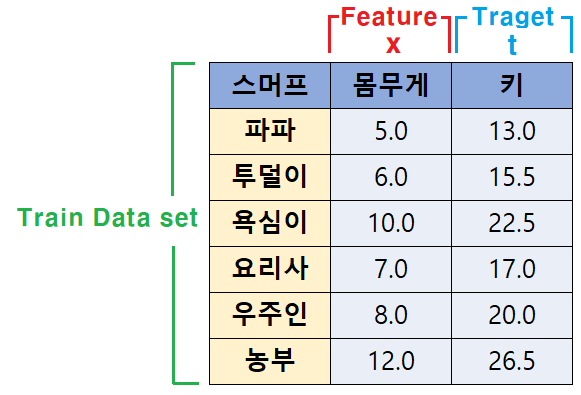
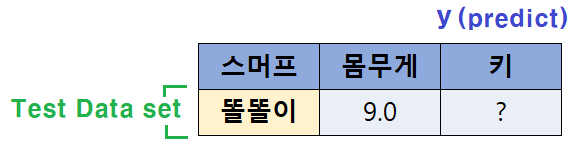
1. **문제 이해**

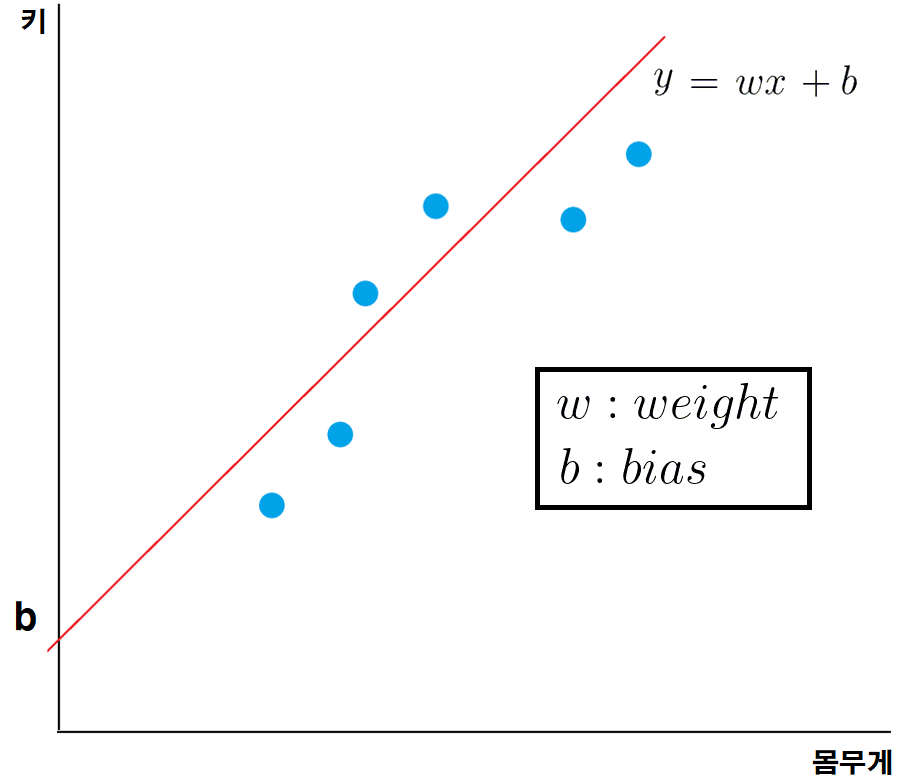
**스머프들의 몸무게에 대한 키 데이터가 아래와 같이 주어졌다.**

****

아래와 같이 똘똘이 스머프의 몸무게가 주어졌을 때, 키는 얼마일까?



학습 데이터 아래와 같이 그래프로 그렸다. 몸무게는X축, 키는 Y축으로 하였다.

