

## Q4. Agglomerative Clustering

```
import numpy as np
import os

import matplotlib as mpl
import matplotlib.pyplot as plt
from matplotlib.cm import get_cmap

import sklearn
from sklearn.datasets import make_blobs
from sklearn.cluster import AgglomerativeClustering

#####
### Agglomerative clustering on random dataset ###
#####

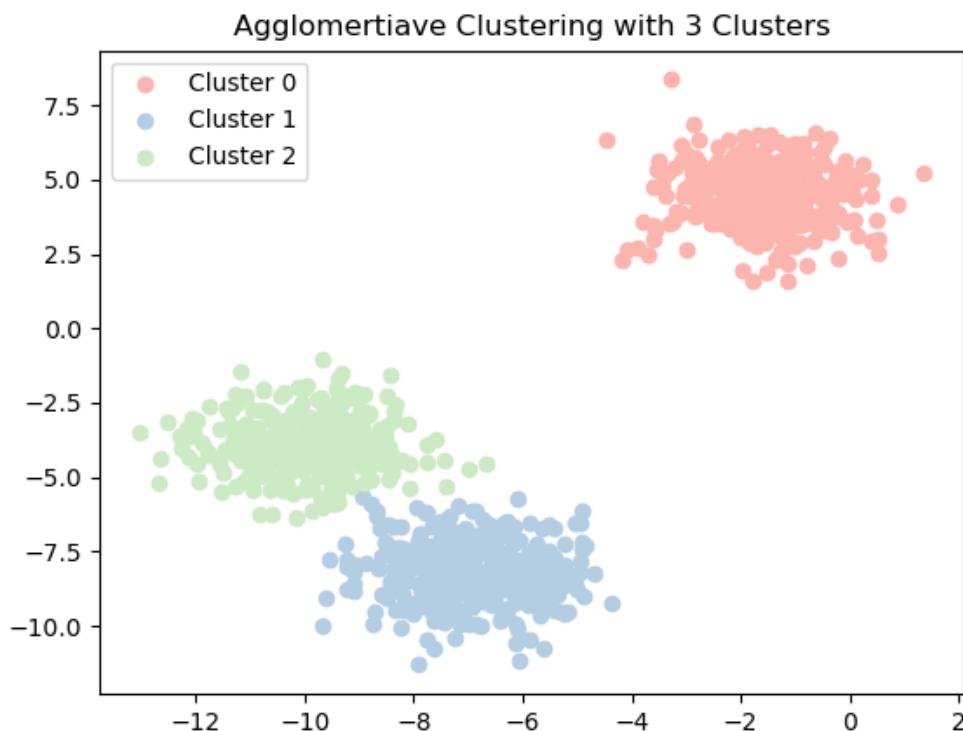
### Prepare random dataset with blob distribution #####
X, y = make_blobs(n_samples=1000, random_state=1)

### Agglomerative Clustering Preparation #####
agg = AgglomerativeClustering(n_clusters=3) # Agglomerative clusterer with 3 clusters

cluster = agg.fit_predict(X) # Fit agglomerative clusterer with current random dataset

### Plot clustering results of agglomerative clustering on given random dataset #####
cmap = get_cmap('Pastel1') # Prepare color map / Each cluster uses a distinctive color
legend = []
for label in range(agg.n_clusters_):
    # Plot only the points that correspond to certain cluster label using X[cluster==label]
    # Assign the color to the points in the dataset according to their labels
    plt.scatter(X[cluster==label][:, 0], X[cluster==label][:, 1], c=cmap.colors[label], label='Cluster ' + str(label))
    legend.append('Cluster ' + str(label))

plt.legend(legend, loc='best')
plt.title('Agglomerative Clustering with 3 Clusters')
plt.show()
```



- Agglomerative Clustering 은 거리 기준으로 데이터를 Clustering 하는 알고리즘임
- Agglomerative Clustering 은 근거리 Cluster 를 Merge 하는 기능을 가졌기 때문에 Cluster 중심점을 기준으로 목표 거리까지 Cluster 가 확장할 수 있음 .
- 거리를 기준으로 Cluster 를 구성하기 때문에 Cluster 중심점을 기준으로 목표 거리 내에 원형구조로 Cluster 가 생기는 것을 볼 수 있음 .