

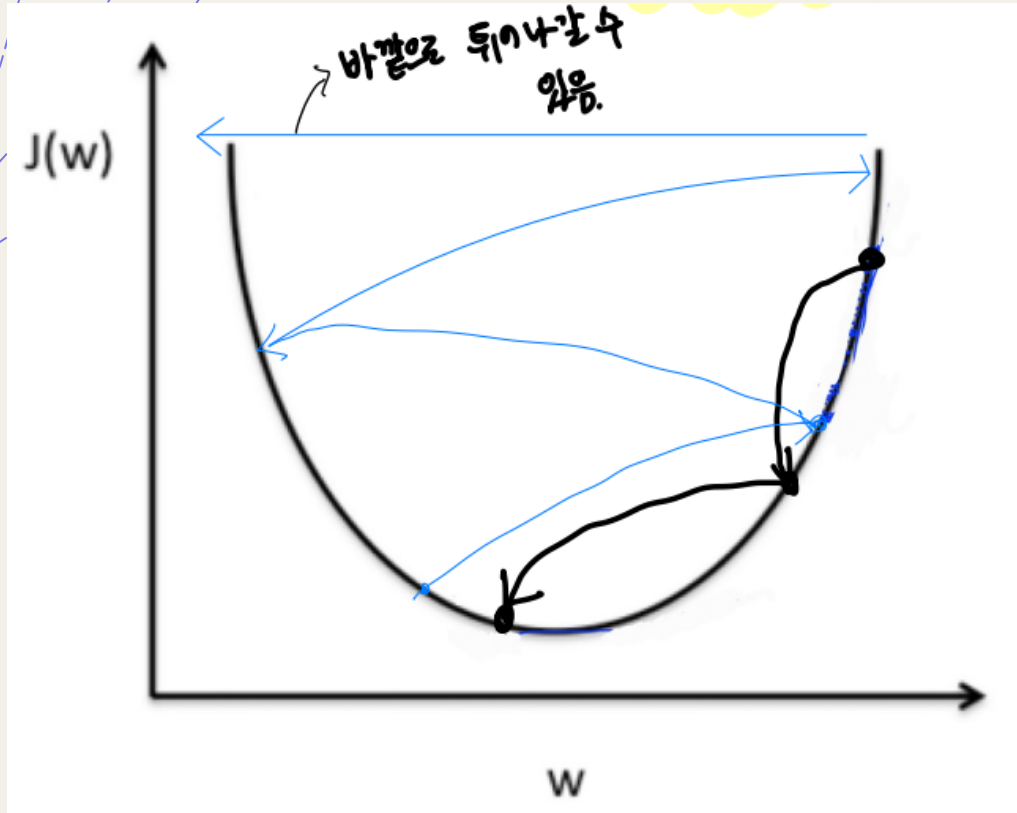
Tips

(Learning rate, data preprocessing,
overfitting)

