

11/30

AI → 지도학습 (Supervised Learning)

→ Machine Learning

→ Regression / Linear Regression

(Model) (모델) ~ Classical Linear Regression Model

$$\Rightarrow \hat{y} = \beta_0 + \sum_{i=1}^n \beta_i x_i$$

$$y = \beta_0 + \beta_1 x_1$$

$$\check{y} = w x + b$$

$$\rightarrow Y = X \cdot w + b$$

Linear Regression Model의 구현방법

① Error(오류)

→ 각 데이터에 대해 [
Target
Label
결과값] 가 있는데

우리 모델이 계산해주는 값을 동양화 예측값



② loss 층

선형회귀기사용

MSE를 LOSS로
사용해요

오류 (Error)

x	t	y	Error
1	10	9	1
2	30	20	10
5	70	50	20
7	90	100	-10

LOSS는 동양화
직어야 좋아요

③ 경사하강법 (Gradient Descent) ~ Optimizer

$$\underline{\underline{w}} = \underline{\underline{w}} - \alpha \cdot \frac{\partial \text{loss}}{\partial \underline{\underline{w}}}$$

경사하강법

SGD $\rightarrow \begin{pmatrix} \dots \\ \dots \\ \dots \end{pmatrix} \alpha$

learning rate (학습률) \rightarrow 학습률

hyperparameter

이값을 고려하여 만들면
될거예요

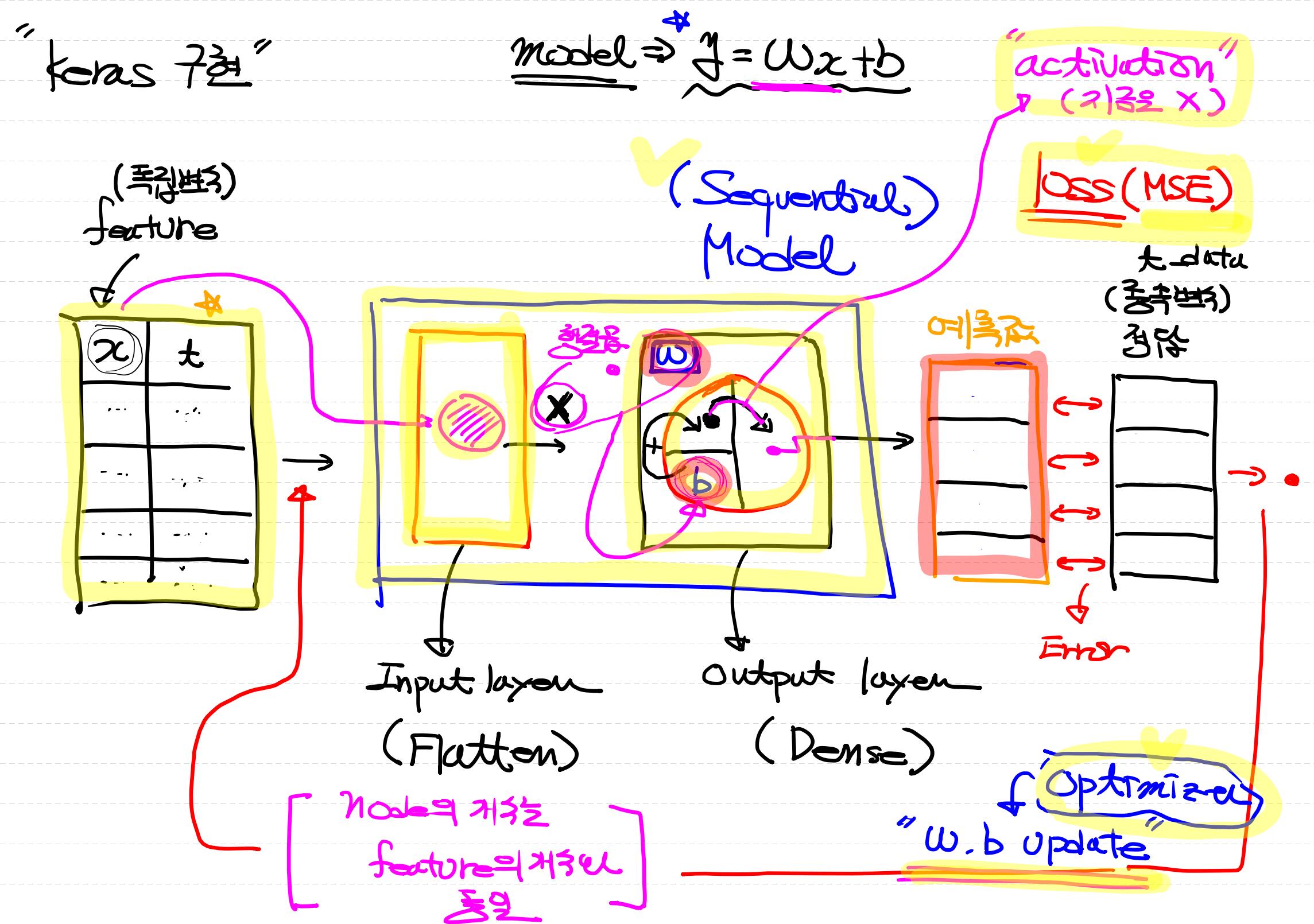
$$\underline{\underline{b}} = \underline{\underline{b}} - \alpha \cdot \frac{\partial \text{loss}}{\partial \underline{\underline{b}}}$$

