



7. 데이터 스케일링을 해야 하는 이유



1. 데이터 스케일링을 해야 하는 이유

데이터 스케일링이란?

데이터 전처리 과정

데이터 특성의 범위 또는 분포를 같게 만드는 작업

1. 데이터 스케일링을 해야 하는 이유

데이터 스케일링이란?

x_1 , x_2 특성이 있을 때

x_1 은 0부터 1 사이의 값을 갖고,

x_2 는 100000부터 10000000000000 사이의 값을 갖는다고 한다면?

1. 데이터 스케일링을 해야 하는 이유

StandardScaler

```
from sklearn.preprocessing import StandardScaler
```

모든 특성을 평균을 0으로 표준편차를 1로 변환

이상치에 민감

1. 데이터 스케일링을 해야 하는 이유

MinMaxScaler

```
from sklearn.preprocessing import MinMaxScaler
```

특성 중 가장 작은 값을 0, 가장 큰 값을 1로 변환하여 0 ~ 1 사이의 값으로 만들
이상치에 매우 민감

1. 데이터 스케일링을 해야 하는 이유

RobustScaler

```
from sklearn.preprocessing import RobustScaler
```

중앙값과 사분위 값을 사용하여 중앙값을 빼고, 사분위 값으로 나눔
이상치의 영향을 최소화할 수 있음