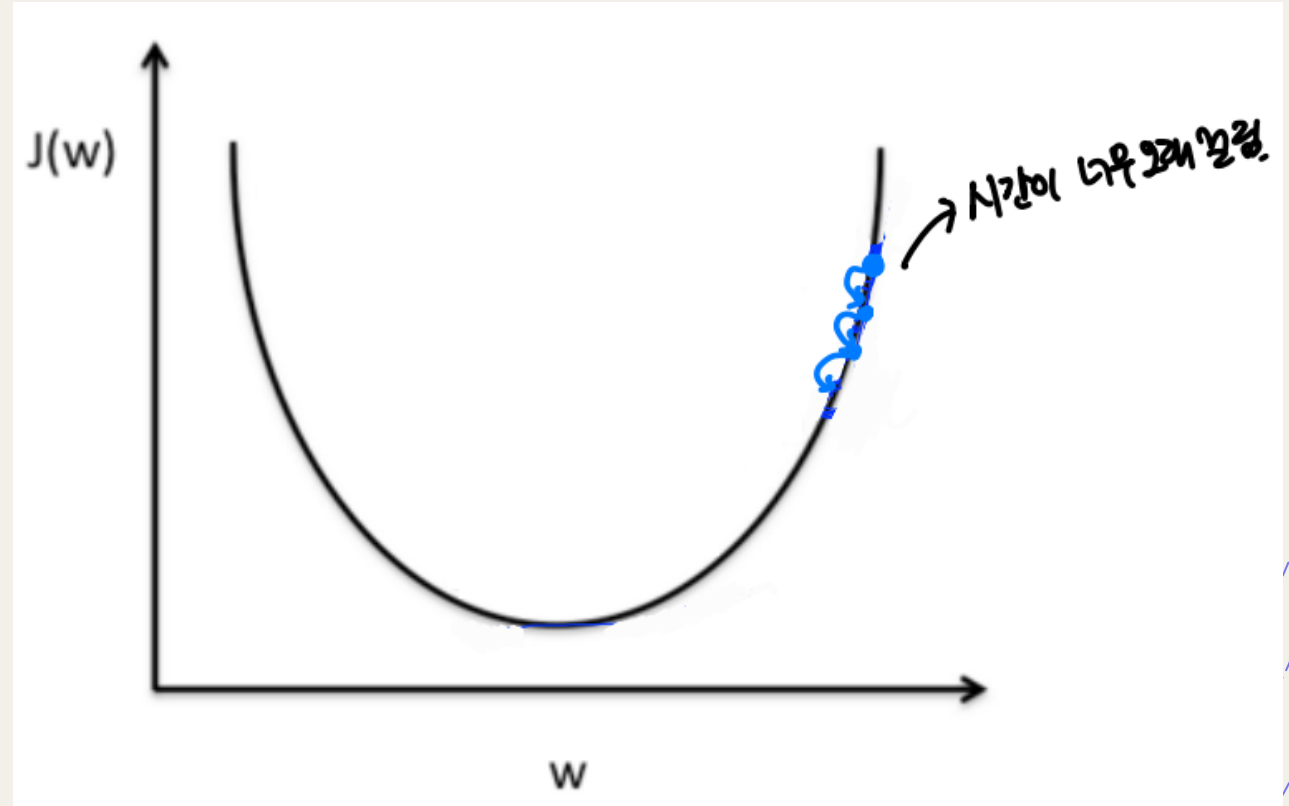
+

Learning rate

Learning rate를 크게 잡았을 때 :
overshooting

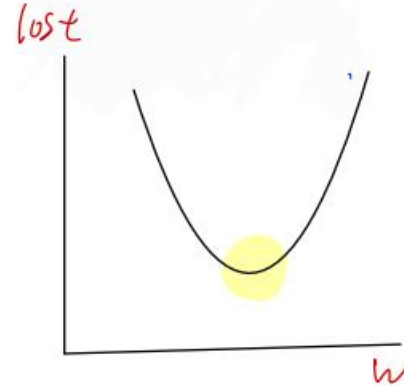
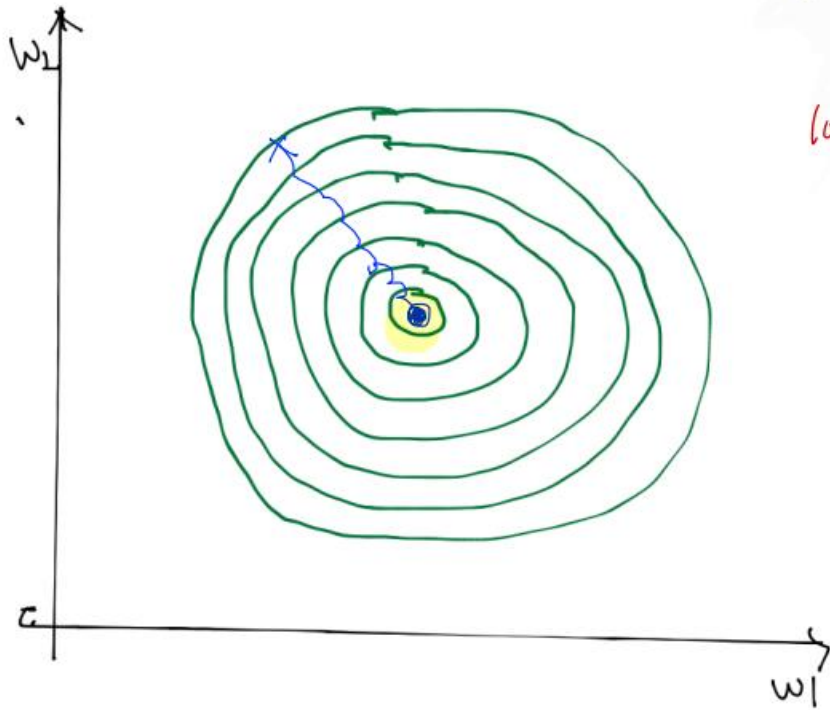


