

Machine Learning

Linear Regression

선형 회귀

x(hour)	y(score)
9	90
8	80
4	40
2	20

시험성적 데이터

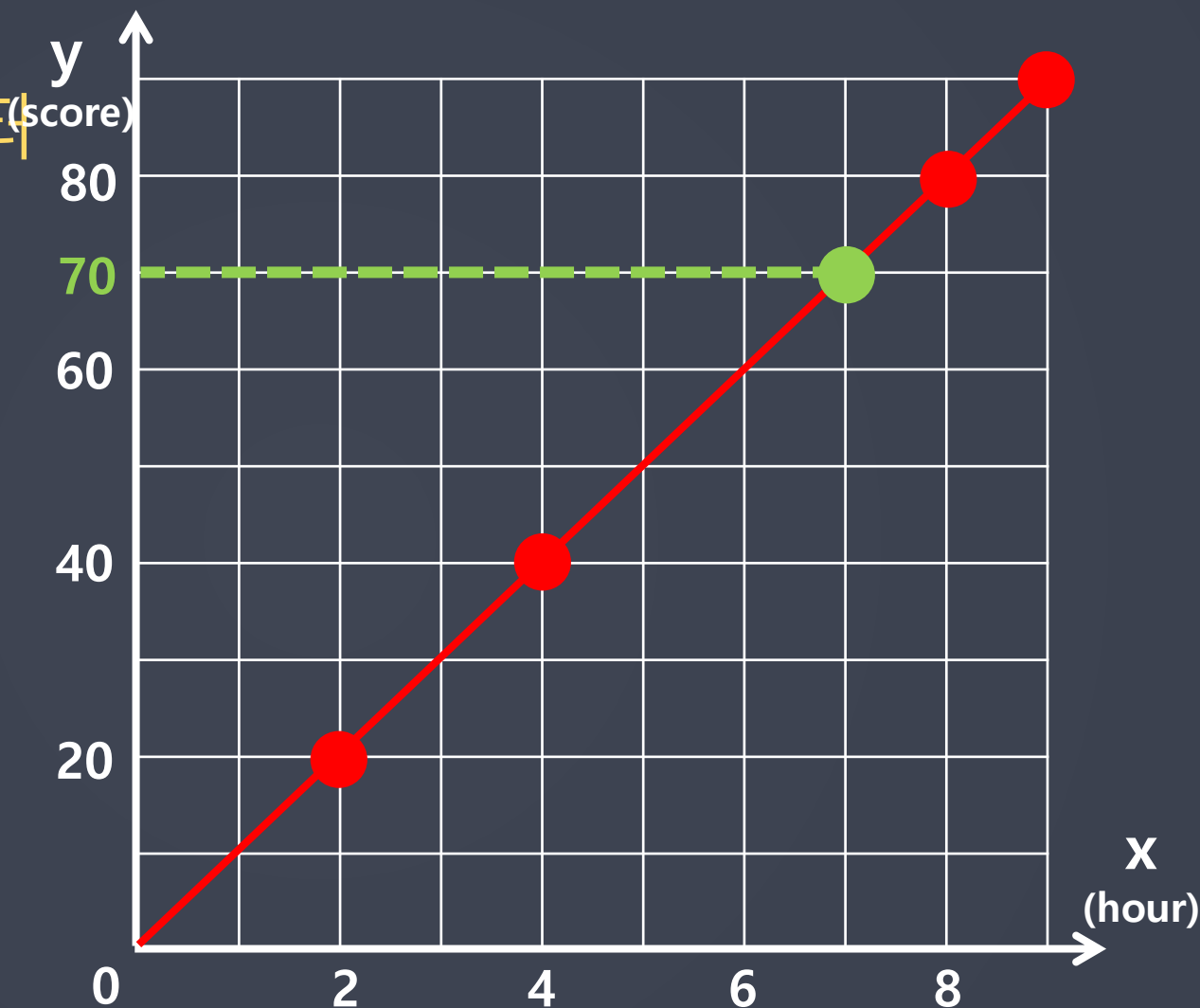
7시간 공부 할 경우
성적은 몇 점 일까?

