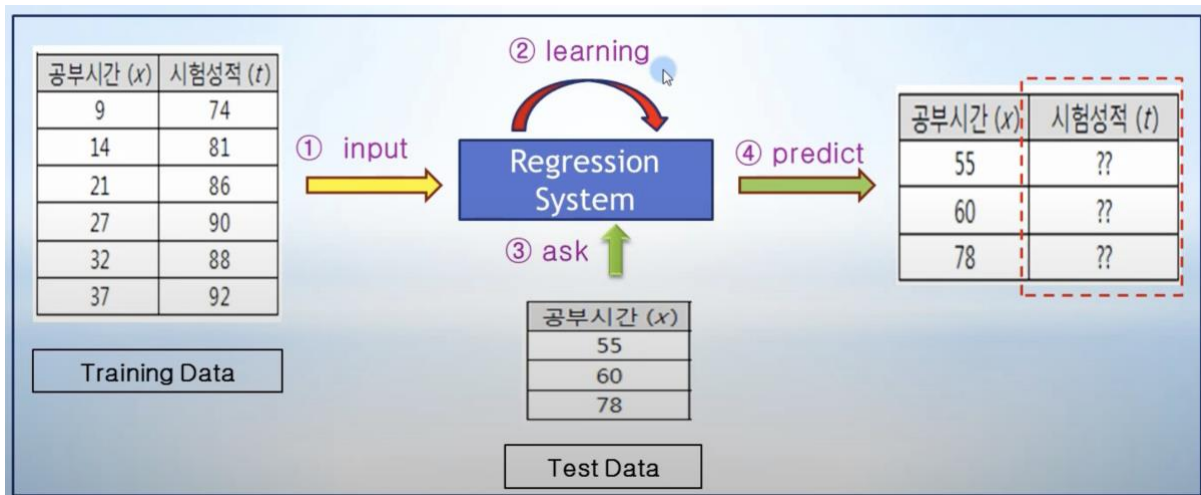


Review – 회귀(Regression)

회귀(Regression)

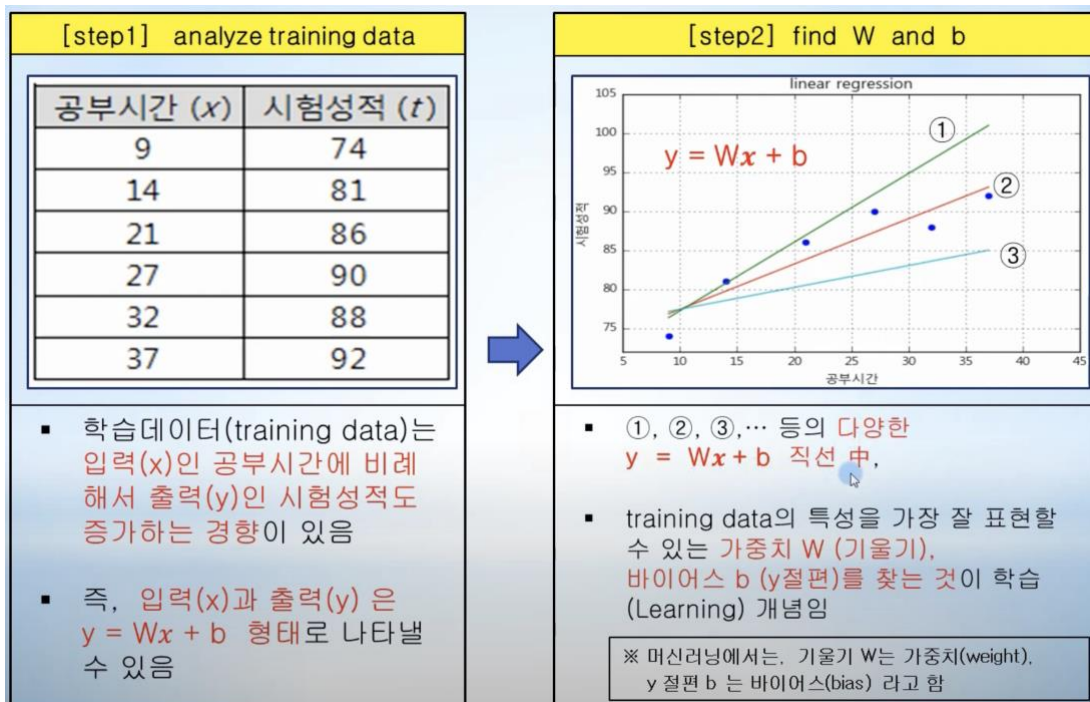
- Training Data를 이용하여 데이터의 특성과 상관관계 등을 파악하고, 그 결과를 바탕으로 Training Data에 없는 **미지의 데이터가 주어졌을 경우에, 그 결과를 연속적인 (숫자) 값으로 예측하는 것**

ex) 공부시간과 시험성적의 관계, 집 평수와 집 가격의 관계 등

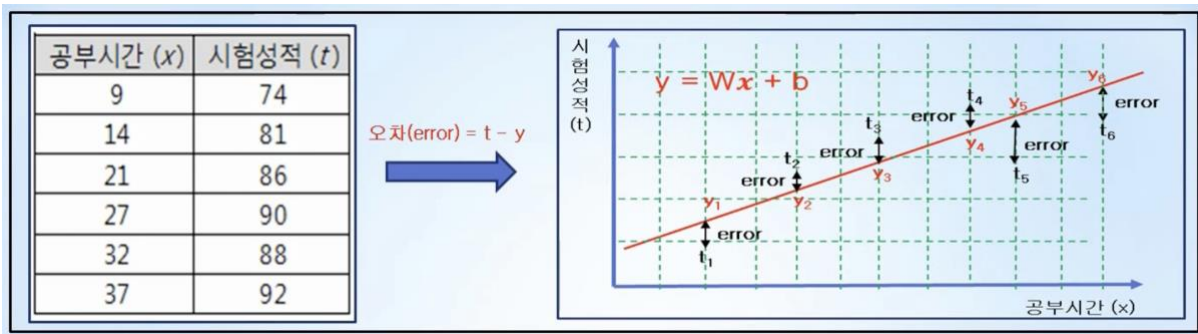


회귀(Regression)