****

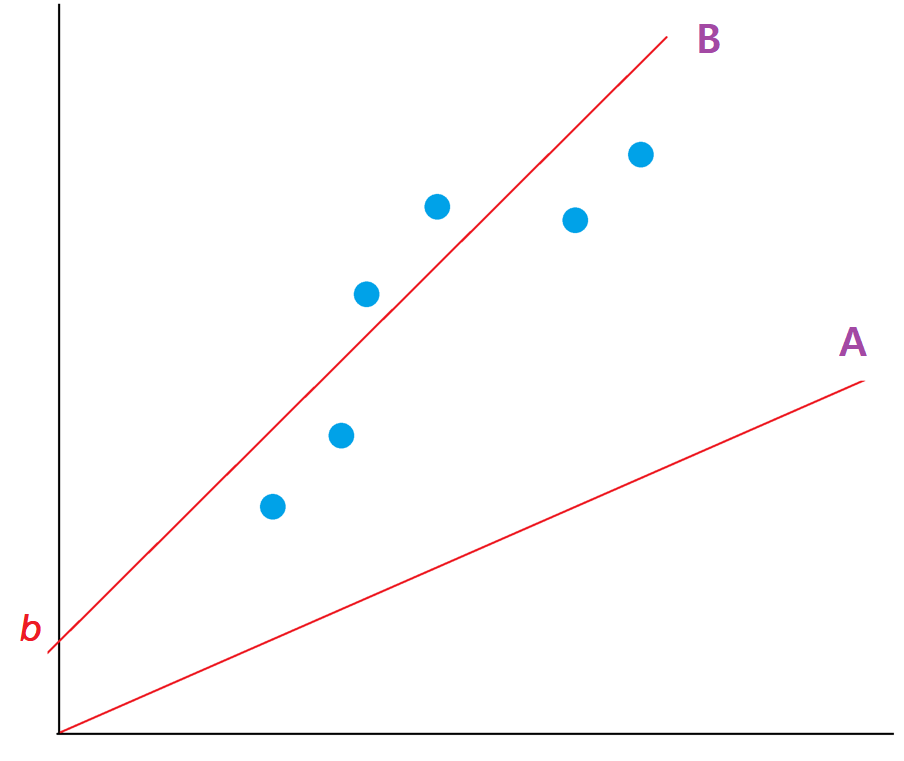
몸무게 값에 대해 키 값이 선형으로 증가 한다는 것을 알 수 있다. 때문에 몸무게를 x, 키를 y로 하면, 아래와 같이 선형함수 공식이 성립 한다.



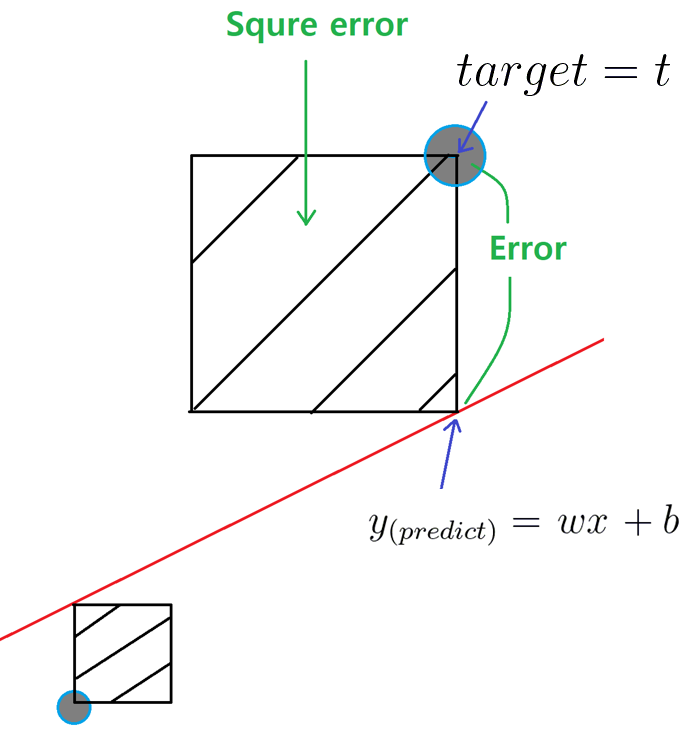
때문에 먼저 학습 데이터로 최적의 w (weight), b (bias)를 찾아 선형 함수를 만든다. 다음으로 선형 함수로 똘똘이 스머프의 몸무게(x 값)을 대입하여 키(y 값)을 구한다.