시험 성적 데이터

x(hour)	y(score)
9	90
8	80
4	40
2	20

$$y = ax + b$$

$$y = 10x + 0$$



$$y = wx + b$$

$$y = w_1x_1 + w_2x_2 + \dots + w_{13}x_{13} + b$$

모델 파라미터(Model Parameter)

모델이 학습하며 스스로 찾아내는 파라미터

w : 가중치, 계수

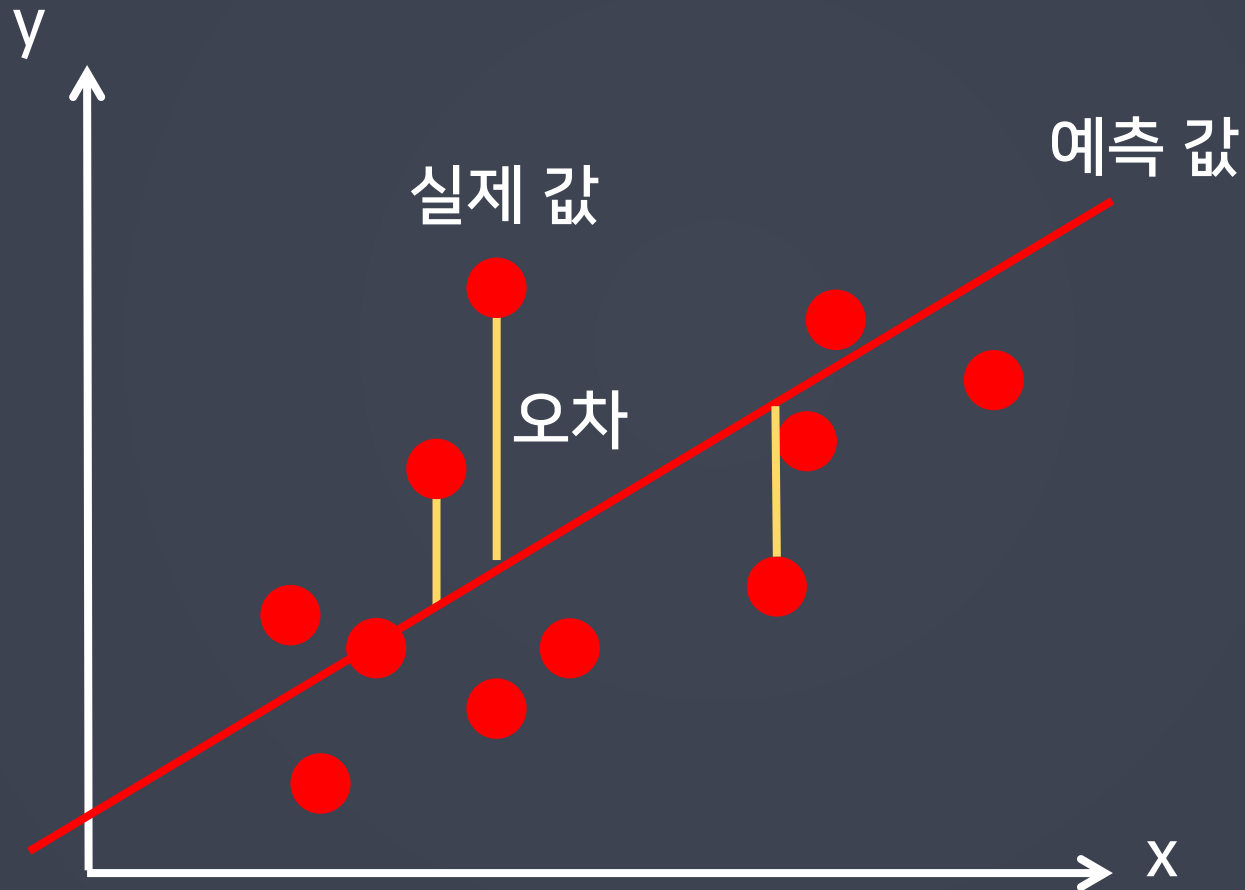
b : 절편, 편향

하이퍼파라미터(HyperParameter)

