경사하강법(Gradient decent algorithm) - review loss function $E(W, b)$

$$y = Wx + b$$

$$\text{loss function} = E(W,b) = \frac{1}{n}\sum_{i=1}^{n}[t_i - y_i]^2 = \frac{1}{n}\sum_{i=1}^{n}[t_i - (Wx_i + b)]^2$$

- 손실함수는 오차의 평균값을 나타내기 때문에, 손실함수가 최소값을 갖는다는 것은 실제 정답과 계산 값의 차이인 오차가 최소가 되어, 미지의 데이터에 대해서 결과를 더 잘 예측할 수 있다는 것을 의미

- 이러한 손실함수는 W, b의 영향을 받기 때문에, 손실함수가 최소가 되는 가중치 W와 바이어스 b를 찾는 것이 regression system을 구현하는 최종 목표

경사하강법(Gradient decent algorithm) – 손실함수(loss function) 계산

- 계산을 쉽게 하고 손실함수의 모양을 파악하기 위해 $E(W,b)$에서 바이어스 b = 0으로 가정

Ex) 다음과 같은 Training Data에서, W값에 대한 손실함수 $E(W,b)$ 계산

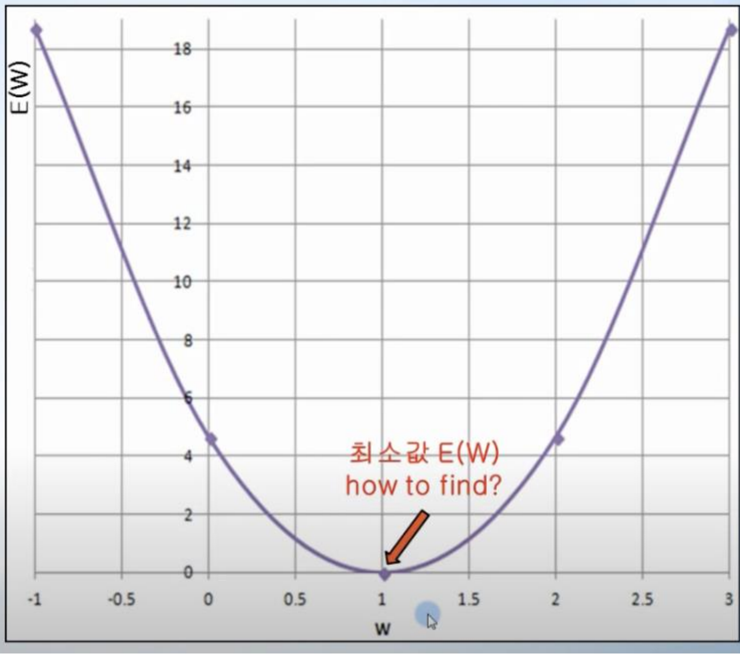
training data

x	t
1	1
2	2
3	3

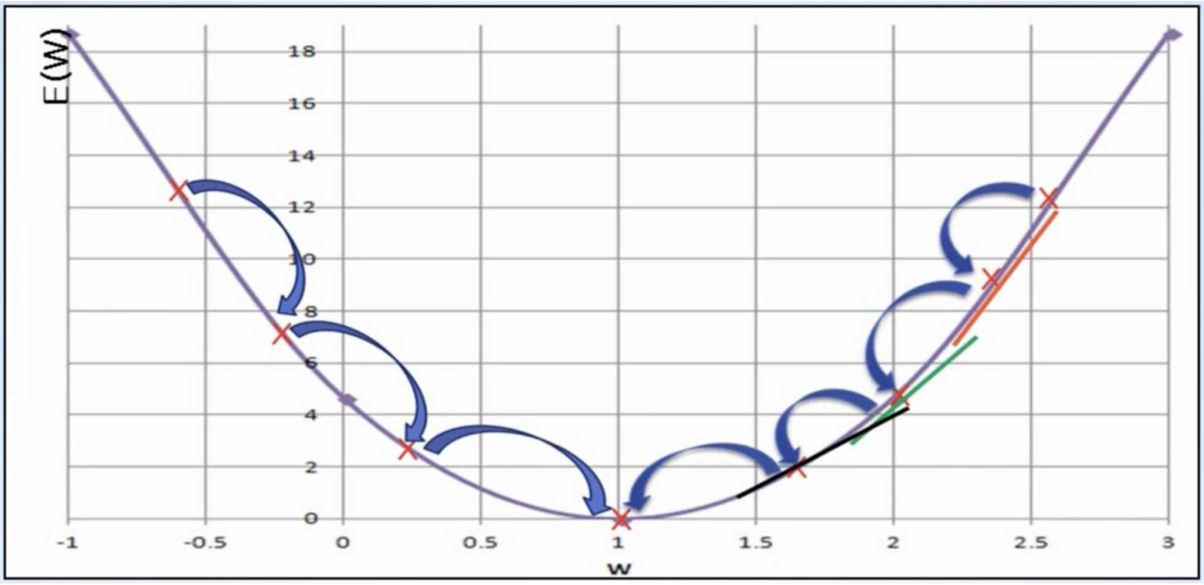
W = -1	$E(-1,0) = \frac{1}{3}\sum_{i=1}^{3}[t_i - (-1 \cdot x_i + 0)]^2$	$= \frac{[1-(-1\cdot1+0)]^2 + [2-(-1\cdot2+0)]^2 + [3-(-1\cdot3+0)]^2}{3} = 18.7$
