• 03/08

Linear Regression
↓
Logistic Regression
↓ sigmoid Activation function
↓ Softmax Activation Function

Data Preprocessing (데이터 전처리)

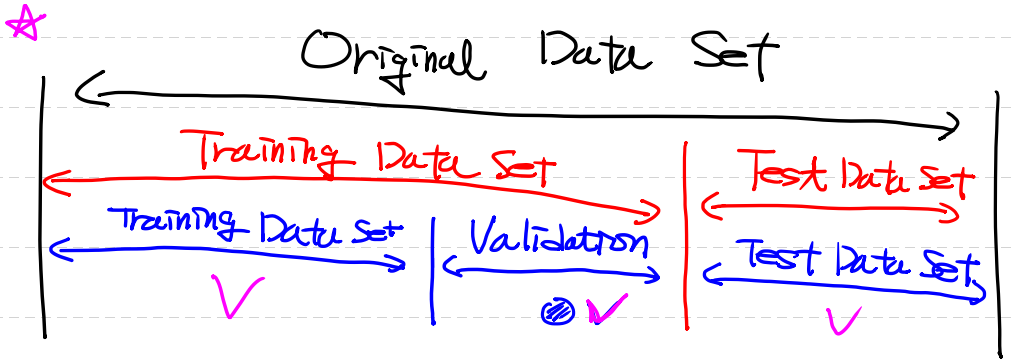
① 머신러닝에 입력될 데이터 → 결측처리 / 이상치처리 / 정규화처리 (Normalization)
- Min-Max Scaling (이상치에 민감) → 0~1 사이의 실수로
- Standardization (표준화) → 동일한 척도로 Scaling 출수없어요. (상대적으로 이상치에 둔감)
- Student's T 분포

② Model을 만들어요.
- python ~> 속도가 느려요 (x)
- Sklearn ~> 전반적으로 !! (O)
- tensorflow
- pytouch
  - 1.x 버전으로 공부
  - 2.x 구현할예정.

③ Evaluation (평가) : Metric
① Precision
② recall
③ *** accuracy
④ F1
⑤ Fall-out

• Training Data Set을 이용해서 Accuracy를 측정하면 안되요!

- 성능평가 Data Set



Training Data Set: 학습용도

Test Data Set: 최종 성능 평가.

Validation Data Set: 모델 개선 평가의 Test용 data set

- Data Set이 충분히 많지 않을 경우 ⓞ

→ Training Data Set 잘라올 수밖에 없어요 ⇝ Underfitting (과소적합)

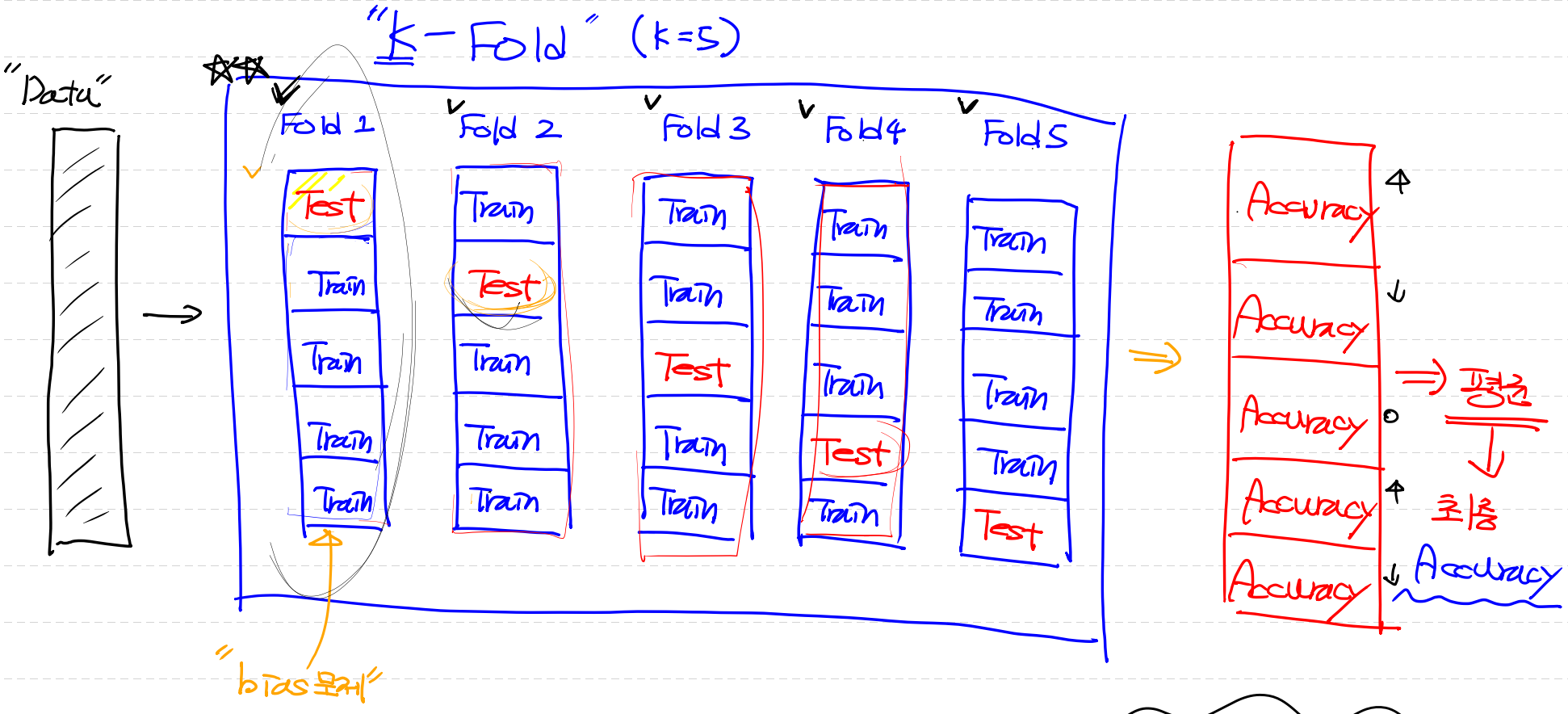
↓

epoch을 늘려서 학습 → 우리 Model이 해당 data만 잘 표현하고 일반적인 data에 대해서는 잘 적용이 안되는 현상 (과대적합) → Overfitting

"k-fold Cross Validation" (CV)

"교차검증"

↓



구현 ∞ → "BMI 예제" · [전처리 / 모델생성 / accuracy]