

14강. 전처리



전처리 (pre-processing)

데이터 분석에 적합하게 데이터를 가공/ 변형 /처리/클리닝

Garbage in, Garbage out!



전처리 (pre-processing)

분석가의 **80%** 시간을 데이터 수집 및 전처리에
사용한다고 합니다

전처리 (pre-processing)

1. 결측치 – Imputer
2. 이상치
3. 정규화 (Normalization)
4. 표준화 (Standardization)
5. 샘플링 (over/under sampling)
6. 피처 공학 (Feature Engineering)
 - feature 생성/ 연산
 - 구간 생성, 스케일 변형..

정규화, 표준화

정규화 (Normalization)

- 0~1사이의 분포로 조정

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

표준화

- 평균을 0, 표준편차를 1로 맞춤

$$x_{new} = \frac{x - \mu}{\sigma}$$