1. **최적의 w (weight), b (bias)의 기준**

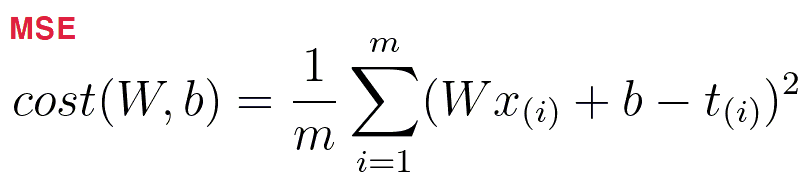
**아래와 같이, 두 선형 함수, A와 B 가 있다. 둘 중에 어느 선형 함수가 더 최적의 w, b를 가진, 즉 예측하기 좋은 함수 일까? 데이터와 거리가 A는 멀고 B는 반대로 짧다. 때문에 B 가 A 에 비해 예측하기 더 좋은 함수이다.**

****

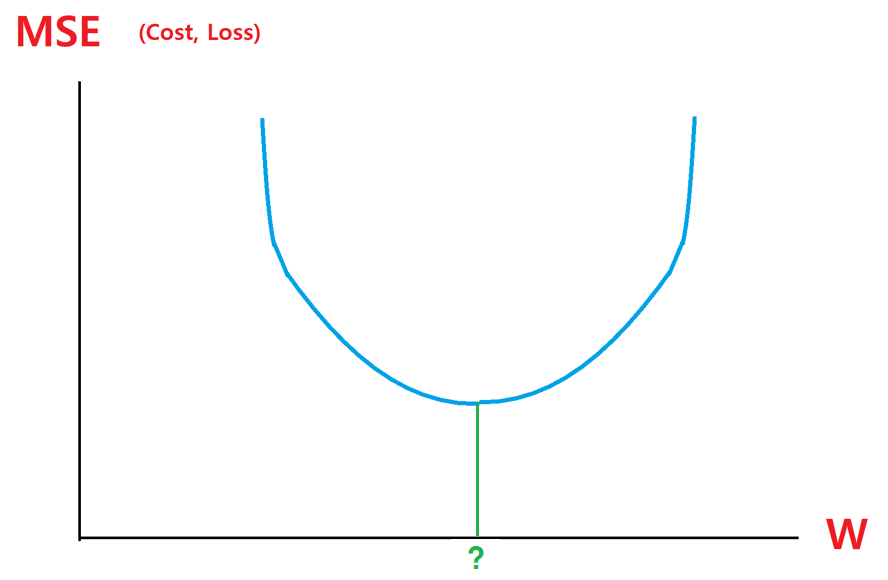
**데이터의 y축의 값을 t(target)로 하자, t와 선형 함수( y (predict) = wx + b) 거리를 Error 라고 한다. 이 Error의 제곱을 Square error 라고 한다. 즉 아래와 같이, 두 변의 길이가 Error를 가진 정 사각형 넓이가 된다. 그리고 모든 데이터와의 Square error를 합의 평균을 MSE(Mean Square Error) 라고 한다.**



