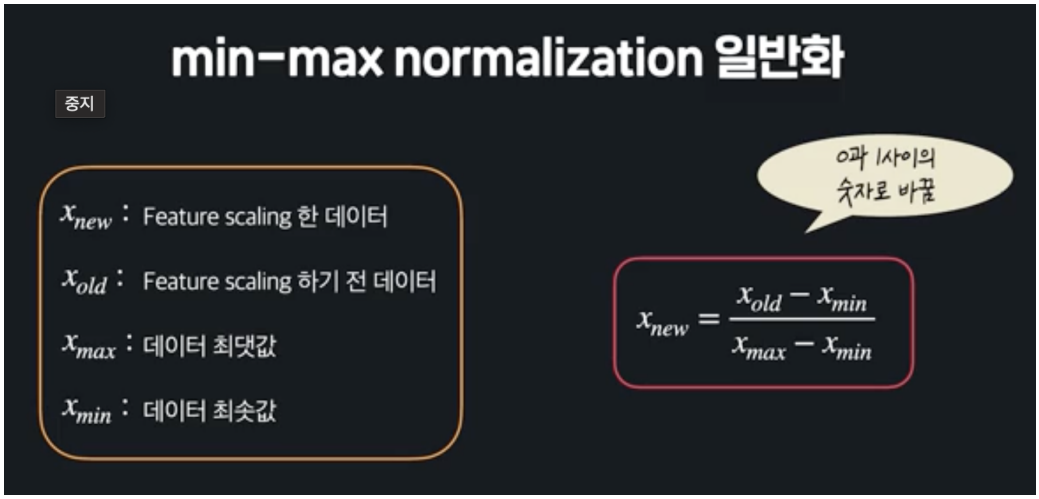
데이터 전처리

- 데이터 전처리
  - 데이터를 그대로 사용하지 않고, 가공해서 모델을 학습 시키는데 좀 더 좋은 형식으로 만들어주는 것
- Feature Scaling
  - 머신 러닝 모델에 사용할 입력 변수들의 크기를 조정해서 일정 범위 내에 떨어지도록 바꾸는 것
  - 경사 하강법을 좀 더 빨리할 수 있게 도와준다
  - 방법
    - min-max normalization
      - 데이터의 최소, 최댓값을 이용해서 데이터의 크기를 0과 1사이로 바꿔준다



        - 선형 회귀뿐만 아니라 경사 하강법을 사용하는 모든 알고리즘의 속도를 빠르게 해준다.
    - 정규화 또한 가능
- 머신 러닝에서 사용하는 데이터
    - 수치형 데이터 : 나이, 몸무게, 키
    - 범주형 데이터 : 혈액형, 성별 (→수치형 데이터로 바꿔줘야!)
        - One-hot Encoding
            - 각 카테고리의 열을 새로운 열로 만들어주는 방법