**교차검증 - K-Fold와 Stratified K-Fold**

[교차 검증]

* 학습 데이터 세트
* 학습 데이터를 다시 분할하여 학습 데이터와 학습된 모델의 성능을 일차 평가하는 검증 데이터로 나눔.
* 학습 데이터 ---분할-🡪 학습 데이터 세트 | 검증 데이터 세트
* 테스트 데이터 세트
* 모든 학습/검증 과정이 완료된 후 최종적으로 성능을 평가하기 위한 데이터 세트

[K 폴드 교차 검증]

* K개의 폴드 세트에 K번의 학습과 검증 평가 반복 수행 (K: 폴드 교차 검증 횟수)
* 교차 검증 최종 평가 = (검증 평가 전체의) 평균
* 일반 K 폴드
* Stratified K 폴드
* 불균형한 분포도를 가진 레이블(결정 클래스) 데이터 집합을 위한 K 폴드 방식
* 학습 데이터와 검증 데이터 세트가 가지는 레이블 분포도가 유사하게 검증데이터 추출

|  |
| --- |
| kfold = KFold(n\_splits=5) |

5개의 폴드 세트로 분리하는 K폴드 객체 생성 (기본은 3개 분리)

|  |
| --- |
| kfold.split(features) // features=iris.data |

폴드 별 학습용, 검증용 테스트의 로우 인덱스를 array로 반환

|  |
| --- |
| skfold.split(features, label) |

StratifiedKFold의 split( ) 호출 시 반드시 레이블 데이터 셋도 추가 입력 필요