Learning rate를 작게 잡았을 때 :
small learning rate



Learning rate - cost function을 고려해 보통 0.01로 잡음, 그 후 learning rate를 조금씩 줄이거나 늘려나가면서 알맞은 learning rate를 잡음

Data preprocessing

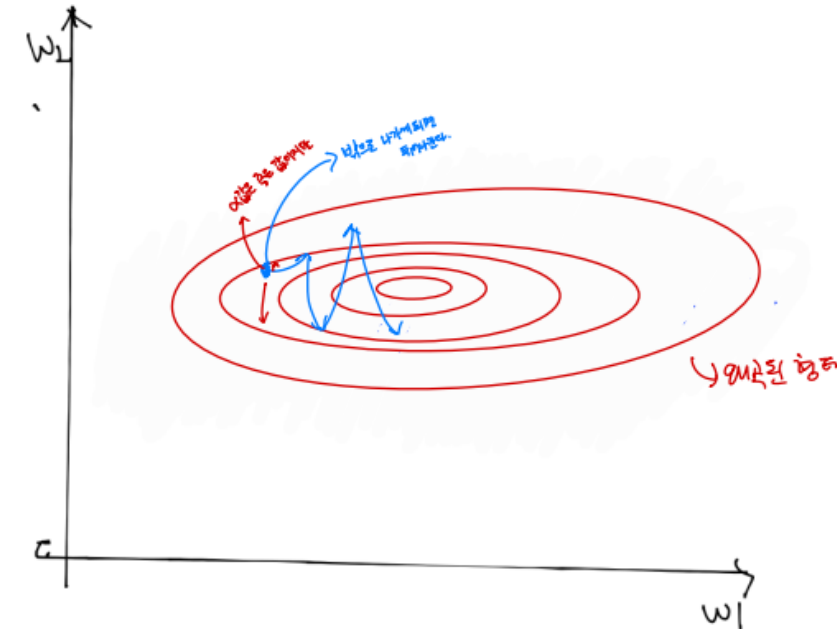


일반적인 등고선 : gradient descent 알고리즘을 이용해 최저점으로 내려올 수 있음

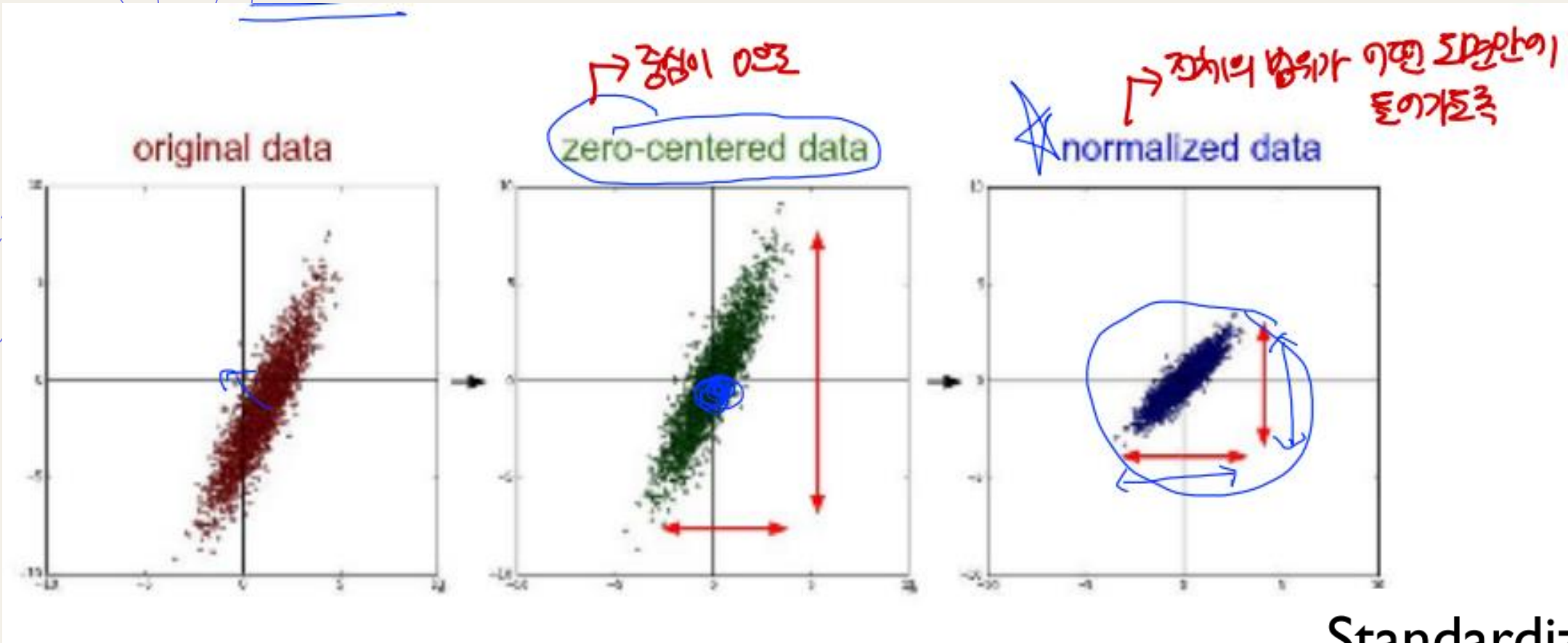
W값의 차이가 크게 나는 등고선 : 왜곡된 형태로 나타나고 learning rate를 잘잡아도 밖으로 튕어나가는 현상발생

