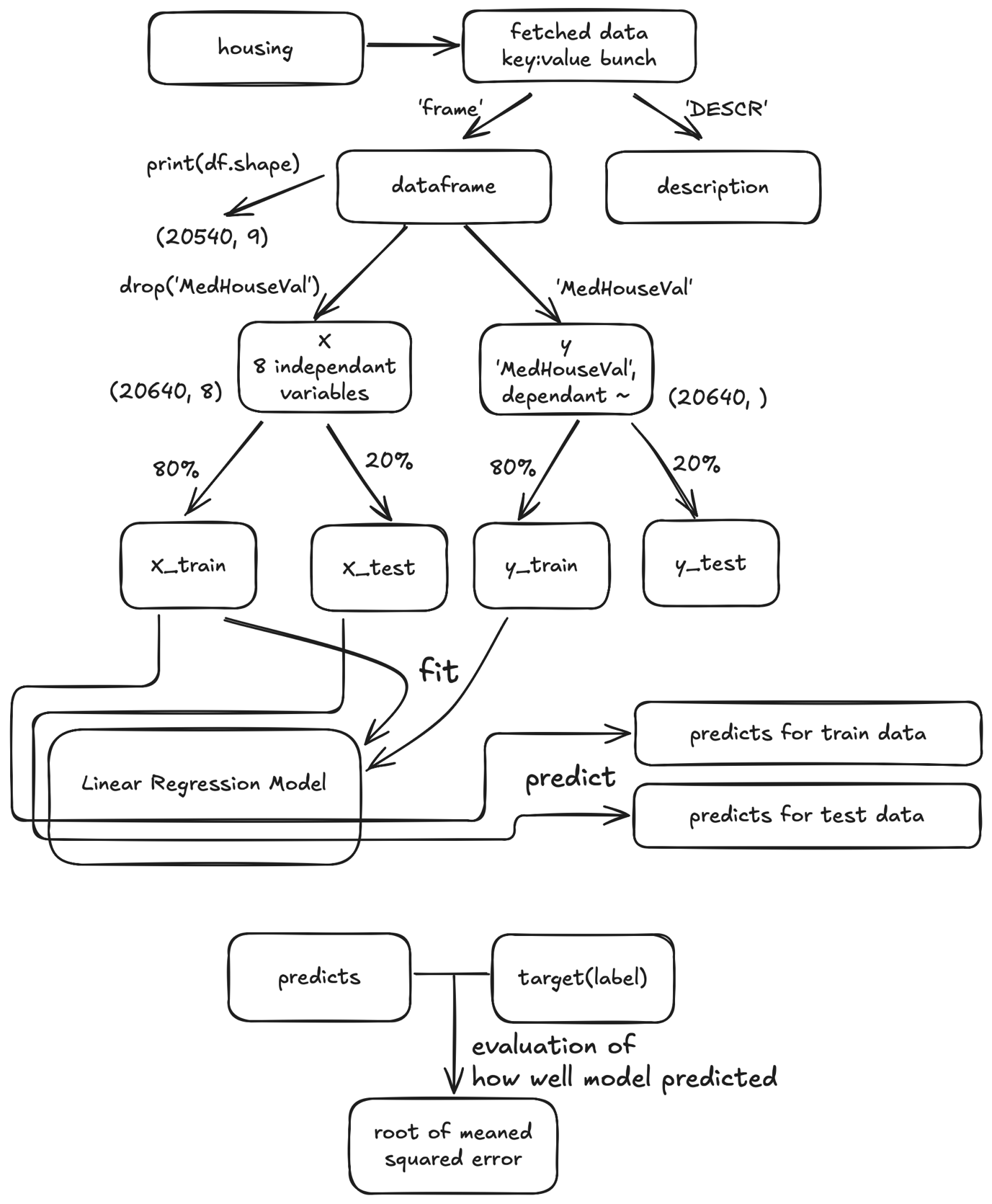
1주차 결과 보고서

전공: 컴퓨터공학 학년: 2학년 학번: 20221559 이름: 박준우

1. 

1) dataset을 fetch해온다.

2) fetch한 데이터 중 dataframe인 ‘frame’항목의 값을 가져다 사용하고, 설명을 출력해본다.