모델 학습 전에 사람이 직접 지정해 주어야 하는 파라미터

knn에서 k

MSE(Mean Squared Error)



MSE(Mean Squared Error)

$$Cost = \frac{1}{m} \sum_{i=1}^m (y_m - \hat{y}_m)^2$$

y_m : 실제값

$\hat{y}_m = wx + b$: 예측값

가중치에 따른 MSE 그래프 그리기

Gradient descent algorithm

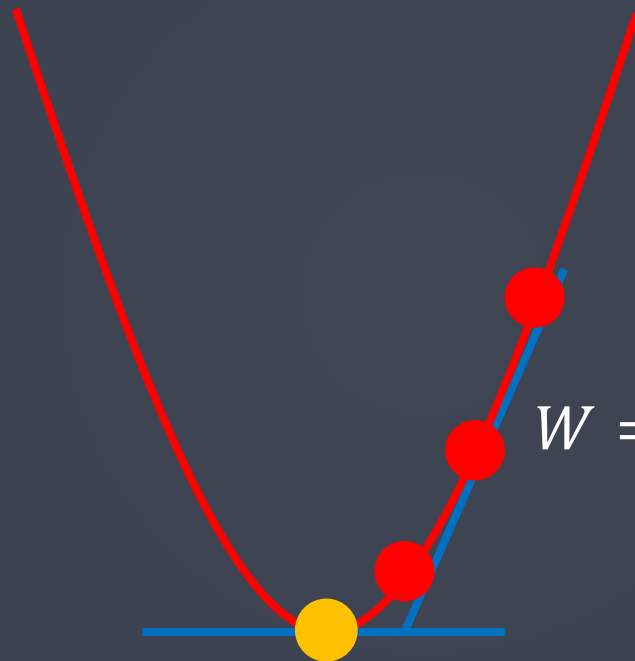


Gradient descent algorithm

함수의 기울기(경사)를 구하여 기울기가 낮은 쪽으로
계속 이동하여 값을 최적화 시키는 방법

경사하강법(Gradient descent algorithm)

