ICT 빅데이터의 이해 - 과제

다음 두 개의 알고리즘을 Python으로 실행해보시오.

단, 실행에 필요한 데이터는 본인이 직접 파일로 작성하고, 작성된 파일을 읽어들여 분석하시오.

실행 코드와 실행 결과를 제시하시오.

1) K-means 알고리즘을 실행해보시오.

**<실행 코드>**

%matplotlib inline

import matplotlib.pyplot as plt

import seaborn as sns; sns.set()

import numpy as np

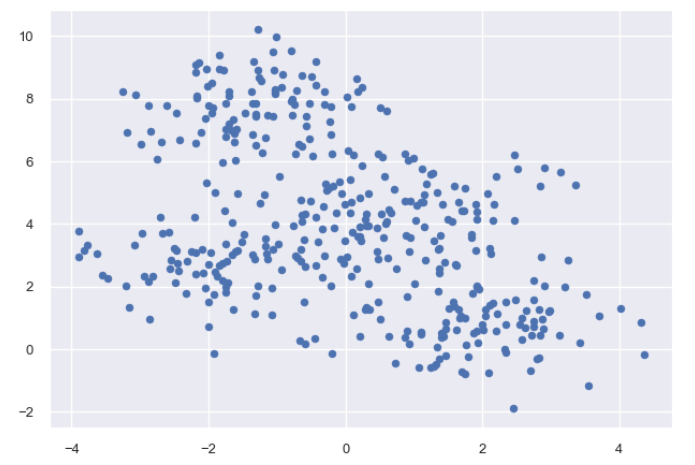
from sklearn.datasets.samples\_generator import make\_blobs

from sklearn.cluster import KMeans

X, y\_true = make\_blobs(n\_samples=400, centers=4, cluster\_std=1, random\_state=0)

plt.scatter(X[:, 0], X[:, 1], s=50)

plt.show()



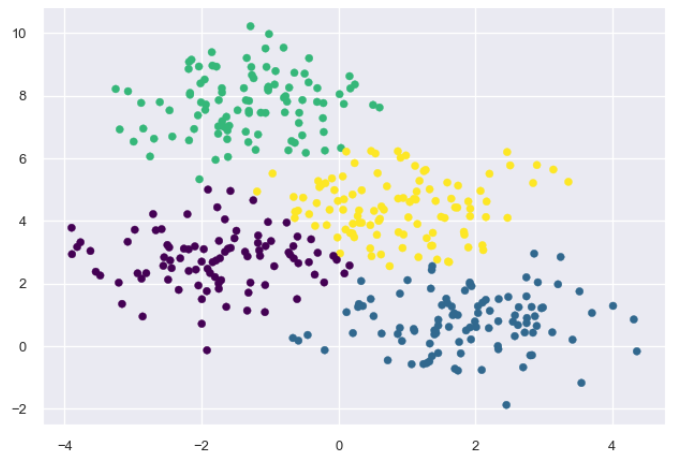
[그림 1] 실행결과

kmeans = KMeans(n\_clusters=4)

kmeans.fit(X)

y\_kmeans = kmeans.predict(X)

plt.scatter(X[:, 0], X[:, 1], c=y\_kmeans, s=50, cmap='viridis')



[그림 2] 실행결과

2) DBSCAN 알고리즘을 실행해보시오.

**<실행 코드>**

%matplotlib inline

import matplotlib

import matplotlib.pyplot as plt

from matplotlib import style

import numpy as np

from sklearn.datasets import make\_moons

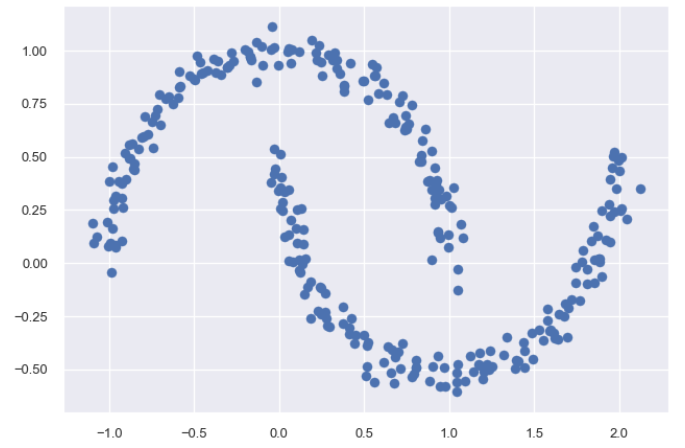
from sklearn.cluster import DBSCAN

style.use('seaborn-talk')

X, y = make\_moons(n\_samples=300, noise=0.05, random\_state=0)

plt.scatter(X[:, 0], X[:, 1])

plt.show()



[그림 ] 실행결과

def plotResult(X, y, title='Clustering Result'):

plt.scatter(X[y==0,0], X[y==0,1], c='blue', marker='o', s=30, label='cluster 1')

plt.scatter(X[y==1,0], X[y==1,1], c='red', marker='o', s=30, label='cluster 2')

plt.title(title)

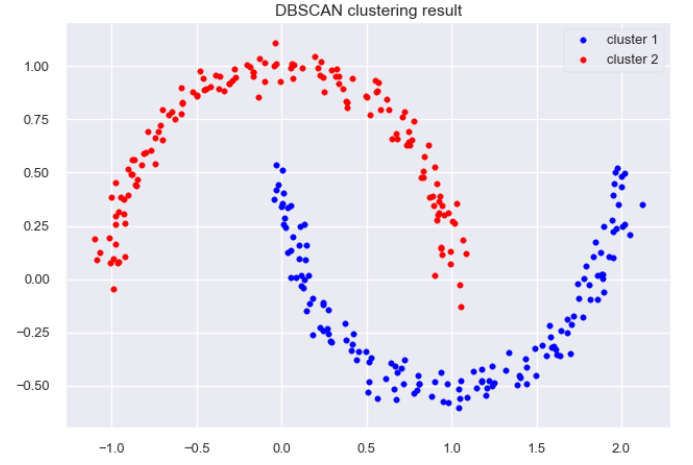
plt.legend()

plt.show()

db = DBSCAN(eps=0.3, min\_samples=4, metric='euclidean')

y\_db = db.fit\_predict(X)

plotResult(X, y\_db, title='DBSCAN clustering result')



[그림 ] 실행결과