W = 0	$E(0,0) = \frac{1}{3}\sum_{i=1}^{3}[t_i - (0 \cdot x_i + 0)]^2$	$= \frac{[1-(0\cdot1+0)]^2 + [2-(0\cdot2+0)]^2 + [3-(0\cdot3+0)]^2}{3} = 4.67$
W = 1	$E(1,0) = \frac{1}{3}\sum_{i=1}^{3}[t_i - (1 \cdot x_i + 0)]^2$	$= \frac{[1-(1\cdot1+0)]^2 + [2-(1\cdot2+0)]^2 + [3-(1\cdot3+0)]^2}{3} = 0$
W = 2	$E(2,0) = \frac{1}{3}\sum_{i=1}^{3}[t_i - (2 \cdot x_i + 0)]^2$	$= \frac{[1-(2\cdot1+0)]^2 + [2-(2\cdot2+0)]^2 + [3-(2\cdot3+0)]^2}{3} = 4.67$
W = 3	$E(3,0) = \frac{1}{3}\sum_{i=1}^{3}[t_i - (3 \cdot x_i + 0)]^2$	$= \frac{[1-(3\cdot1+0)]^2 + [2-(3\cdot2+0)]^2 + [3-(3\cdot3+0)]^2}{3} = 18.7$

손실함수(loss function)의 형태

W	E(W)
-1	18.7
0	4.67
1	0
2	4.67
3	18.7



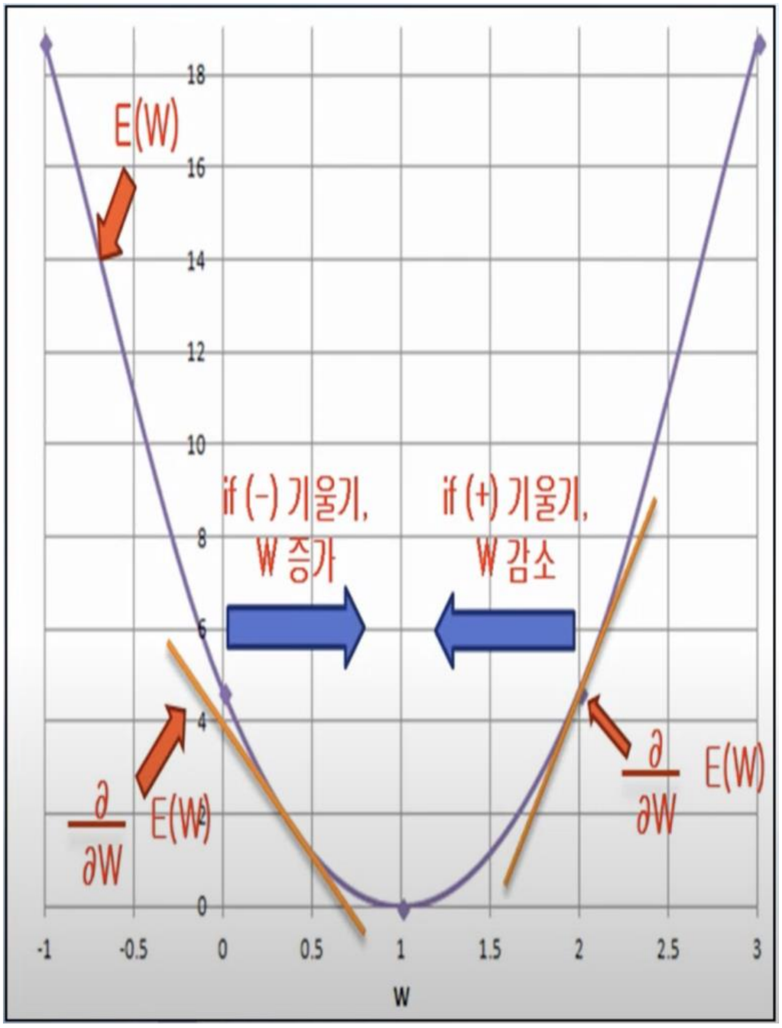
경사하강법(Gradient decent algorithm) – 경사하강법 원리