비용함수

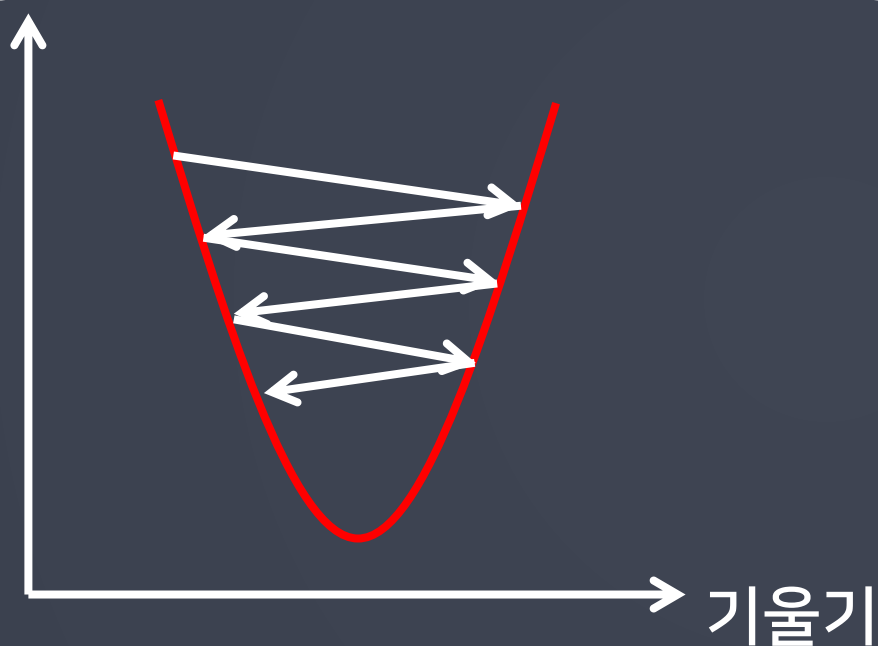
*learning rate*

$$W = W - \alpha \frac{\partial}{\partial W} cost(W)$$

기울기

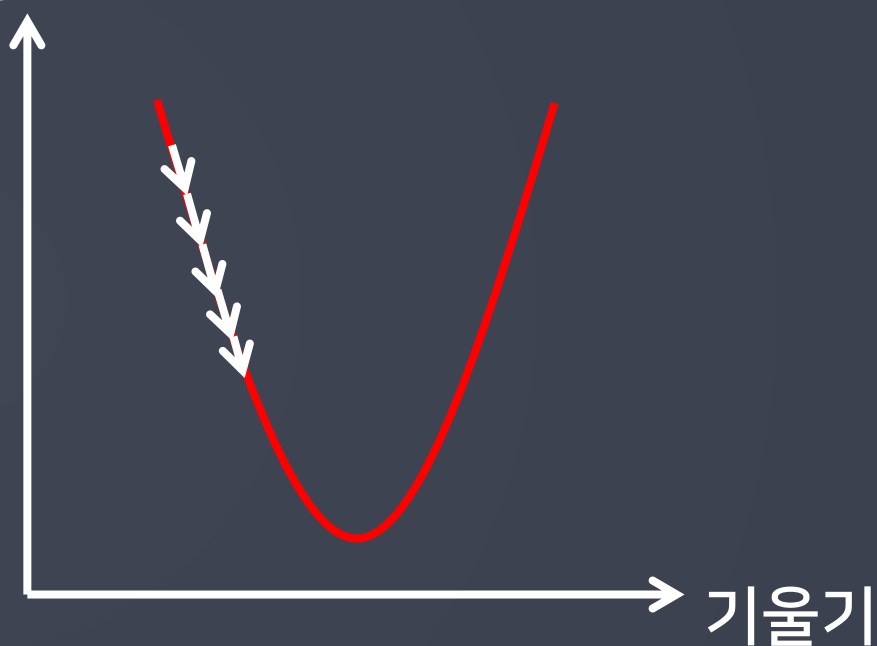
학습률이 큰 경우

비용함수



학습률이 작은 경우

비용함수



Gradient descent algorithm을
사용해서 가중치를 찾아보자

장단점

- 특성이 많은 데이터 세트라면 훌륭한 성능을 낼 수 있다.
- 모델의 복잡도를 제어할 방법이 없어 과대적합 되기 쉽다.



모델 정규화(Regularization)을 통해 과대적합을 제어한다.

정규화

- 가중치(w)의 값을 조정하여 제약을 주는 것.
- L1 규제 : Lasso
 w 의 모든 원소에 똑같은 힘으로 규제를 적용하는 방법.
 특정 계수들은 0이 됨.
 특성선택(Feature Selection)이 자동으로 이루어진다.
- L2 규제 : Ridge
 w 의 모든 원소에 골고루 규제를 적용하여 0에 가깝게 만든다.

정규화 : cost 함수

alpha hyperparameter로 조정

L1 규제 : Lasso

$$J(w)_{LASSO} = \sum_{i=1}^n (y^{(i)} - \hat{y}^{(i)})^2 + \lambda \sum_{j=1}^m |w_j|$$

L2 규제 : Ridge

$$J(w)_{Ridge} = \sum_{i=1}^n (y^{(i)} - \hat{y}^{(i)})^2 + \lambda \sum_{j=1}^m w_j^2$$