머신 러닝 (Machine Learning) 기초

2 sklearn 이란?

scikit-learn 이란?

Python 대표
머신 러닝 라이브러리

“**데이터 준비 > 머신 러닝 학습 > 테스트 > 성능 평가 및 시각화** 지원”

“분류, 회귀, 군집 분석, 차원 축소 등 **머신 러닝 알고리즘이 모두 포함**”

“**어떤 알고리즘도 같은 방식으로 사용**이 가능”

“초심자 학습을 위한 **코드 샘플 및 데이터(토이) 세트 제공**”

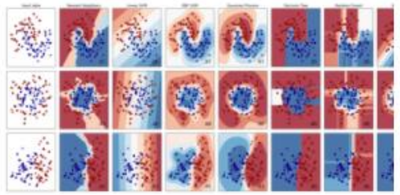
scikit-learn
공식 홈페이지

https://scikit-learn.org/stable/

Classification

Identifying which category an object belongs to.

Applications: Spam detection, image recognition.
Algorithms: SVM, nearest neighbors, random forest, and more...

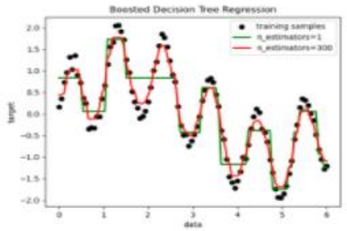


Examples

Regression

Predicting a continuous-valued attribute associated with an object.

Applications: Drug response, Stock prices.
Algorithms: SVR, nearest neighbors, random forest, and more...

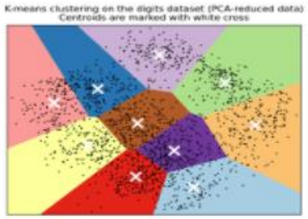


Examples

Clustering

Automatic grouping of similar objects into sets.

Applications: Customer segmentation, Grouping experiment outcomes
Algorithms: k-Means, spectral clustering, mean-shift, and more...

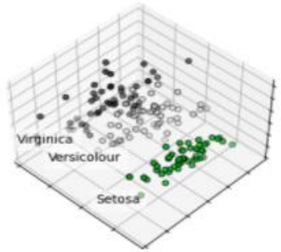


Examples

Dimensionality reduction

Reducing the number of random variables to consider.

Applications: Visualization, Increased efficiency
Algorithms: k-Means, feature selection, non-negative matrix factorization, and more...

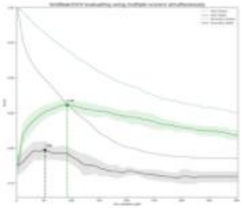


Examples

Model selection

Comparing, validating and choosing parameters and models.

Applications: Improved accuracy via parameter tuning
Algorithms: grid search, cross validation, metrics, and more...

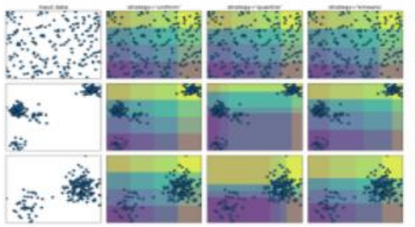


Examples

Preprocessing

Feature extraction and normalization.

Applications: Transforming input data such as text for use with machine learning algorithms.
Algorithms: preprocessing, feature extraction, and more...