- ① 임의의 가중치 W 선택
  ② 그 W에서의 직선의 기울기를 나타내는 미분 값 (해당 W에서의 미분, ∂E(W)/ ∂W)를 구함
  ③ 그 미분 값이 작아지는 방향으로 W 감소(또는 증가)시켜 나가면
  ④ 최종적으로 기울기가 더 이상 작아지지 않는 곳을 찾을 수 있는데, 그곳이 손실함수 E(W)의 최소값임을 알 수 있음

- 이처럼, W에서의 직선의 기울기인 미분 값을 이용하여, 그 값이 작아지는 방향으로 진행하여 손실함수 최소값을 찾는 방법을 경사하강법(gradient decent algorithm)이라고 함

경사하강법(Gradient decent algorithm) – W값 구하기



W 에서의 편미분 ∂E(W)/ ∂W
해당 W에서의 기울기(slope)를 나타냄

- ⇨ ∂E(W)/ ∂W 양수 (+) 값을 갖는다면, W는 왼쪽으로 이동시켜야만(감소), 손실함수 E(W) 최소값을 찾음

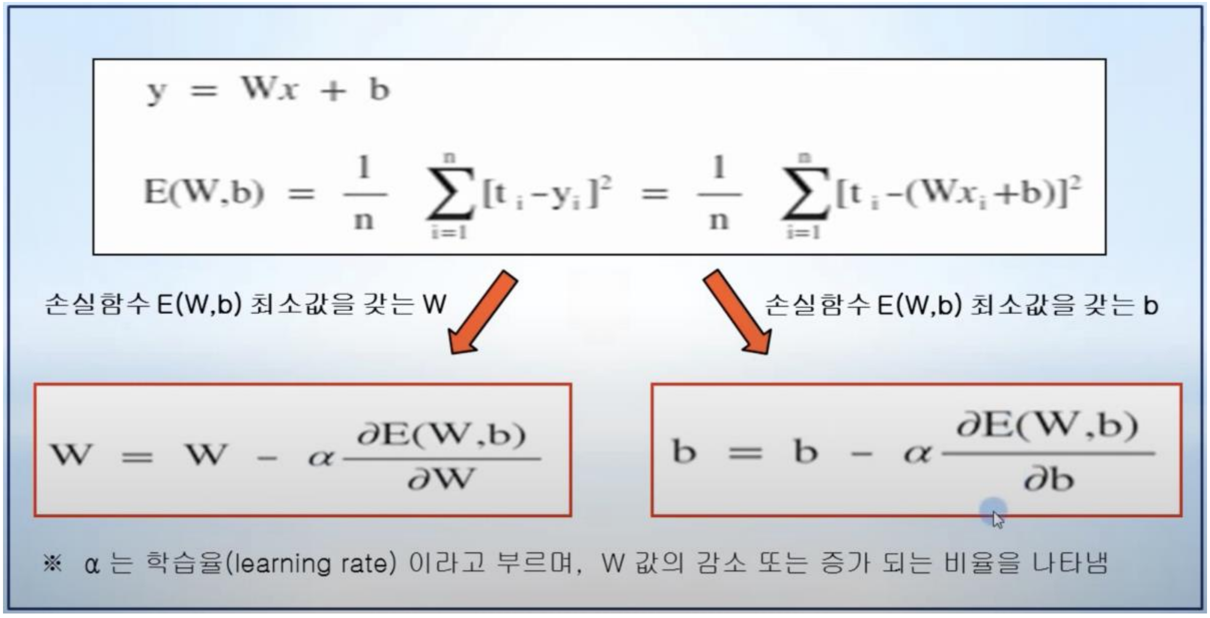
- ⇨ ∂E(W)/ ∂W 음수 (-) 값을 갖는다면, W는 오른쪽으로 이동시켜야만(증가), 손실함수 E(W) 최소값을 찾음

