

사이킷 런(Scikit-learn) 소개

2007년도 구글 하계 코드 프로젝트 모임에 참여한 몇몇 개발자들이 중심이 되어 시작된 라이브러리

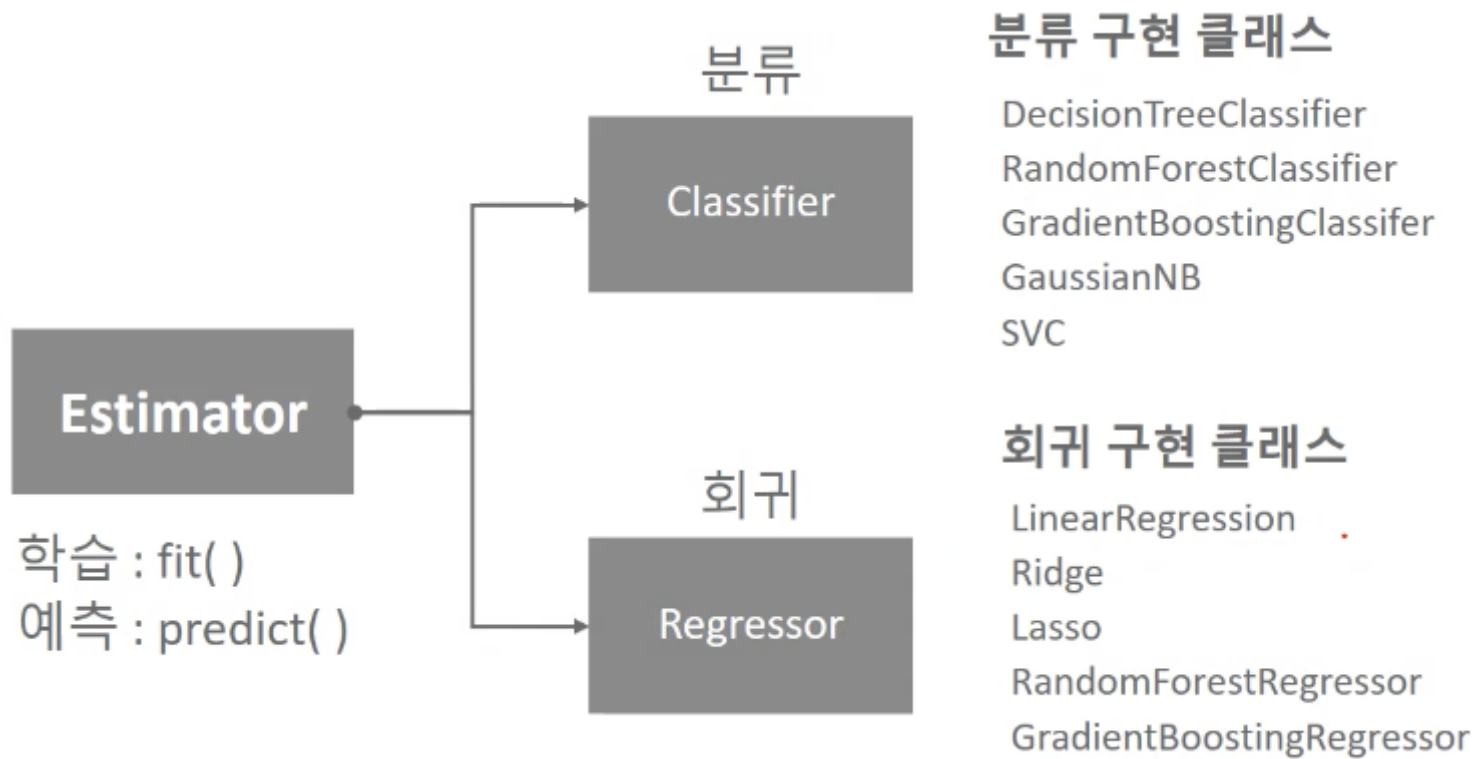
지도학습, 비지도학습을 위한 다양한 모델을 제공

이 모델을 위한 시각화 도구, 교차 검증 도구들까지 매우 광범위한 기능을 제공

하지만 딥러닝은 제공 안함

#01. sklearn의 핵심 기능

1) 주요 기능 분류



2) 주요 하위 모듈

분류	모듈명	설명
예제 데이터	sklearn.datasets	사이킷런에 내장되어 예제로 제공하는 데이터 세트
데이터 분리, 검증 & 파라미터 튜닝	sklearn.model_selection	교차 검증을 위한 학습용/테스트용 분리, 그리드 서치(Grid Search)로 최적 파라미터 추출 등의 API 제공
피처 처리	sklearn.preprocessing	데이터 전처리에 필요한 다양한 가공 기능 제공(문자열을 숫자형 코드 값으로 인코딩, 정규화, 스케일링 등)
	sklearn.feature_selection	알고리즘에 큰 영향을 미치는 피처를 우선순위 대로 선택 작업 수행하는 다양한 기능 제공
	sklearn.feature_extraction	<p>텍스트 데이터나 이미지 데이터의 벡터화된 피처를 추출하는 데 사용됨.</p> <p>예를 들어 텍스트 데이터에서 Count Vectorizer 나 Tf-Idf Vectorizer 등을 생성하는 기능 제공.</p> <p>텍스트 데이터의 피처 추출은 sklearn.feature_extraction.text 모듈에, 이미지 데이터의 피처 추출은 sklearn.feature_extraction.image 모듈에 지원 API가 있음.</p>
피처 처리 & 차원 축소	sklearn.decomposition	차원 축소와 관련한 알고리즘을 지원하는 모듈임. PCA, NMF, Truncated SVD 등을 통해 차원 축소 기능을 수행할 수 있음

분류	모듈명	설명
