bagging classifier

# 라이브러리 링크

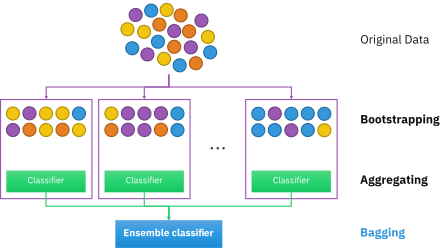
**install :**

<https://scikit-learn.org/stable/install.html>

**github :**

https://github.com/scikit-learn/scikit-learn/blob/82df48934eba1df9a1ed3be98aaace8eada59e6e/sklearn/ensemble/\_bagging.py

# 기초 설명



Bagging (from bootstrap aggregating) 은 statistical classification 과 regression 에 사용되는 머신러닝 알고리즘의 안정성과 정확도를 향상시키기 위해 설계된 머신러닝 ensemble meta-algorrithm이다.

# 버전 정보

* NumPy >= 1.14.6 (pip install numpy)
* Scipy >= 1.1.0 (pip install scipy)
* Joblib >= 0.11 (pip install joblib
* Threadpoolctl >= 2.0.0 (pip install threadpoolctl)
* pandas >= 1.2.4 (pip install pandas)
* matplotlib == 3.22 (pip install matplotlib)
* sklearn == 1.0.2 (pip install sklearn)

# 데이터셋 설명 및 출처

* diabetes.scv: This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage.
* Sources : https://www.kaggle.com/uciml/pima-indians-diabetes-database

# 코드 설명

* Sklearn 의 BaggingClassifier 모델을 사용하여 decision tree 를 bagging 방식으로 모델링 한다.
* 데이터셋을 train 데이터셋과 val 데이터셋으로 나눈다.
* Train 데이터셋을 BaggingClassifier모델과 Decision 트리 모델에 학습한다.
* 두 모델의 학습결과 검증한뒤 이를 점수화 하여 출력한다.
* 예측검증을 위한 테스트 데이터를 생성한다.
* 두 모델의 테스트 데이터에 대한 에측 결과를 출력한다.

# 검증 방법

* 코드 내부에서 데이터셋을 학습용 데이터셋과 검증용 데이터셋으로 나누어 이를 검증함.

(test\_size = 0.2, random\_state = 0)