



# 모델 평가와 하이퍼파라미터 고르기

- k겹 교차 검증

- k : 데이터 셋을 몇 개로 나눌 건지 정하는 변수 - 일반적으로 5

데이터 셋을 k개로 나눈다,  $k = 5$ ,  $m = 1000$

집 크기 (평)	집 값 (억)
데이터 200 개	성능: 80%
데이터 200 개	성능: 70%
데이터 200 개	성능: 85%
데이터 200 개	성능: 85%
데이터 200 개	성능: 80%

→ 평균 성능: 80%

- 하이퍼 파라미터

- 학습을 하기 전에 미리 정해 줘야 하는 변수 또는 파라미터들
- `model = Lasso(alpha=0.01, max_iter=1000)`
  - 어떤 값을 넣어주냐에 따라 결과가 크게 달라짐
- Grid Search
  - 직관적인 방법
    - `model = Lasso(alpha=?, max_iter=?)`
    - 정해줘야 하는 각 하이퍼 파라미터에 넣어보고 싶은 후보 값을 몇 개씩 정함
    - 모든 후보 값의 조합으로 모델을 학습시켰을 때 성능이 가장 좋았던 하이퍼 파라미터 조합을 고름

	0.1	1	10
100 0	alpha=0.1, max_iter=1000 일 때 성능	alpha=1, max_iter=1000 일 때 성능	alpha=10, max_iter=1000 일 때 성능
200 0	alpha=0.1, max_iter=2000 일 때 성능	alpha=1, max_iter=2000 일 때 성능	alpha=10, max_iter=2000 일 때 성능
300 0	alpha=0.1, max_iter=3000 일 때 성능	alpha=1, max_iter=3000 일 때 성능	alpha=10, max_iter=3000 일 때 성능