Examples

pip install scikit-learn

from sklearn.model_selection import train_test_split

From sklearn.linear_model import LinearRegression

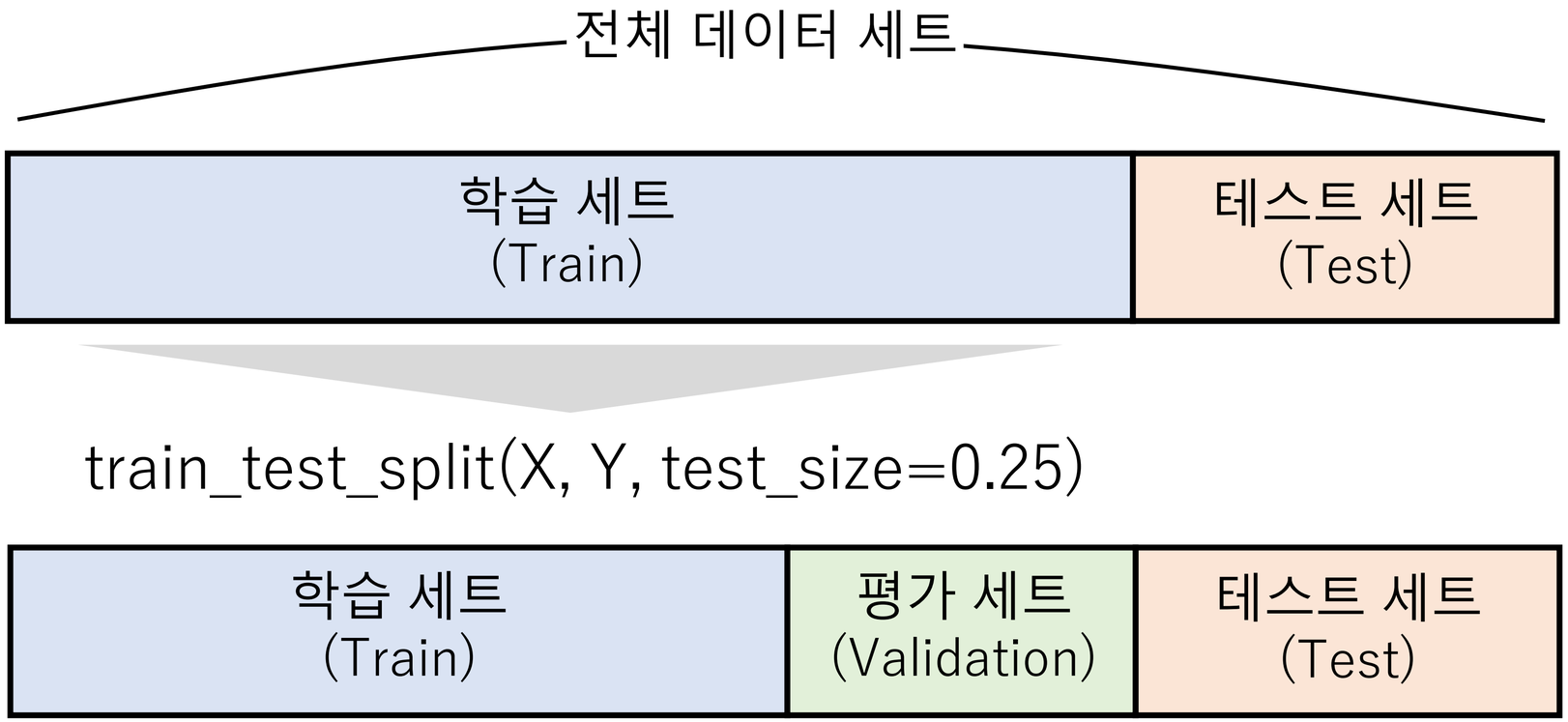
model = LinearRegression()

model.fit(X, Y)

predictions(Y') = model.predict(X')

