Pam clustering

# 라이브러리 링크

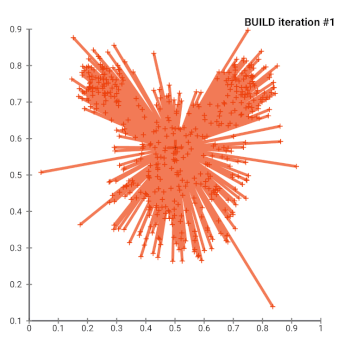
**install :**

https://scikit-learn-extra.readthedocs.io/en/stable/install.html

**source :**

https://scikit-learn-extra.readthedocs.io/en/stable/\_modules/sklearn\_extra/cluster/\_k\_medoids.html#KMedoids

# 기초 설명

**\**

**(gif image :** [**https://upload.wikimedia.org/wikipedia/commons/thumb/e/e1/K-Medoids\_Clustering.gif/350px-K-Medoids\_Clustering.gif**](https://upload.wikimedia.org/wikipedia/commons/thumb/e/e1/K-Medoids_Clustering.gif/350px-K-Medoids_Clustering.gif) **)**

PAM (Partitioning Around Medoids) 는 NP-hard 문제인 K-medoids의 heuristic solutions중 하나로 그리디 서치 방법을 사용하여 문제에 대한 정확한 해는 찾을수 없을수도 있으나 전체탐색 알고리즘 보다 빠른 속도를 가진다.

# 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)
* sklearn == 1.0.2 (pip install sklearn)

# 데이터셋 설명 및 출처

* segmentation data.scv: Customer Segmentation is the subdivision of a market into discrete customer groups that share similar characteristics. Customer Segmentation can be a powerful means to identify unsatisfied customer needs. Using the above data companies can then outperform the competition by developing uniquely appealing products and services.
* You are owing a supermarket mall and through membership cards, you have some basic data about your customers like Customer ID, age, gender, annual income and spending score. You want to understand the customers like who are the target customers so that the sense can be given to marketing team and plan the strategy accordingly.

<https://www.kaggle.com/dev0914sharma/customer-clustering?select=segmentation+data.csv>

# 코드 설명

* 데이터셋 파일에서 데이터를 읽어들인후 데이터 내부의 나이와 소득에 대한 데이터를 별도로 빼내 이 데이터의 정보를 PAM 클러스터링 모델에 입력후 클러스터링 된 결과물을 군집별로 색상을 다르게 하여 그래프로 나타낸다.

# 검증 방법

* x