구현해보아요!"

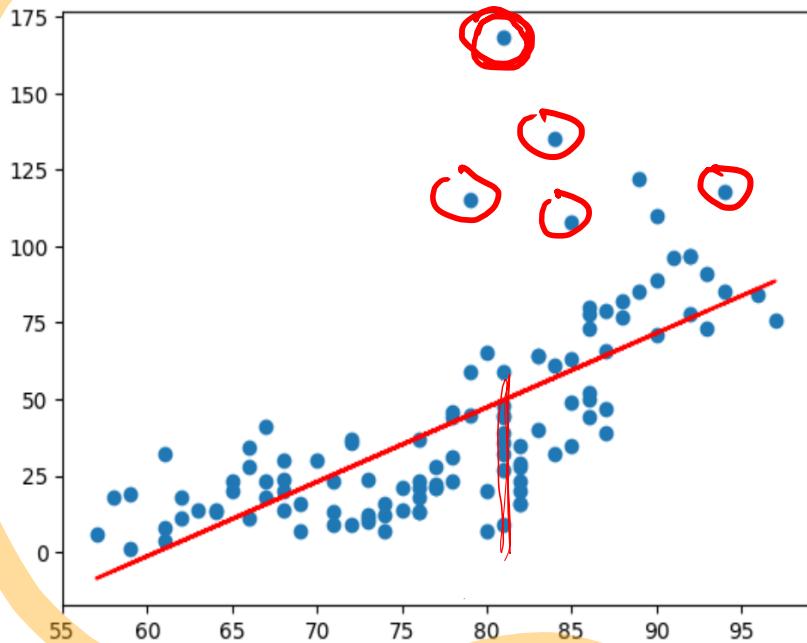
Python (어려워요)