**MSE를 Cost 함수 라고 한다. MAE, RMSE 와 같은 다른 함수도 있다. 어쨌든 Cost 함수 MSE를 m 개의 데이터에 대해서 식으로 나타내면 아래와 같다.**

****

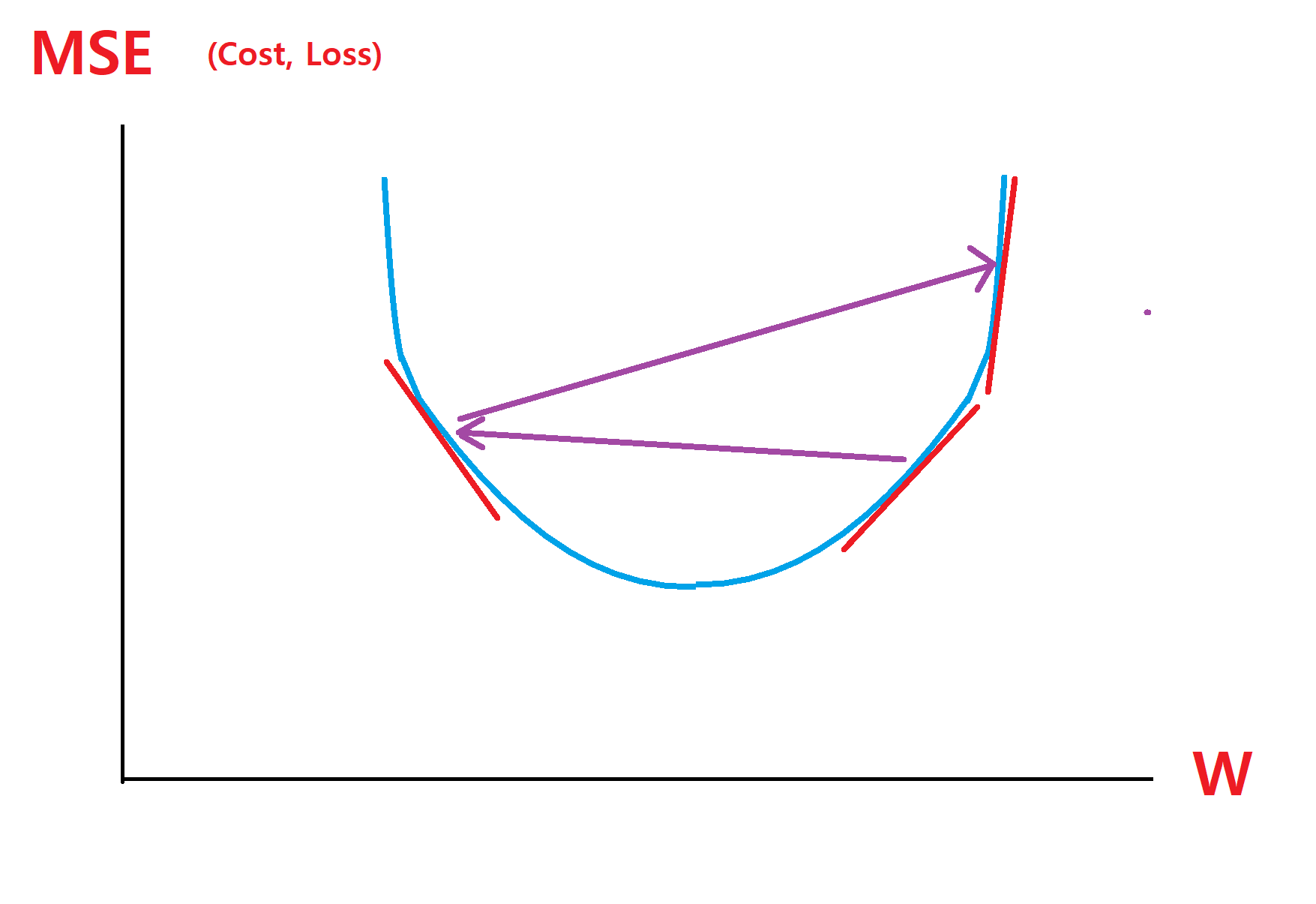
**Cost 함수 값이 최소에 해당하는w, b값을 가진 선형 함수가 예측하기 가장 좋다. Cost 함수는 2차 함수로써 아래 그림과 같은 형태일 것이다. Cost 함수가 최소가 되는 W는 “?” 위치에 해당 할 것이다. 그러면 아래 그림의 “?’ 에 해당하는 W 값은 어떻게 찾을 것인가? 경사하강 법으로 찾을 수 있다.**



