해보자.
- 각 경우의 loss를 기록하고 비교하시오.

- 방법 1

업데이트1:

(0.5, 1, -1) 사용

업데이트2:

(0.3,1.4, -0.7) 사용

업데이트3:

(-0.2, -0.8, 1.2) 사용

$$L = \frac{1}{N} \frac{1}{m} \sum_{i=1}^m (\hat{y}_i - y_i)^2$$

- 방법2

업데이트1:

(0.5, 1, -1)

(0.3,1.4, -0.7) 두 개 사용

업데이트2:

(0.3,1.4, -0.7)

(-0.2, -0.8, 1.2) 두 개 사용

업데이트3:

(-0.2, -0.8, 1.2)

(0.5, 1, -1) 두 개 사용

- 방법3

(0.5, 1, -1)

(0.3,1.4, -0.7)

(-0.2, -0.8, 1.2) 세 개 사용

업데이트2:

(0.5, 1, -1)

(0.3,1.4, -0.7)

(-0.2, -0.8, 1.2) 세 개 사용

업데이트3:

(0.5, 1, -1)

(0.3,1.4, -0.7)

(-0.2, -0.8, 1.2) 세 개 사용

Stochastic Gradient Descent Algorithm

- (문제2) 앞에서 실험한 각각의 경우의 장단점과 특징을 통계적
인 관점에서 설명하고, 실제 학습할 시 어떤 것이 가장 적절한지
설명하시오. (힌트 Deeplearning(I. Goodfellow 책 SGD파트)
- 1번처럼 데이터를 한 개씩 업데이트하는 경우
- 2번처럼 일부를 묶어서 평균적으로 업데이트하는 경우
- 3번처럼 전체 데이터로 계속 업데이트 하는 경우