

- 04/08 Tensorflow 2.x 7월
- * Linear Regression
↓
Logistic Regression
↓
Multinomial Classification

- Linear Regression

→ Ozone.csv

↓
결측치 처리

① 삭제

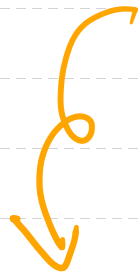
② Imputation (결측치 보간)

동일변수 + 종속변수

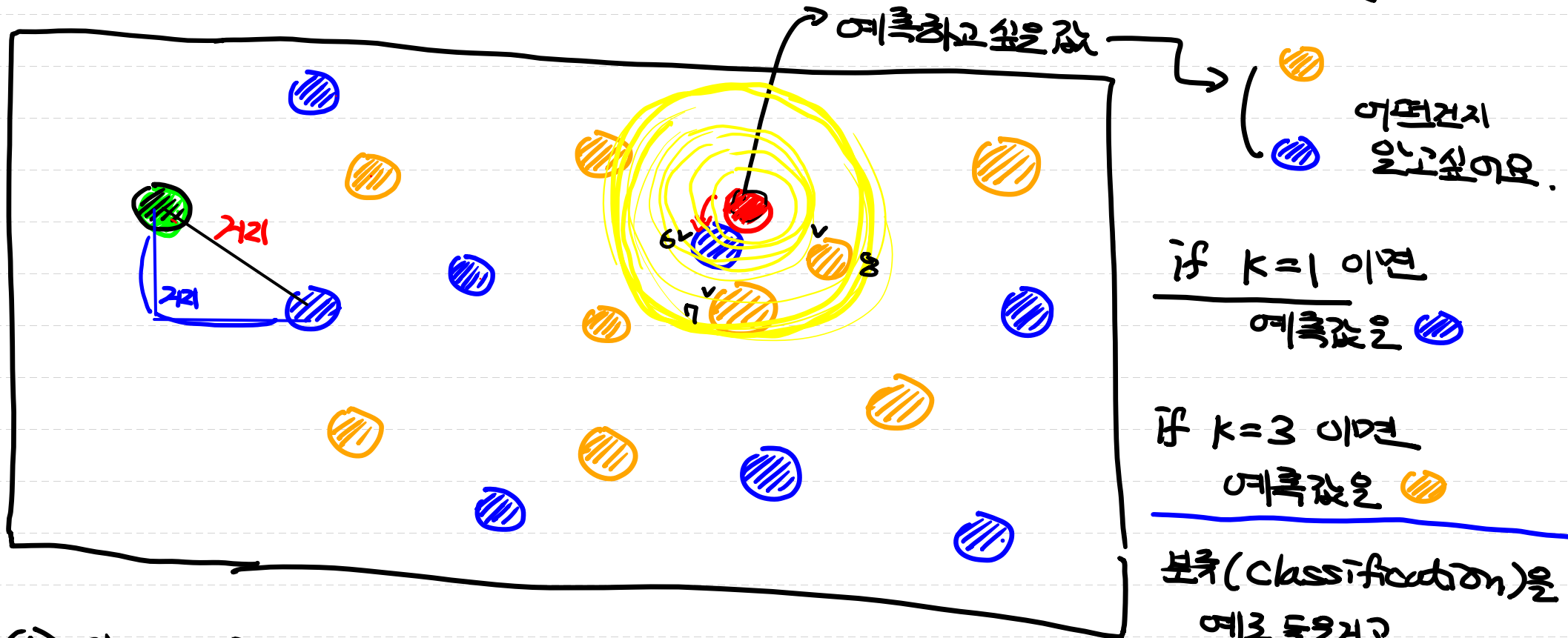
평균기법 (평균, 중앙값, 최빈값, 최대, 최소)
머신러닝기법

KNN

↓
종속변수를 대상



① KNN (k-Nearest Neighbors) → k-최근접 이웃 이라는
Machine Learning 알고리즘



① 마땅치 않음이라고 부를 수 있는 과정이 있습니다!!

[Instance-based Learning (Lazy Model)
Model-based Learning
↳ Regression

② KNN은 2개의 hyperparameter를 사용

① k

거리 측정 방법

↳ 유클리디언 distance
↳ 맨하튼 distance
↳ 마홀라노비스 distance

① $k=1$ 일때 어느정도의 성능을 보장!!

② 반드시 전처리를 진행.

③ 단점도 시간이 오래 걸릴 수 있어요. → * 모든 데이터에 대해 거리를 계산해야 해요!!
↳ 현재는 속도가 개선

Sklearn 으로 구현하고 넘어가요!!

→ Ozone 문제로 sklearn 데이터 처리부터 해보아요!!

- ★
 - ① 독립변수 (Solar.R, Wind, Temp)에 대한 Missing value는 median으로 처리
 - ② 독립변수의 이상치를 찾아내기 mean으로 처리
 - ③ 종속변수의 이상치를 찾아내기 mean으로 처리
 - ④ 전처리 진행
 - ⑤ 종속변수에 대한 Missing value는 KNN을 이용해 Imputation

* ④ 4/11(수) \Rightarrow machine Learning 필강편지
<오포북, 60점 이상>

④ 4/17(일) \Rightarrow machine Learning 수행평가 지킴
(구글 드라이브에 수행평가 내용 공유되어 왔어요)