1. **경사하강법**

|  |  |
| --- | --- |
|  |  |

**위 그림에서, 찾으려는 최적의 w=2이다. 먼저w를 임의 값을 준다. 그림에서는 1을 주었다. 다음으로, W=1일 때의, MSE 미분 \* learning rate 값을 구한다.  
그림에서 w=1 일 때, 처음 구해지는 값은 -0.5가 된다. 그 값을 현재의 w에 반영하여 새로운w를 구한다. 그림에서 새로 구해진 w는 1.5가 된다.  
다시 다음으로 W=1.5 일 때, (MSE 미분 \* learning rate 값)=-0.2을 구한다. 그리고 다시 w 에 반영한다. 이처럼 계속 계산을 반복하다 보면, 최적의 w=2로 찾아간다.  
최적의 w 에 더 가까울수록 MSE를 미분하면 0에 가까워 지므로, 최적의 W=2벗어나지 않고 수렴 할 수 있다.**

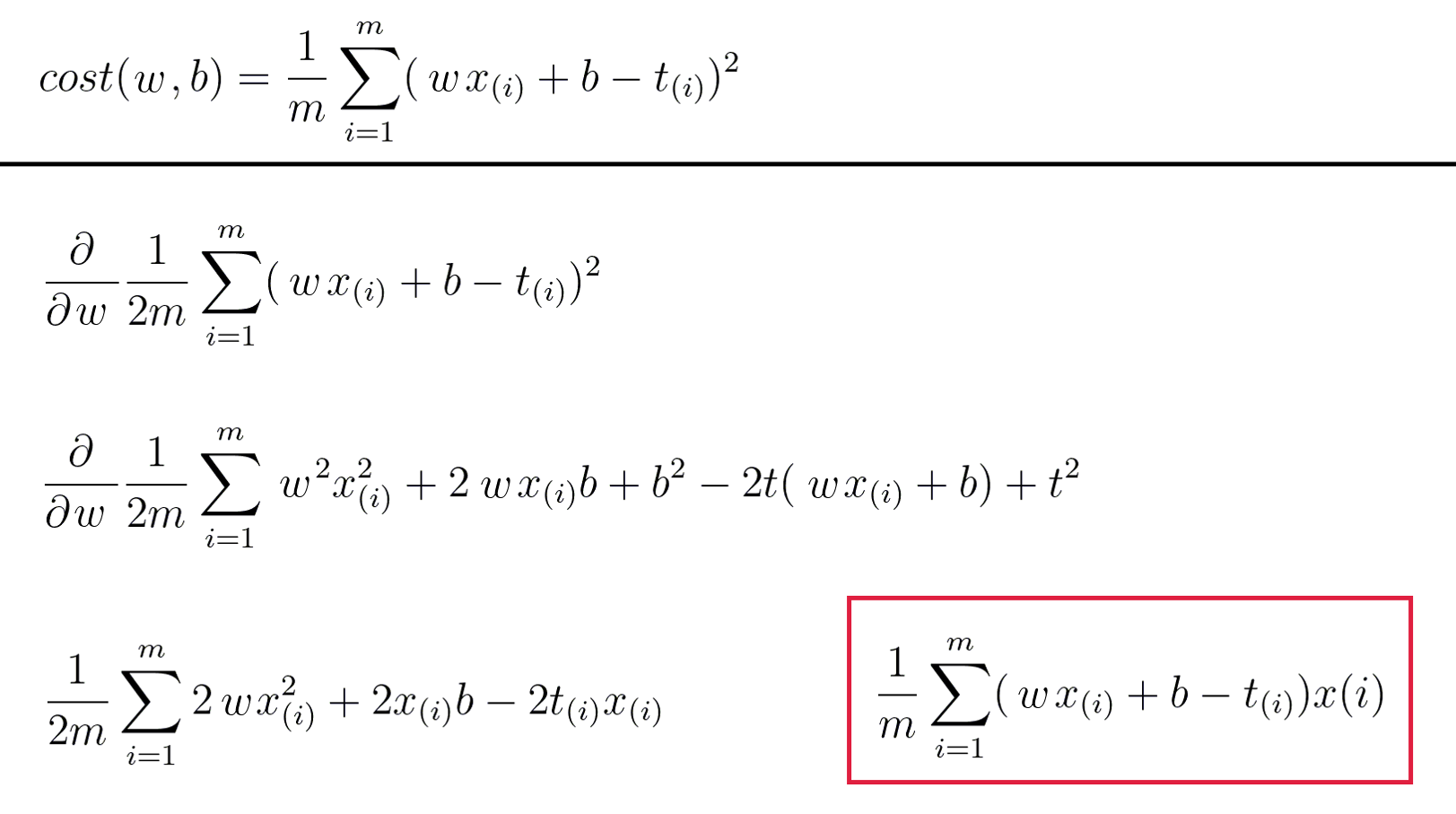


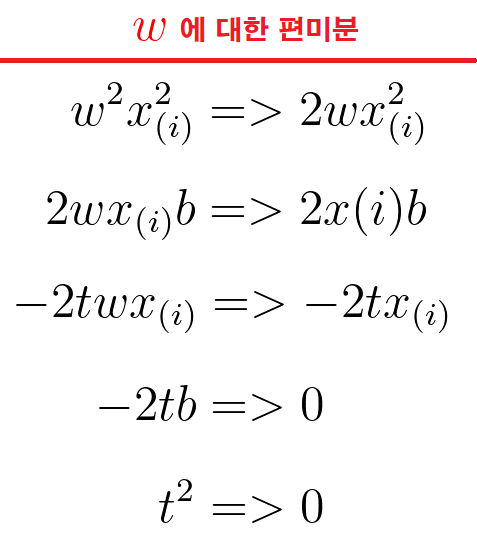
**learning rate 너무 크면, W의 변화폭이 너무 커서 발산하는 문제가 발생한다.**

