

표준화 & 정규화

표준화 & 정규화가 필요한 이유

- 데이터들을 평가하기 때문에 비교가 편하다

- 토익(990점) 시험(100점)처럼 측정 특성의 단위도 다르고 범위도 달라 직접적으로 비교가 힘들다.
- scale 범위가 너무크면 노이즈 데이터가 생성되거나 overfitting이 될 가능성이 높다.
- 머신러닝에서 scale이 큰 feature의 영향이 비뚤어지는 것을 방지
- 딥러닝에서 Local-Minimum에 빠질 위험감소 (학습속도 향상)

정규화 Normalization

- 요약

- 값의 범위를 0~1사이로 바꿔주는 것
- scikit-learn ➡ MinMaxScaler

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

표준화 Standardization

- 요약

- 값의 범위를 평균 '0' 분산 '1'이 되도록 바꿔준다
- 정규분포를 표준정규분포로 변환하는 것과 같다.
- scikit-learn ➡ StandardScaler

$$\frac{x - \mu}{\sigma}$$