- $W = W - \alpha \frac{\partial E(W,b)}{\partial W}$

α 는 학습율(learning rate)이라고 부르며, W 값의 감소 또는 증가되는 비율을 나타냄

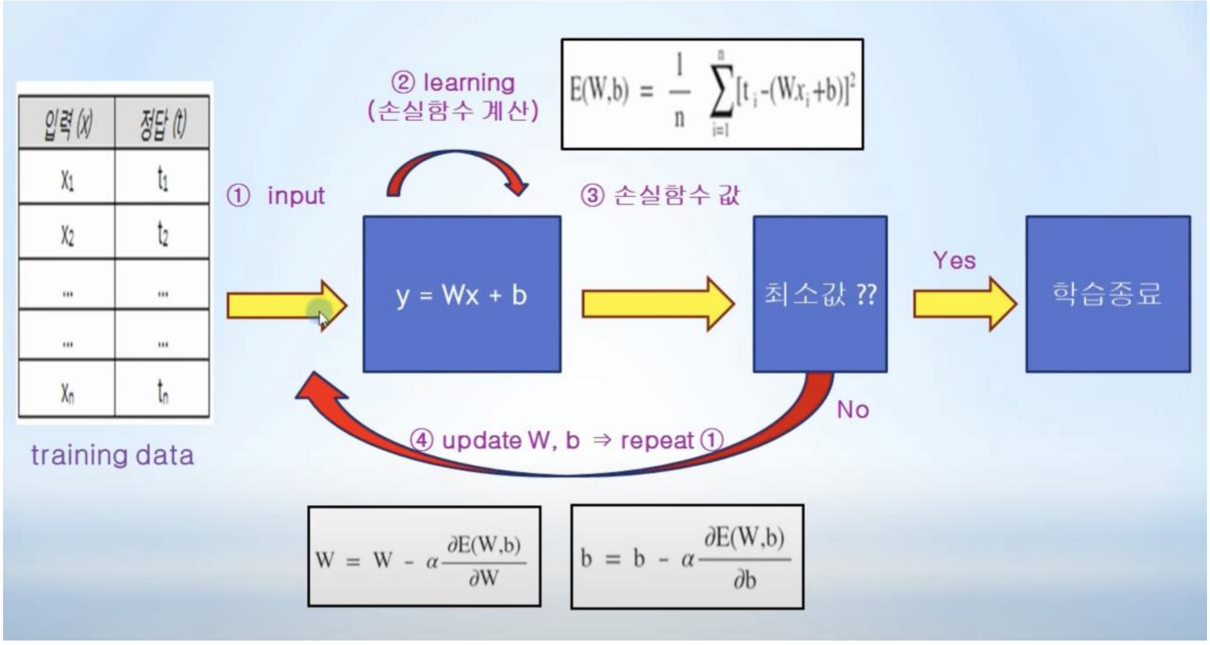
경사하강법(Gradient decent algorithm) – 손실함수 $\boldsymbol{E(W,b)}$ 최소값이 되는 W, b

• linear regression의 목표는 training data의 특성/분포를 가장 잘 나타내는 임의의 직선 $y = Wx + b$에서의 [W, b]를 구하는 것



경사하강법(Gradient decent algorithm)

- **최적의 [W, b] 계산 프로세스**



⇨ 임의의 직선 $y = Wx + b$를 가정

⇨ 주어진 트레이닝 데이터를 이용하여 손실함수를 계산

⇨ 손실함수 값이 최소값이면 학습 종료

⇨ 손실함수 값이 최소가 아니라면, 주어진 W와 b를 미분값을 이용하여 update한 뒤, 다시 처음으로 돌아가서 손실함수 값을 계산해서 최소가 될 때가지 반복