Regression – 학습(learning) 개념



Regression – 오차(error), 가중치(weight) W , 바이어스(bias) b



- Training data의 정답(t)과 직선 $y = Wx + b$ 값의 차이인 오차(error)는,

$\text{오차}(\text{error}) = t - y = t - (Wx + b)$ 로 계산되며,

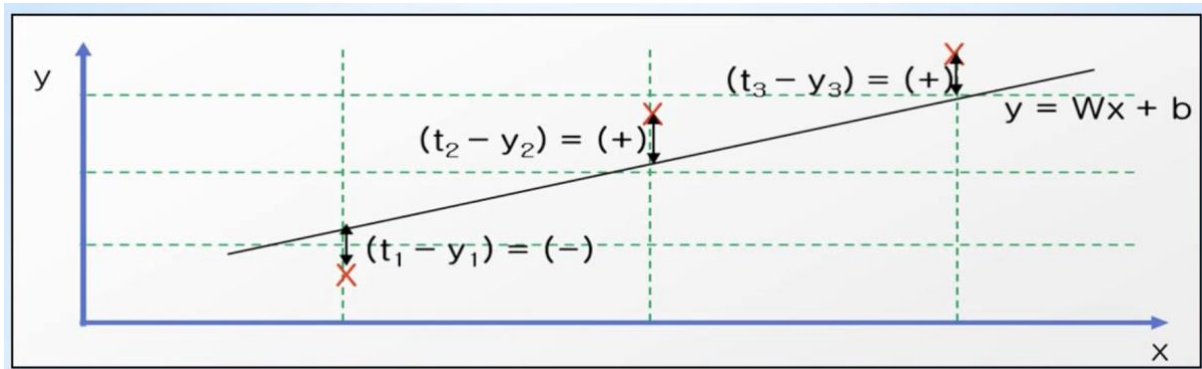
오차가 크다면, 우리가 임의로 설정한 직선의 가중치와 바이어스 값이 잘못된 것이고, 오차가 작다면 직선의 가중치와 바이어스 값이 잘 된 것이기 때문에 미래 값 예측도 정확할 수 있다고 예상할 수 있음

- 머신러닝의 regression 시스템은,

모든 데이터의 오차(error) = $t - y = t - (Wx + b)$ 의 합이 최소가 되어서, 미래 값을 잘 예측할 수 있는 가중치 w 와 바이어스 b 값을 찾아야 함

Regression – 손실함수(loss function)

- 손실함수(loss function / cost function)는, training data의 정답(t)과 입력(x)에 대한 계산 값 y 의 차이를 모두 더해 수식으로 나타낸 것



- 각각의 오차인 $(t - y)$ 를 모두 더해서 손실함수(loss function)을 구하면 각각의 오차가 $(+)$, $(-)$ 등이 동시에 존재하기 때문에 오차의 합이 0이 나올 수도 있음.
즉, 0이라는 것이 최소 오차 값인지 아닌지를 판별하는 것이 어려움

⇒ 손실함수에서 오차(error)를 계산할 때는 $(t - y)^2 = (t - [Wx + b])^2$ 을 사용함.
즉 오차는 언제나 양수이며, 제곱을 하기 때문에 정답과 계산 값 차이가 크다면, 제곱에 의해 오차는 더 큰 값을 가지게 되어 머신러닝 학습에 있어 장점을 가짐

[수식]

각각의 정답 t 에서 계산 값 y 를 빼서 제곱하고, 제곱 값들을 모두 더한 뒤 training data의 개수로 나눔

즉, 손실함수라는 것은 모든 데이터에 대한 평균 오차 값

$$\begin{aligned}\text{loss function} &= \frac{(t_1 - y_1)^2 + (t_2 - y_2)^2 + \dots + (t_n - y_n)^2}{n} \\ &= \frac{[t_1 - (Wx_1 + b)]^2 + [t_2 - (Wx_2 + b)]^2 + \dots + [t_n - (Wx_n + b)]^2}{n} \\ &= \frac{1}{n} \sum_{i=1}^n [t_i - (Wx_i + b)]^2\end{aligned}$$

$$y = Wx + b$$

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

• x 와 t 는 training data에서 주어지는 값이므로, 손실함수(loss function)인 $E(W, b)$ 는 결국 W 와 b 의 영향을 받는 함수

- $E(W, b)$ 값이 작다는 것은 정답(t , target)과 $y = Wx + b$ 에 의해 계산된 값의 평균 오차가 작다는 의미이며,
- 평균 오차가 작다는 것은 미지의 데이터 x 가 주어질 경우, 확률적으로 미래의 결과값도 오차가 작을 것이라고 추측할 수 있음
- 이처럼 training data를 바탕으로 손실함수 $E(W, b)$ 가 최소값을 갖도록 (W, b) 를 구하는 것이 (linear) regression model의 최종 목적