Tensorflow keras

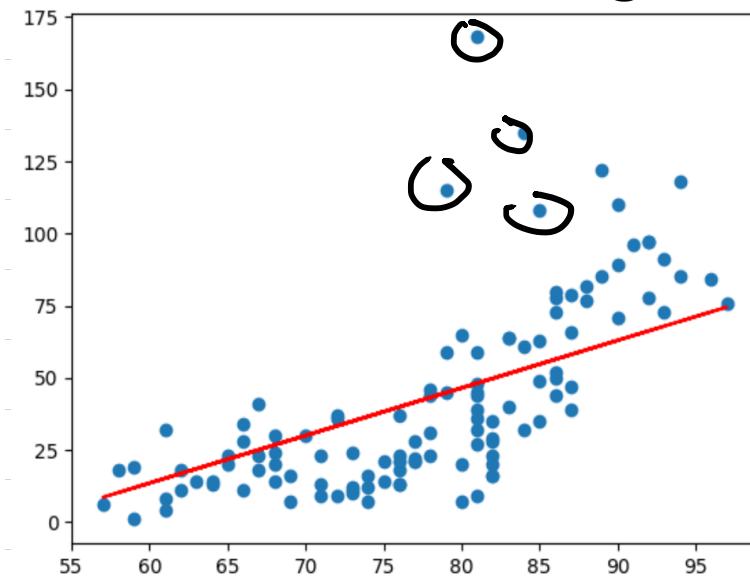
Sklearn



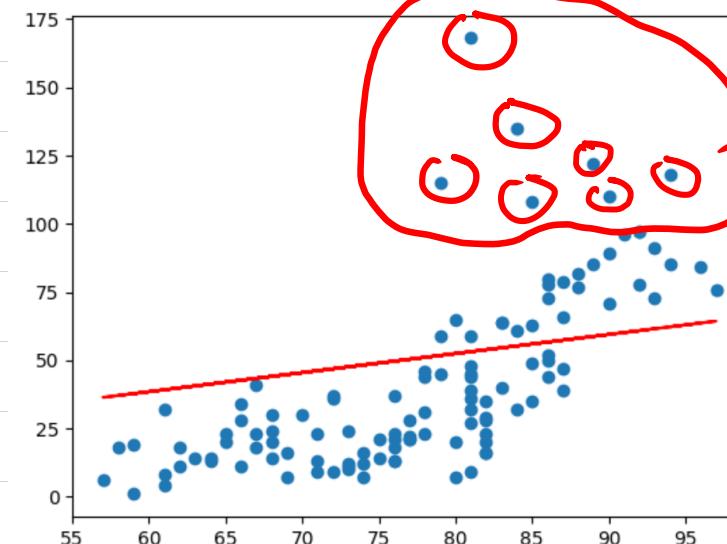
Sklearn 구현 ↗



python 구현 ↗



Tensorflow 구현 ↗



Why ??

이상치

• 뉴이뮨이 물건가 있나요!!

Regression → 조건부 통계! → 이상치가 있으면 좋지 않아요!!

④ 이상치

→ 일반적인 값이 비례 통계가 흐트리다!

↳ 온도변수 (Temp) → 여기서 발생하는 통계가 크면 → 이상치
↳ 층속변수 (Ozone) → 여기서 발생하는 통계가 크면 → 이상치

이상치는 어떻게 검출(코드)하나요?

Outlier

↳ 빈도수 많이 있나요 → 이중 2개만 알아두면 되요!!