3) pandas.dataframe 클래스의 클래스 변수 shape를 한번 출력해본다.

4) label이 될 ‘MedHouseVal’ 열과 나머지 8개의 열을 분리한다.

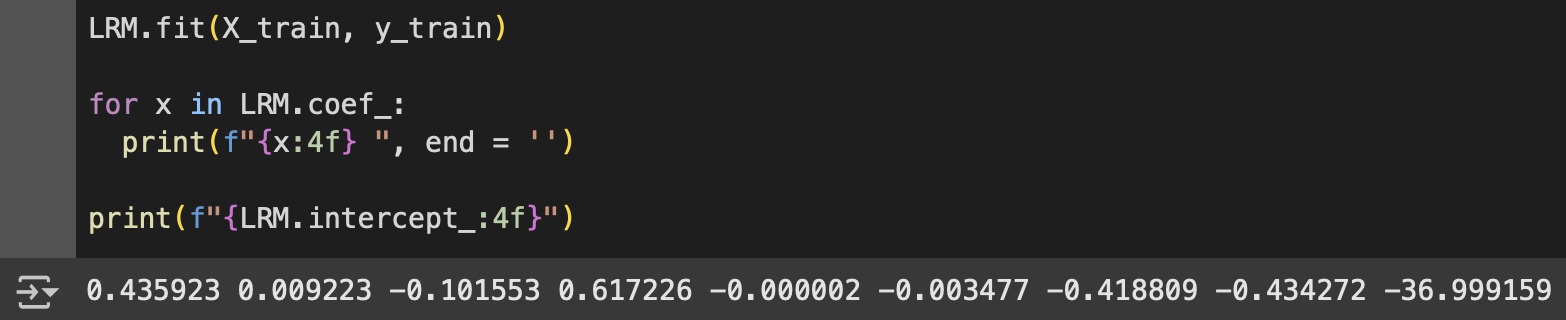
5) 학습용 데이터, 평가용 데이터를 80:20 비율로 분할한다.

6) LinearRegression모델을 하나 정의한 후 학습데이터로 학습시킨다.

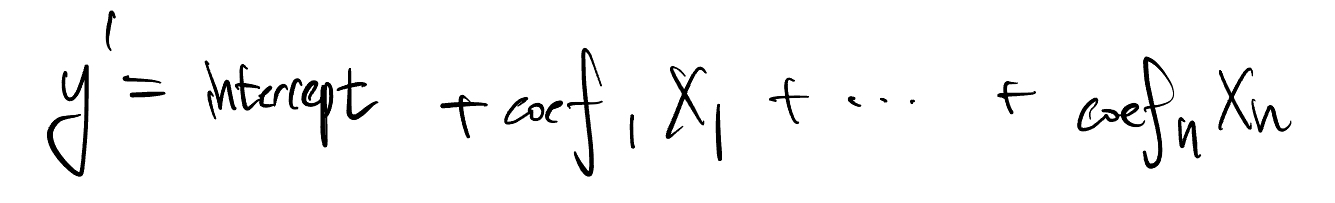
7) 추론한 결과와, 정답의 RMSE를 구해서 모델의 성능을 평가한다.

라이브러리에 전부 구현되어 있는 것을 사용했고, 어떤 기능을 수행했는지 서술했다.

2.



학습 시킨 데이터의 열이 8개이기 때문에, 8개의 계수와 하나의 상수가 나오는 것을 알 수 있다.



모델은 다음과 같이 찾아낸 intercept (y절편, 상수) + 각각 찾아낸 coefficient(상수)와 각 데이터 수치의 곱의 합으로 예상치를 구한다.

3. RMSE는 rooted mean square error인데, 이는 평균 오차의 크기를 측정하는 지표로, 예측값과 실제값 간의 평균적인 절대적 오차의 크기를 측정한다고 볼 수 있다.

해당 데이터에서의 label은 1에 10만 달러이기 때문에 0.7 정도라고 함은 예측한 가격과 실제 가격의 차이가 평균적으로 7만 달러 정도임을 나타낸다. 이는 한화로 약 1억이니, 정말 큰 차이라고 볼 수 있을 것 같다.