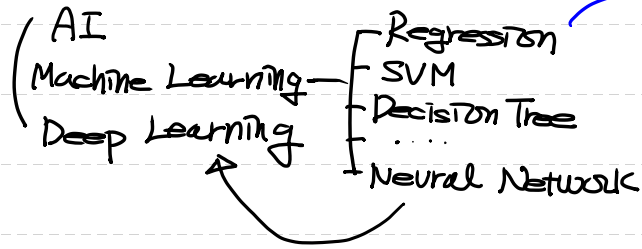


02/25



classical Linear Regression Model

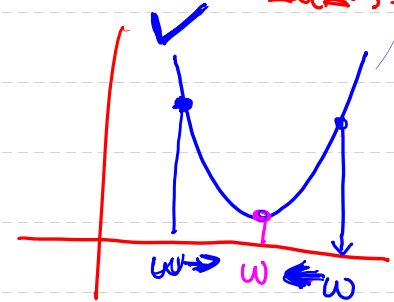
→ Simple Linear Regression (단순 선형 회귀)

↳ 독립변수가 1개..

(loss function을 반복적으로 최소화하면서
'w, b를 update' ⇒ 최소화)

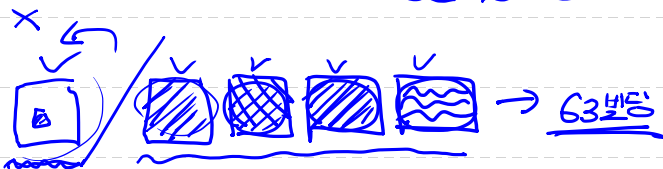
↳ Gradient Descent Algorithm
경사 하강법

loss function은 w와 b의 2차 함수.

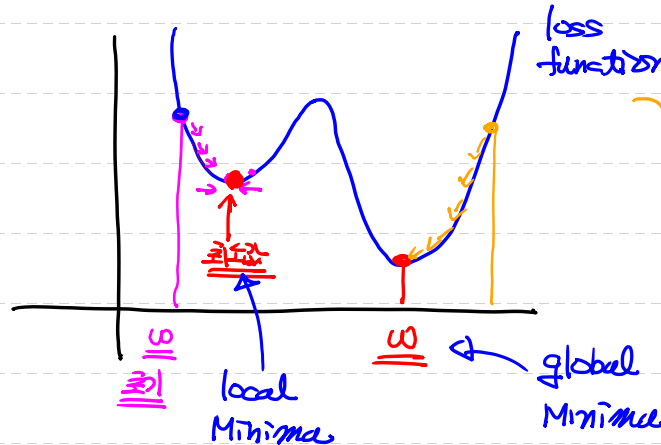


제공된 Data를 이용하여 Simple Linear Regression 구현

- ✓ python으로 구현
- ✓ Tensorflow로 구현
- ✓ sklearn으로 구현 (가장)



만약 여러의 loss function이



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

초기의 w가 어떤값이 선택되더라도
loss를 최소화할 수 있도록
중요한 역할이 있음

↓
볼록 함수
(convex function)

- 머신러닝에서 특성이 잘 되기 위해서는 "오류의 데이터"가 있어야 해요.

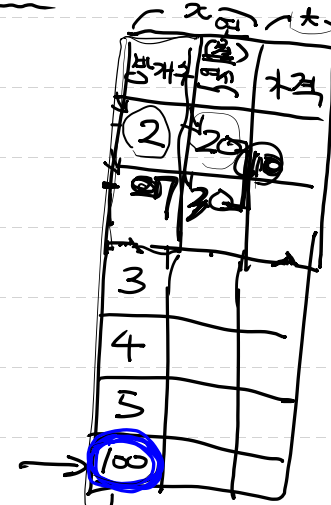
데이터 전처리 (Data preprocessing)

↑
"비데이터"

✓ outlier

이상치 처리

정제 (데이터 feature의 scale이 차이가 나는 경우)



↑
(feature)

(feature)
이상치 (outlier): 주성분에 들어있는 값이 일반적인 값에 비해
훨씬 큰 값을 가짐.

독립변수에 있는 이상치 → 기대점.

종속변수에 있는 이상치 → outlier (이상치)
(label)

이상치 검출 방법

→ 크아야지 처리할 수 있어요. → 방법

이상치 검출 방법

Variance: 분산도.

Likelihood: 메아크 관련 이용 (특종)

Nearest-Neighbor: 모델 데이터와의 거리

Density: 카이제곱분포 이용

....

+ (Tukey Fence
Z-Score)

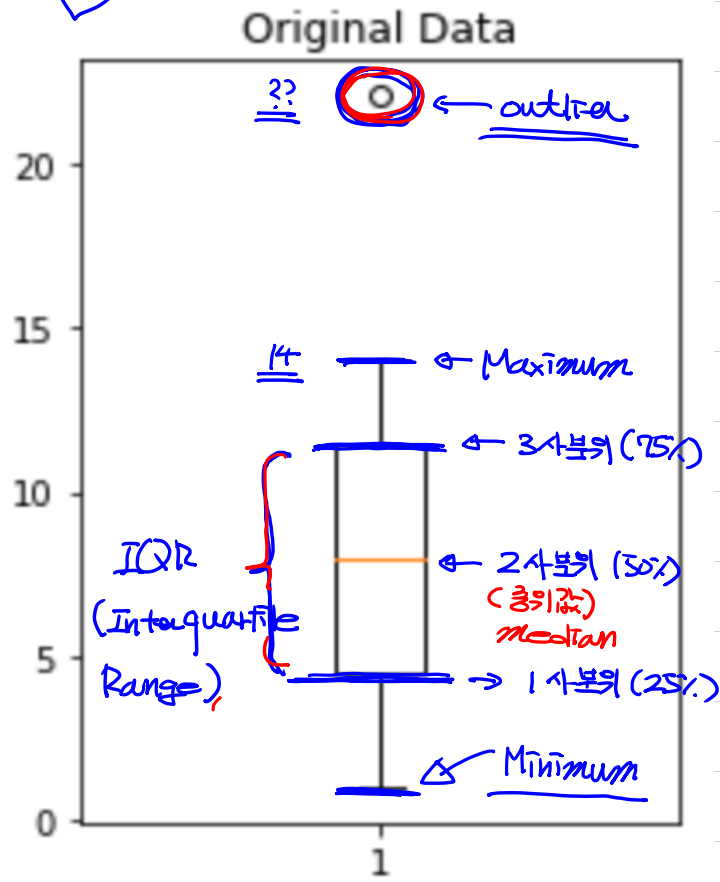
→ 사본

Boxplot을 이용하여

→ 이상치를 graph에
표시

→ 전처리 후 다시 표출된 것.

"Boxplot"



IQR value : 3사분위값 - 1사분위값 \Rightarrow 숫자값!!

(이 범위 \rightarrow 1사분위값 - (IQR value \times 1.5) 초기로 작은 값들
 \rightarrow 3사분위값 + (IQR value \times 1.5) 초기로 큰 값들)

* Tukey Fence

★ Z-Score를 이용한
이상치 구분