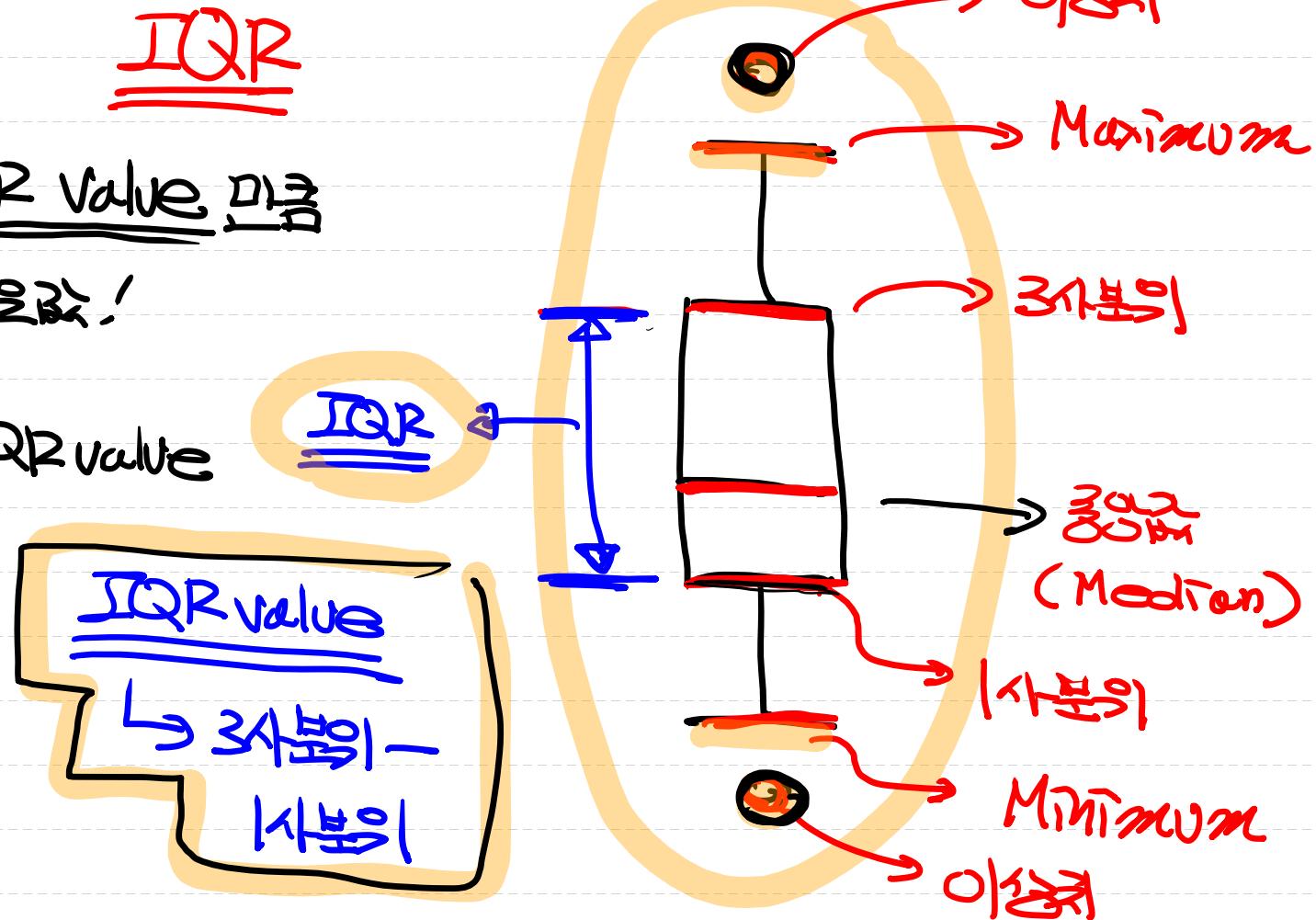
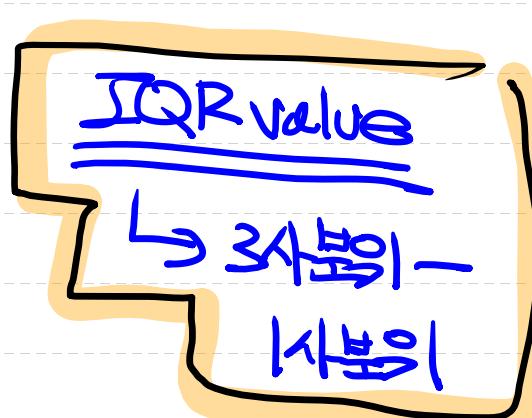
- Tukey's Fence ~~방법~~ → 사분위를 이용
- 정규분포 (Z -score)

Boxplot을 이용하여
우리 data를 사분위를
기준으로 그래프로 표현

- 사분위 보다 $1.5 \times \text{IQR}$ value 많음
로고로 적용!