**반대로, 너무 작으면, W의 변화폭이 너무 작아서 최적의 w를 찾는데 시간이 많이 소요 된다. 때문에 적절한 값을 정해야 한다. 위의 그림은 최적의 w를 찾지 못하고 발산하는 문제를 표현 하였다.**

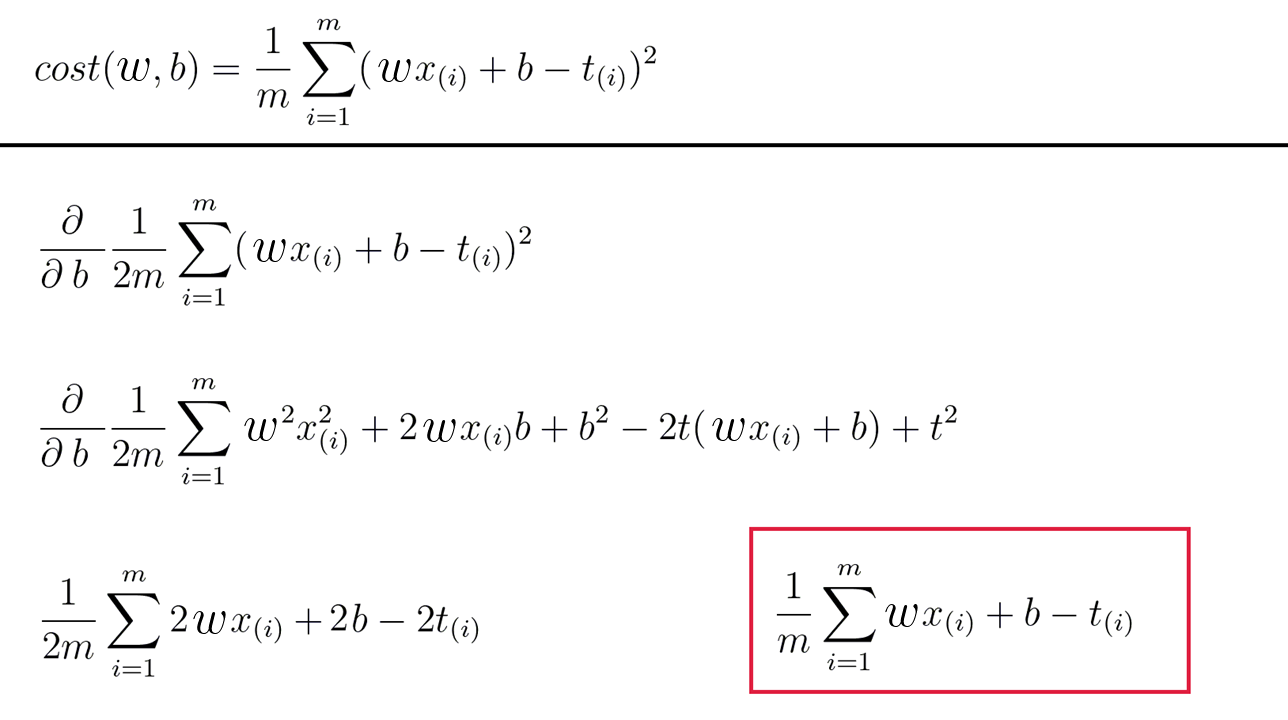
1. **MSE 미분 공식**

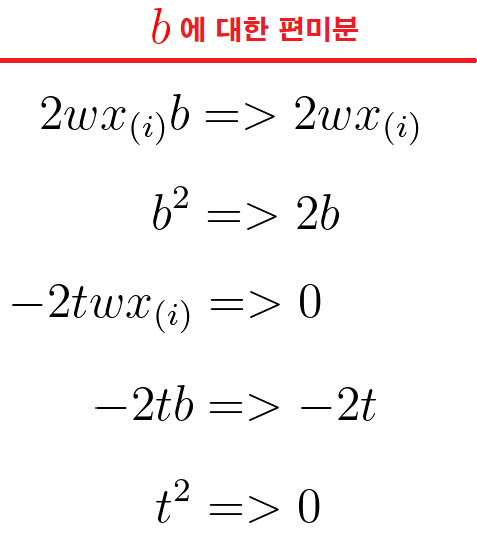
**아래 그림은 Cost 함수 MSE를 w에 대한 편 미분 공식 계산 과정이다.**

****



**아래 그림은 Cost 함수 MSE 에서 b에 대한 편 미분 공식 계산 과정이다.**

****

****

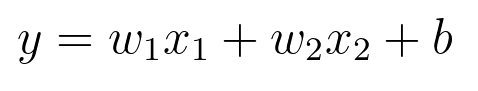
1. **N-차원 데이터**

**지금까지의 설명을 쉽게 하기 위해 몸무게 하나만 x로 두었다. 즉 1차원 x 데이터에 대한 선형함수에 대한 설명 이었다. 사실 선형 함수는 다 차원 x 데이터에 대한 것이다. 아래와 같이 연령 또한 키에 대하여 선형 함수 관계에 있을 수 있다.**

****

****

**x 를 2차원으로 하면 선형함수는 아래와 같다.**

****

**x가 2차원, N차원이 되어도 1차원 x의 선형회귀 원리와 같다.**