train / test 세트 분리

모델의 학습, 평가, 테스트 용도에 따른 데이터 분리

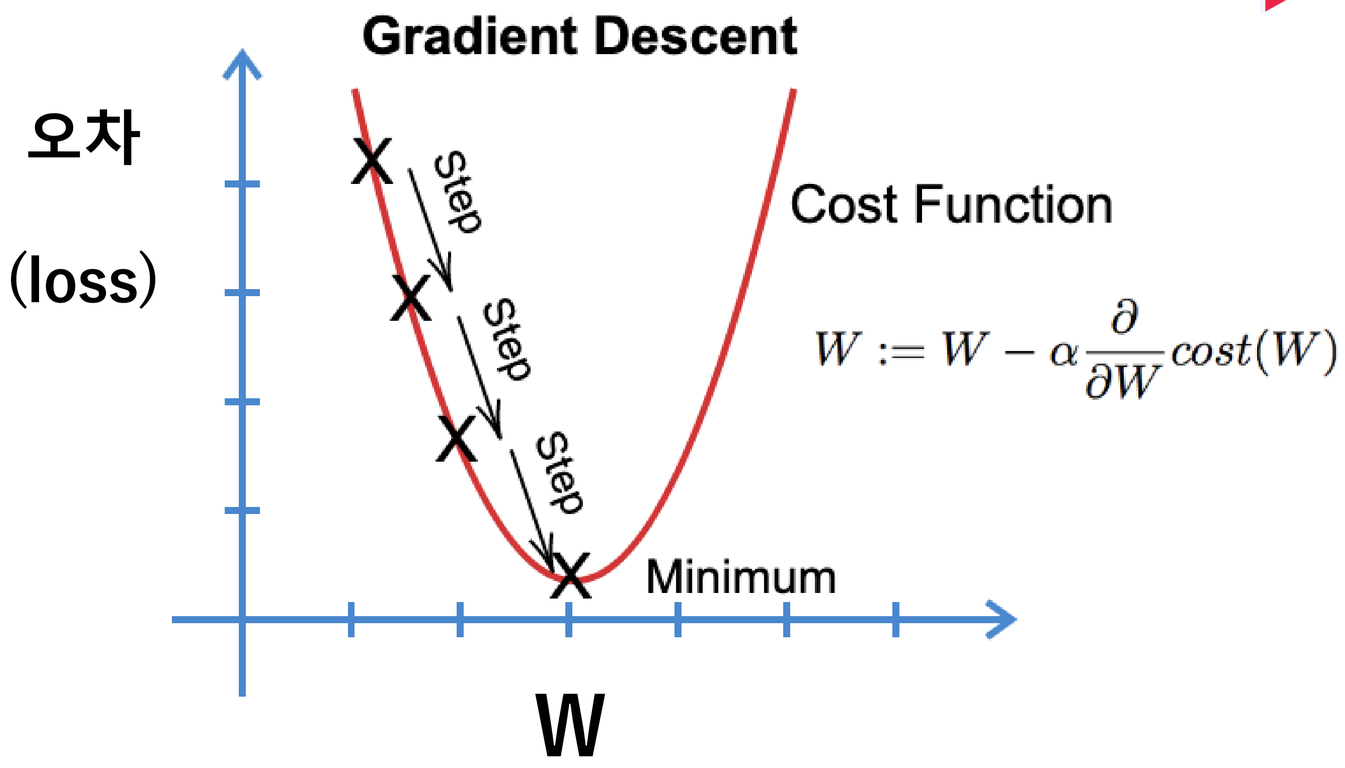


model.fit(X, Y) - 경사 하강법

오차가 가장 작은 최적 W 값을 찾는 방법

정답 입력

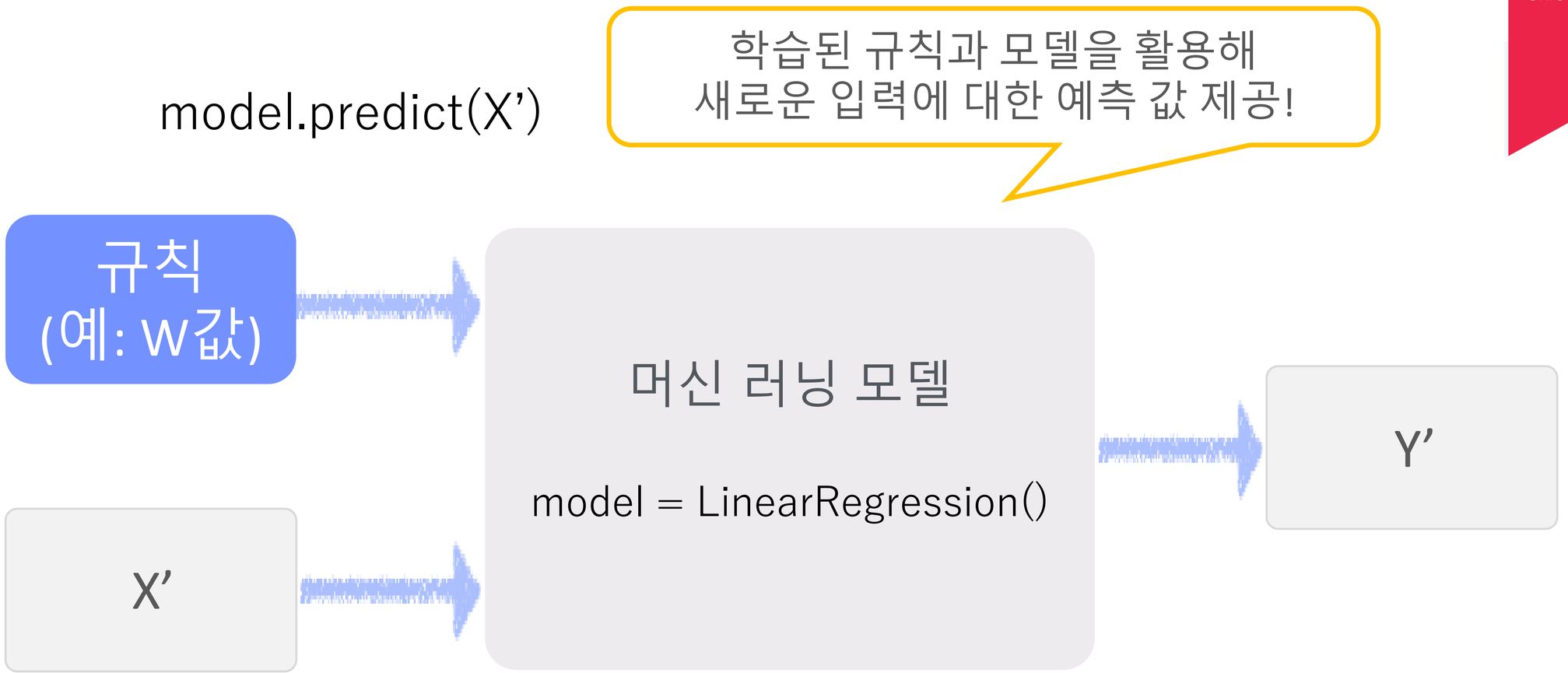
$$Y = WX + b$$



model.fit(X, Y)

경사 하강법과 같은 알고리즘을 사용해
X로부터 Y 를 잘 찾기 위한 모델을 학습!

model.predict(X')



감사합니다.