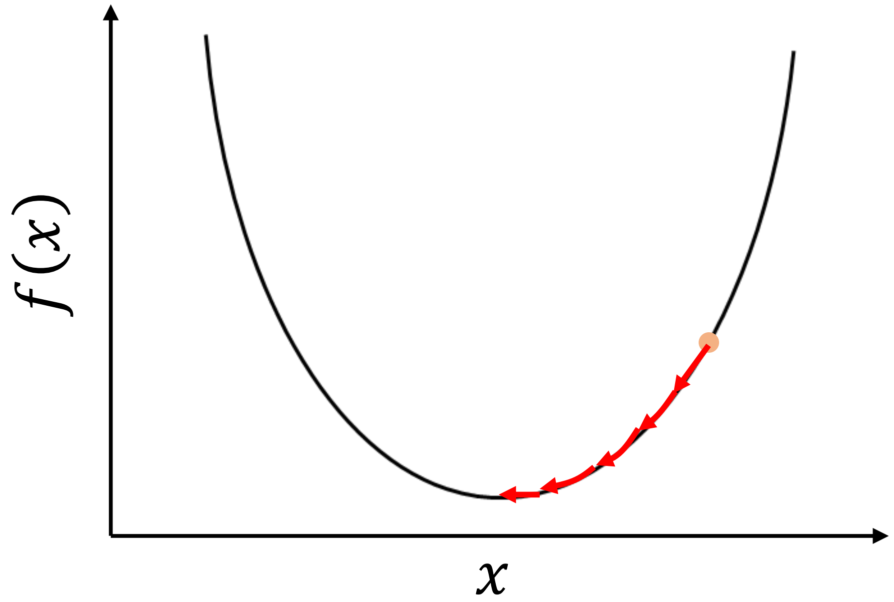
Chapter 03. 쉽게 배우는 경사 하강 학습법

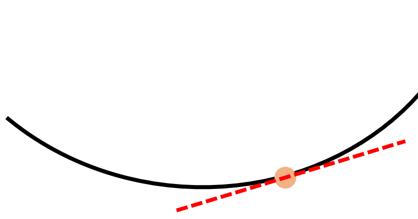
STEP2. 경사 하강 학습법

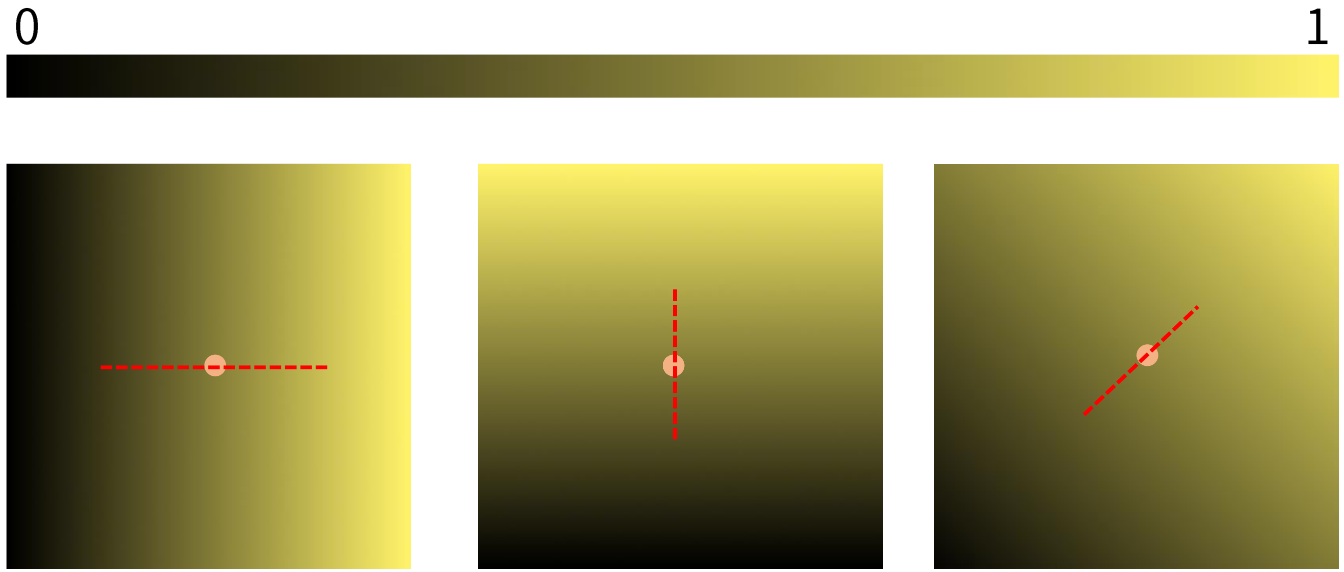
경사 하강법



개념은 이해했으니 보다 깊은 이해를 위해 수학적인 표현을 짚어보고 가자.