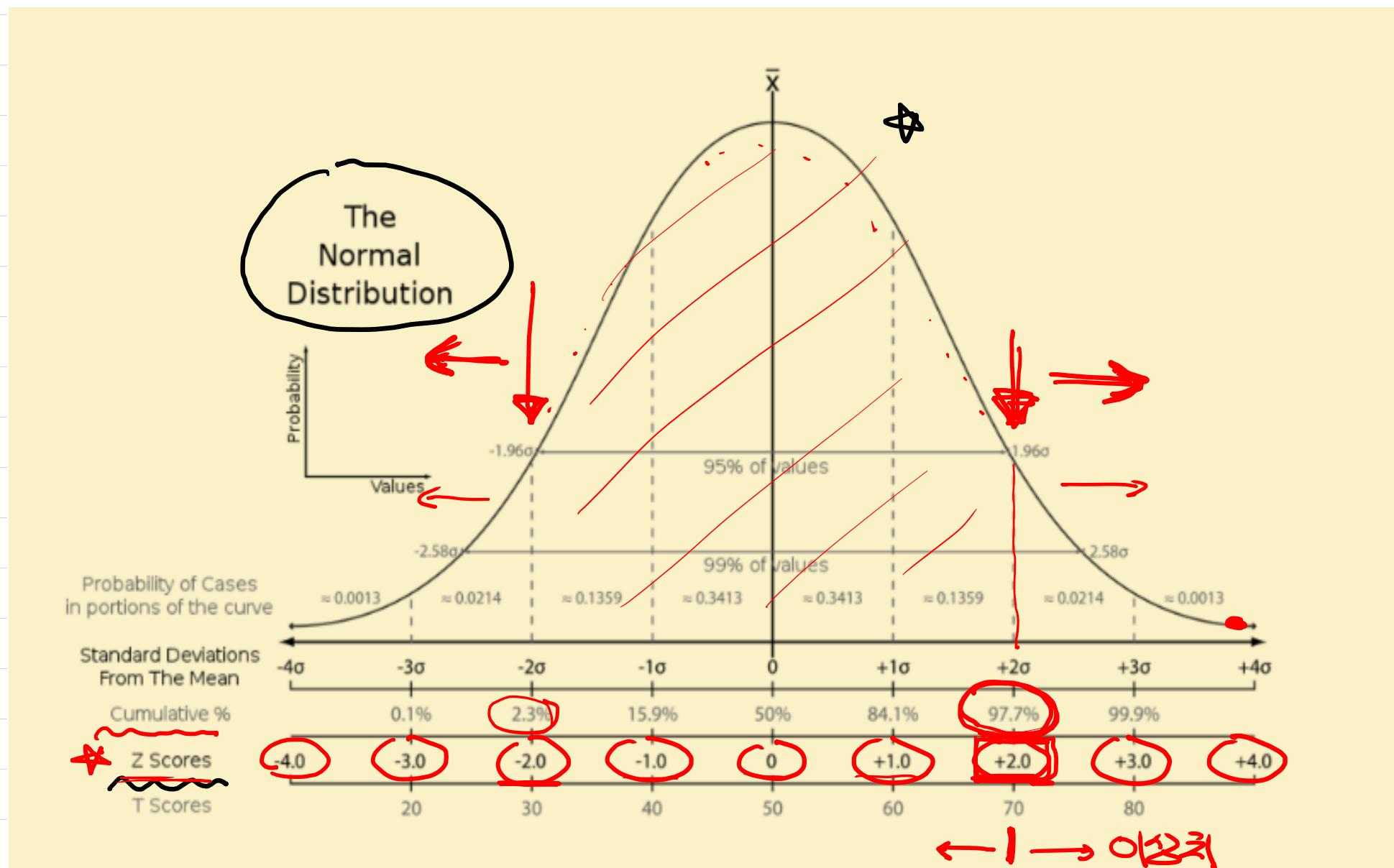
- 3사분위 보다 $1.5 \times \text{IQR}$ value

모듈 초기화
 \Rightarrow





• 0/231 그림 → 정규분포의 Z-Score 를 이용

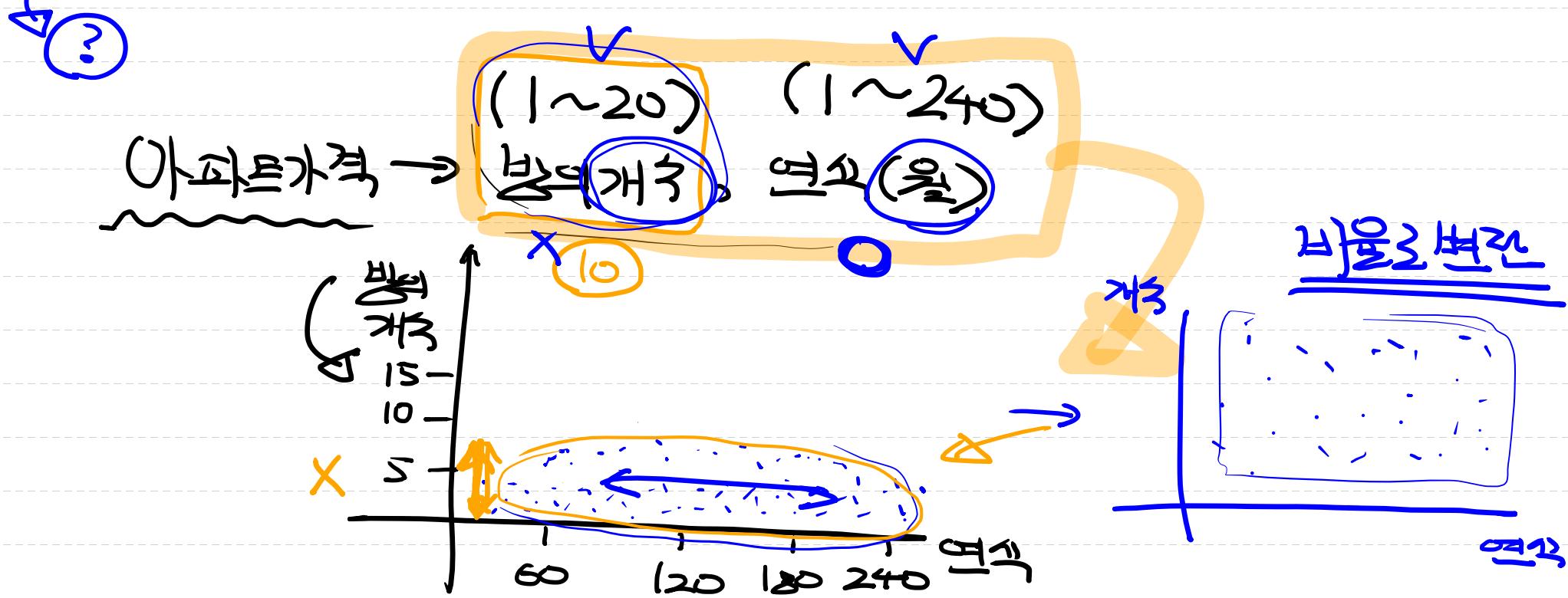


① 이상치도 처리했는지

우리가 원하는 Model이 모델이거나
있어야요!!

Why?

정규화를 이해하세요!! (Normalization)



$$O_2 \text{conc} \rightarrow 0 \text{ C/m}^3$$

$$\text{태양광} \rightarrow \underline{\underline{lux}}$$

$$\begin{array}{l} \text{바람} \rightarrow m^2/s \\ \text{온도} \rightarrow ^\circ C \end{array}$$

데이터 조작

① 결측치 처리

