

사이킷런

- 머신러닝 라이브러리 중 많이 사용됨

딥러닝 라이브러리의 등장

- 사이킷런 인기 하락

분류(classification)은 대표적 지도학습(supervised Learning)

지도학습

- 명확한 정답이 주어진 데이터로 학습한 뒤 미지의 값 예측

하이퍼 파라미터

- 머신러닝 알고리즘별 최적의 학습을 위해 직접 입력하는 파라미터

의사결정트리 알고리즘(Decision Tree)

학습용데이터, 테스트용 데이터 분리 필수

지도학습의 모든 알고리즘을 구현한 클래스(Estimator)

사이킷런 데이터세트는 딕셔너리 구조

학습을 위한 일정한 데이터의 양 보잡

다양한 데이터 기반 예측 성능 평가

과적합(Overfitting)

교차검증

- K폴드 교차검증 (5등분 교차검증)
- stratified K 폴드