미분과 기울기 Gradient



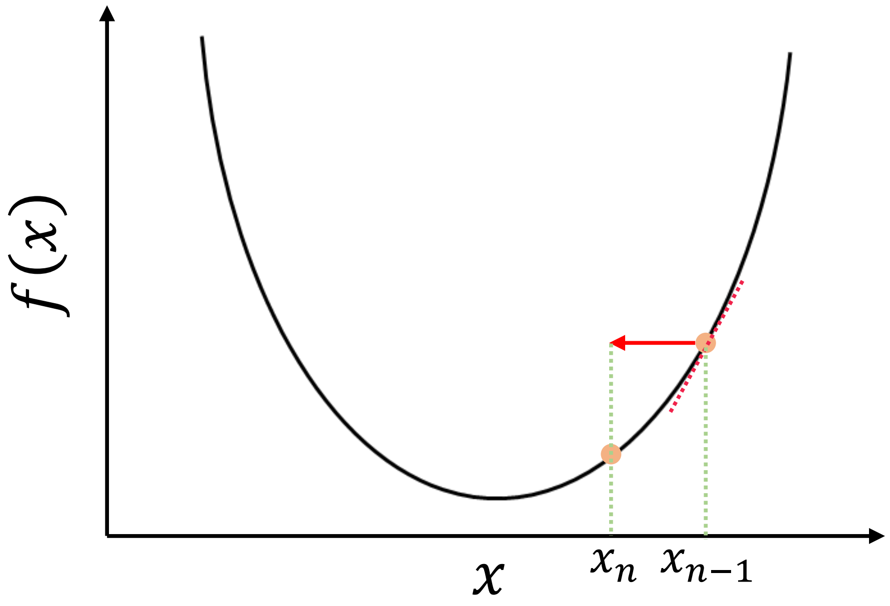


$$\frac{df(x)}{dx} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

$$\nabla f(\boldsymbol{x}) = \left[\frac{df(x_0)}{dx_0}, \frac{df(x_1)}{dx_1}, \ldots, \frac{df(x_{N-1})}{dx_{N-1}}\right]^T$$

기울기(Gradient)는 스칼라를 벡터로 미분한 것이며, 벡터의 각 요소로 미분하면 된다.

경사 하강법



경사 하강법 (Gradient descent)의 한 스텝

1-D의 경우

$$x_n = x_{n-1} - \alpha \frac{df(x_{n-1})}{dx}$$

N-D 의 경우

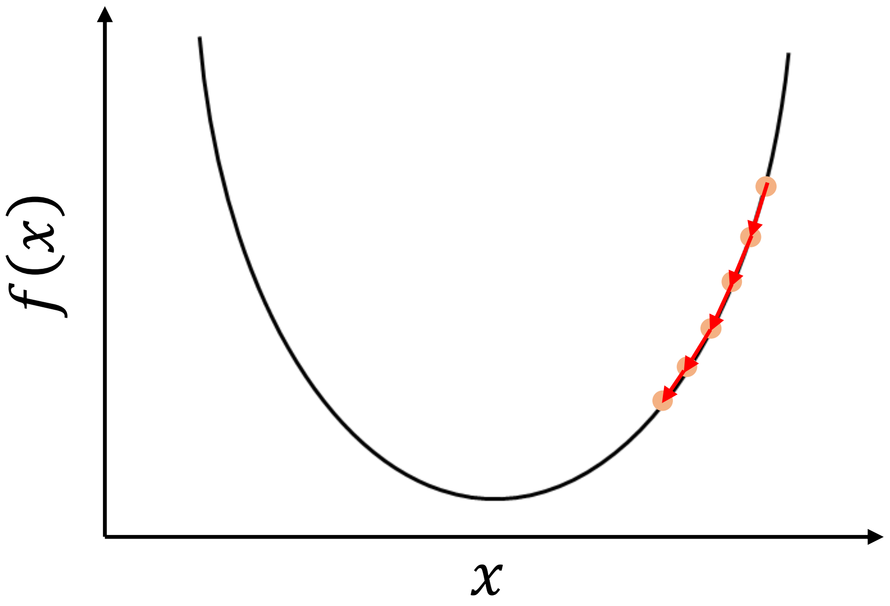
$$\boldsymbol{x}_n = \boldsymbol{x}_{n-1} - \alpha \nabla f(\boldsymbol{x}_{n-1})$$