② 이상치 처리

③ 정규화 처리



정규화를 해야하는
이유

- ① feature scale 조정
- ② 특성 속도의 흐름
- ③ 특정 feature가 가중치가 되어 부여되는
Overfitting을 피할 수 있어요
- ④ 주제 유행성(제작의 종목도가 높아져요)
- ⑤ 여러가지 알고리즘들이 있는데 이들을
풀어 !!

정규화가 중요

(Normalization)



① Min-Max Normalization

How?

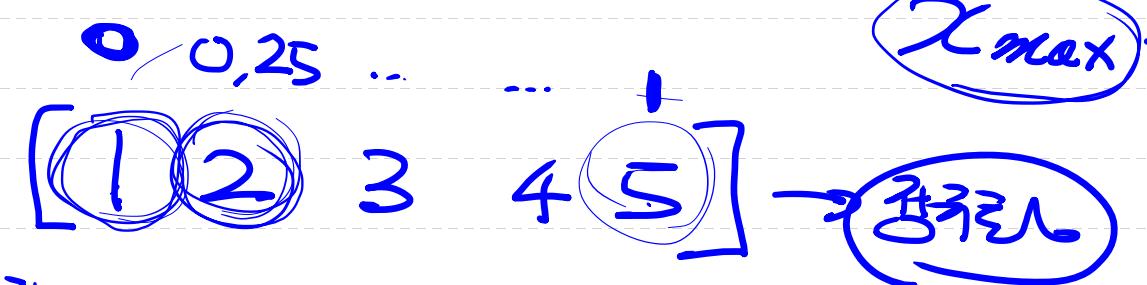
(Scaling)

정규화를 하는 가장 일반적인 방법

모든 feature가 대체 (최소 0)
(최대 1)



$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$



최소 → 1

최대 → 5

$$1 \rightarrow \frac{1 - 1}{5 - 1} = \frac{1}{4}$$
$$\Rightarrow 0$$