가중치 차이가 매우 크다면

	x1	x2	y
1	1	9000	A
2	2	-5000	A
4	4	-2000	B
6	6	8000	B
9	9	9000	C



데이터 값이 큰 차이가 있을 때 -> normalize



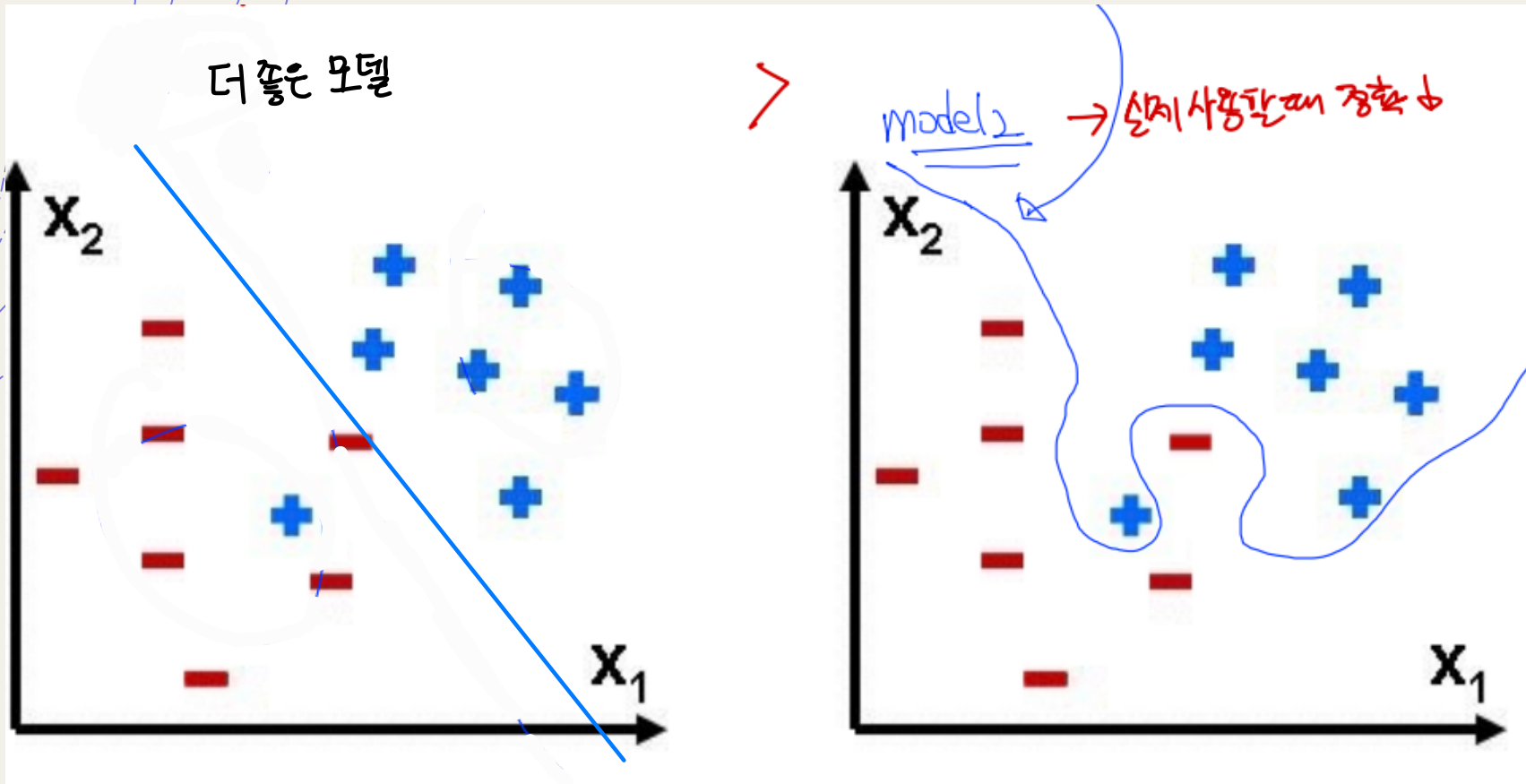
Standardization

$$x'_j = \frac{x_j - \mu_j}{\sigma_j}$$

Handwritten red text: "평균" (mean) points to μ_j and "분산" (variance) points to σ_j . A blue star is next to the formula.

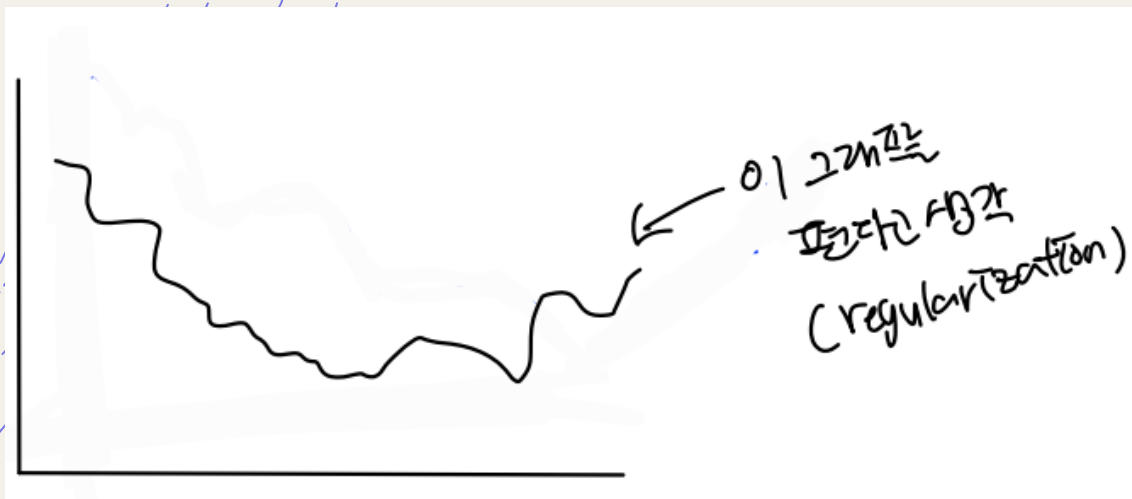
overfitting

- 모델이 너무 정확해서 좋지 않은 형태



Solution - 많은 트레이닝 데이터를 가지고 있는 것.
- 중복된 feature들을 줄이는 것
- 일반화 (regularization)

Solution for overfitting - regularization



LOSS

$$\mathcal{L} = \frac{1}{N} \sum_i \mathcal{D}(s(w x_i + b), L_i) + \lambda \sum W^2$$

TRAINING SET

regularization strength $\rightarrow 0$: regularization (x)
큰 값: regularization 중료

λ 값은 조절해줌.

regularization의 대표적인 방법
- W^2 의 합을 작게 만들어주는 것
 λ 값을 조절해줌.

머신러닝 모델이 얼마나 훌륭한지 평가 가능한가?