평가	sklearn.metrics	분류, 회귀, 클러스터링, 페어와이즈(Pairwise)에 대한 다양한 성능 측정 방법 제공 Accuracy, Precision, Recall, ROC-AUC, RMSE 등 제공
ML 알고리즘	sklearn.ensemble	앙상블 알고리즘 제공 랜덤 포레스트, 에이다 부스트, 그래디언트 부스팅 등을 제공
	sklearn.linear_model	주로 선형 회귀, 릿지(Ridge), 라쏘(Lasso) 및 로지스틱 회귀 등 회귀 관련 알고리즘을 지원. 또한 SGD(Stochastic Gradient Descent) 관련 알고리즘도 제공
	sklearn.naive_bayes	나이브 베이즈 알고리즘 제공. 가우시안 NB, 다항 분포 NB 등.
	sklearn.neighbors	최근접 이웃 알고리즘 제공. K-NN 등
	sklearn.svm	서포트 벡터 머신 알고리즘 제공
	sklearn.tree	의사 결정 트리 알고리즘 제공
	sklearn.cluster	비지도 클러스터링 알고리즘 제공 (K-평균, 계층형, DBSCAN 등)
유틸리티	sklearn.pipeline	피처 처리 등의 변환과 ML 알고리즘 학습, 예측 등을 함께 묶어서 실행할 수 있는 유틸리티 제공

#02. sklearn의 코드 패턴

```
#import
from sklearn.branch import model_name

#create instance
model = model_name()

#fit model
model.fit(X_train, y_train)
```

구분	설명
branch	사용하고자 하는 알고리즘이 구현된 클래스를 담고 있는 패키지
model_name	사용하고자 하는 알고리즘이 구현된 클래스
model	알고리즘에 따라 계산을 수행할 수 있는 객체

model 객체가 갖는 `fit()` 메서드에 독립변수(`x_train`), 종속변수(`y_train`)을 전달하여 독립변수와 종속변수간의 관계를 찾아내는 과정

#03. sklearn을 위한 선행 지식

Tensorflow등을 포함한 모든 머신러닝 패키지에 공통적으로 해당되는 사항

- Numpy
- Scipy
- Pandas
- Matplotlib
- Seaborn