hierarchical Agglomerative clustering

# 라이브러리 링크

**install :**

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

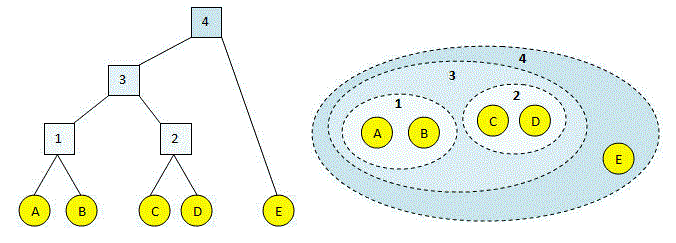
https://scipy.org/download/

**github :**

<https://github.com/scipy/scipy/blob/v1.7.1/scipy/cluster/hierarchy.py>

[https://github.com/scikit-learn/scikit-learn/blob/7e1e6d09b/sklearn/cluster/\_agglomerative.py#L740](https://github.com/scikit-learn/scikit-learn/blob/82df48934eba1df9a1ed3be98aaace8eada59e6e/sklearn/tree/_classes.py)

# 기초 설명

****

Hierarchical clustering 이란 cluster analysis을 위한 방법으로 클러스터의 계층을 만들고자 사용한다. 그리고 hierarchical agglomerative clustering 이란 계층적 클러스터의 “bottom-up”방식의 생성 방법으로 각 단계별로 가장 근접한 두 클러스터간에 페어를 구성하면서 결과적으로 하나의 케이스가 될때까지 클러스터를 묶는 클러스터링 방식이다.

# 버전 정보

* NumPy >= 1.14.6 (pip install numpy)
* Scipy >= 1.1.0 (pip install scipy)
* scikit-learn >= 1.0.2 (pip install sklearn)
* joblib>= 0.11 (pip install joblib)
* threadpoolctl >= 2.0.0 (pip install threadpoolctl)

# 데이터셋 설명 및 출처

* Customers Cluster.scv: Data sets for customers' IDs, gender, age, income, and Spending.
* Sources : <https://www.kaggle.com/tohuangjia/customer-cluster/metadata>

# 코드 설명

* Hierarchical clustering 의 진행 단계를 그래프를 통해 나타내기 위해 임시적으로 x,y 데이터셋을 임으로 생성하고 이를 scipy 의 linkage 와 dendrogram을 사용하여 linkage의 학습 결과를 dendrogram을 통해서 그래프로 시각화 하여 나타낸다.
* 그후 실제 고객들의 Id, 수입, 연령, 성별, 소비 등의 정보를 포함한 데이터셋을 읽어들여 이중 고객의 연령, 수입에 대한 데이터셋을 형성. 그후 이 데이터셋을 동일하게 scipy 의 linkage 와 dendrogram을 사용하여 linkage의 학습 결과를 dendrogram을 통해서 그래프로 시각화 하여 나타낸다. 또한 sklearn의 Agglomeric Clustering 모델에 데이터셋을 학습 시킨후 이에대한 결과를 색이 다른 scatter plot을 통해서 나타낸다.

# 검증 방법

* X