Size	Price
2104	400
1600	330
2400	369
1416	232
3000	540
1985	300
1534	315
1427	199
1380	212
1494	243

→ Bad made 1

같은 데이터를
시험하는 것은
의미가 없음.

training
set

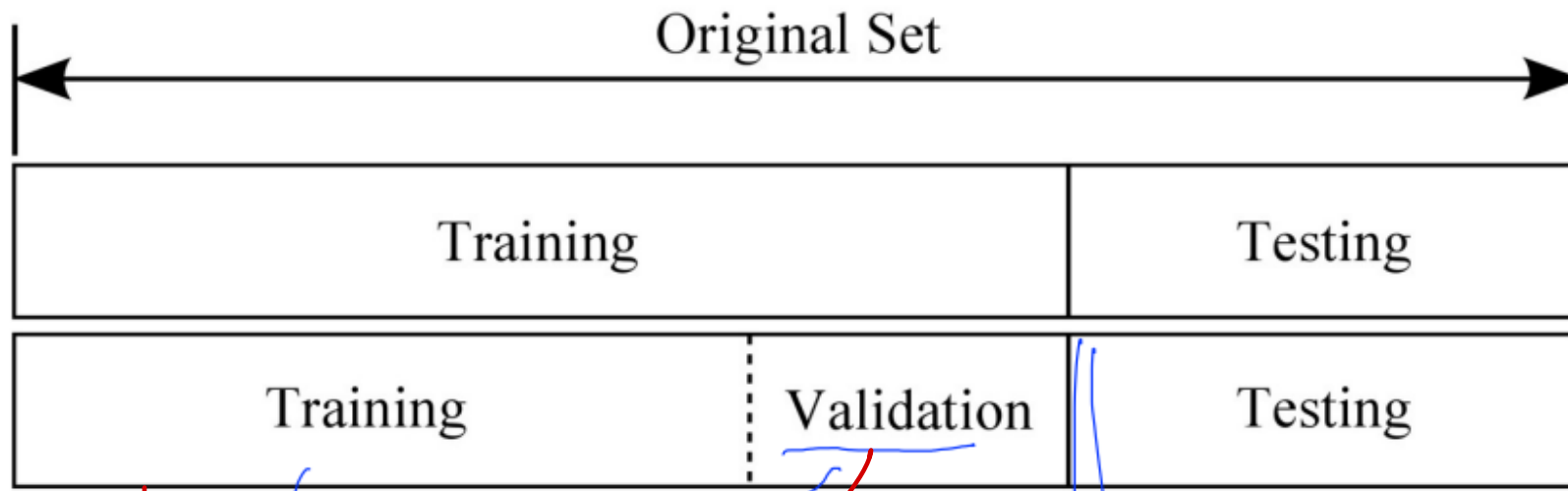
test
set

Size	Price
2104	400
1600	330
2400	369
1416	232
3000	540
1985	300
1534	315
1427	199
1380	212
1494	243

학습시킨 모델로
예측값 도출 \hat{Y}

$\hat{Y} \approx Y$ 의 유사한 정도

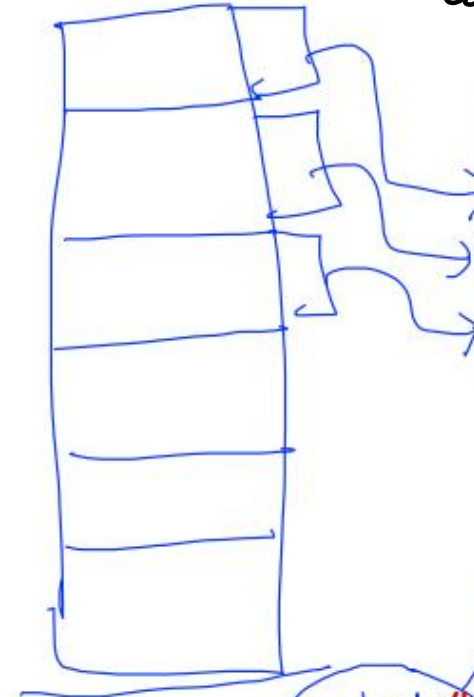
실제데이터 Y



→ 학습

α, λ
값을 튜닝

100만개



10만개씩

Online learning

많은 데이터들을 조각조각 점차 학습.

model



→ 학습된 결과가 필요해야 함.
(추가하면서 학습하는 방법)