α: 학습률 (Learning rate)

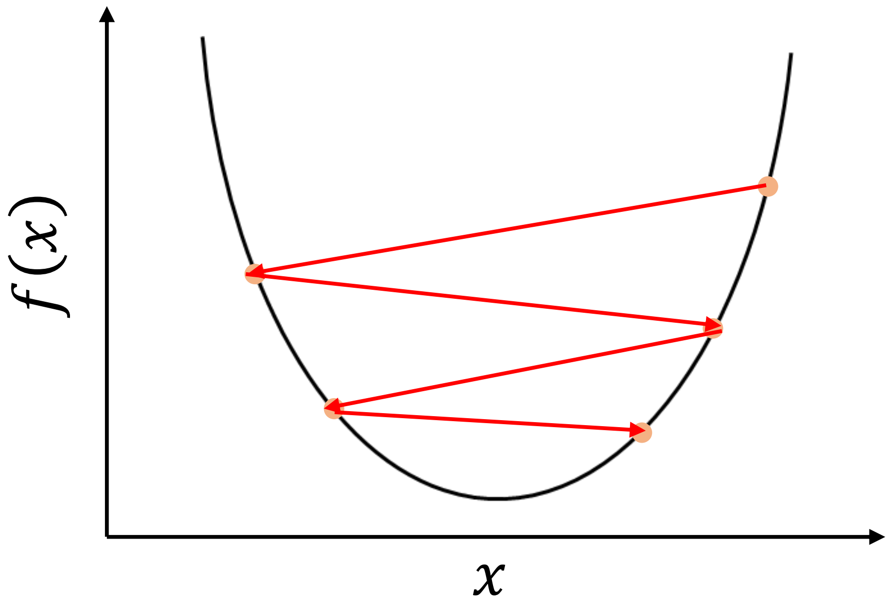
경사 하강법은 $f(x)$의 값이 변하지 않을 때 까지 스텝을 반복한다.

학습률의 선택

$$\boldsymbol{x}_n = \boldsymbol{x}_{n-1} - \underline{\alpha}\nabla f(\boldsymbol{x}_{n-1})$$



α가 너무 작은 경우 (수렴이 늦음)

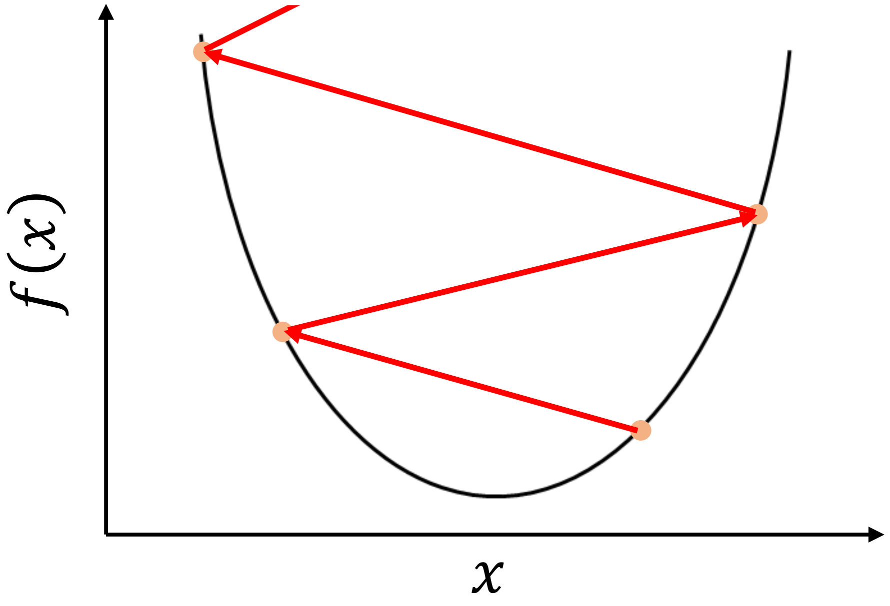


α가 너무 큰 경우 (진동)

이제 학습률은 단지 기울기에 곱하는 상수임을 알 수 있다.

학습률의 선택

$$\boldsymbol{x}_n = \boldsymbol{x}_{n-1} - \alpha \underline{\nabla f(\boldsymbol{x}_{n-1})}$$



한 스텝의 크기는 기울기의 크기에도 비례하므로, 학습률이 극단적으로 크면 값이 발산한다.