Knowledge Discovery and Data Mining

Lab 9 DBSCAN

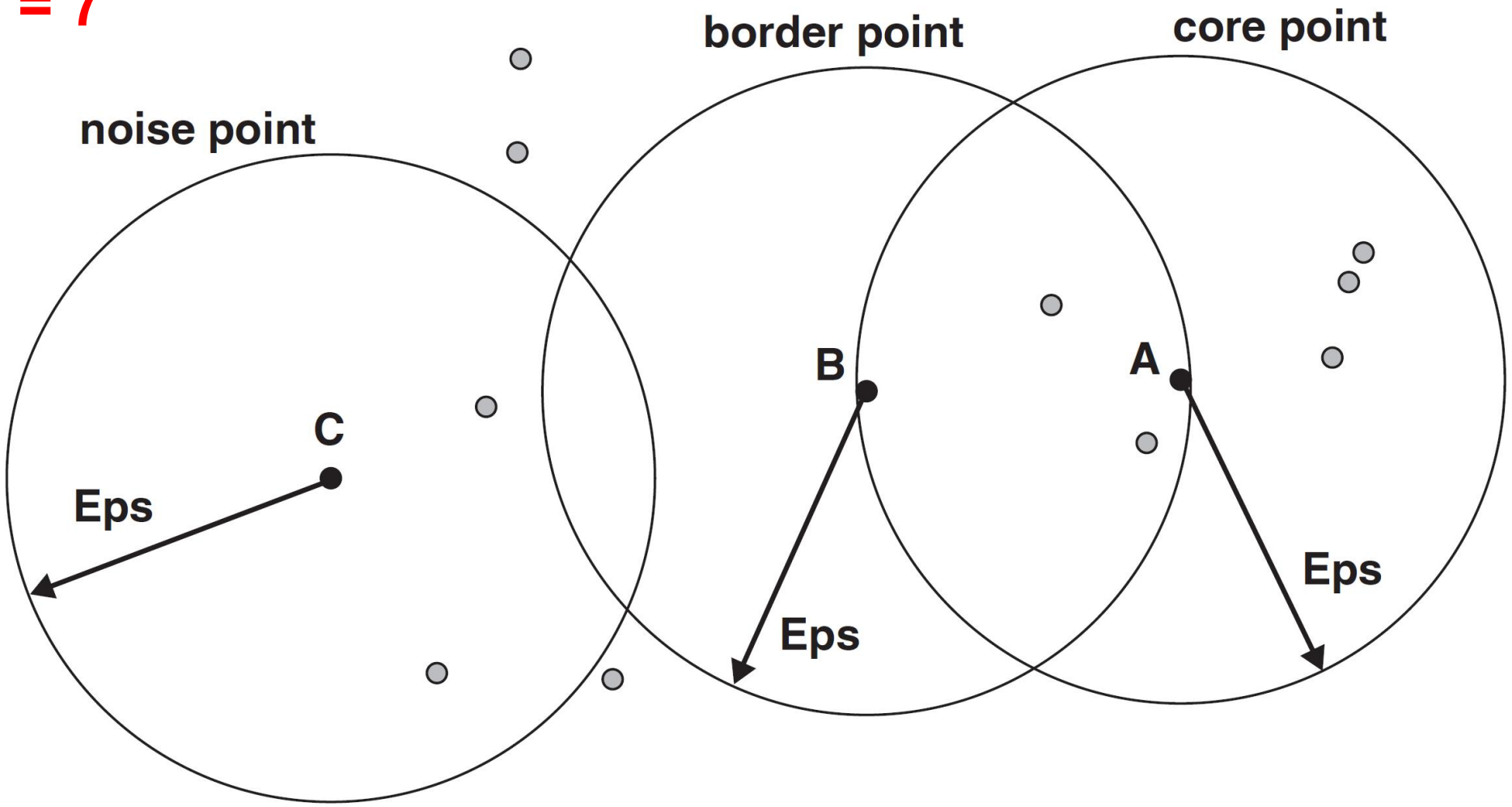
Xuan Song
Songx@sustech.edu.cn

Topics

Implement DBSCAN with scikit-learn

DBSCAN Algorithm

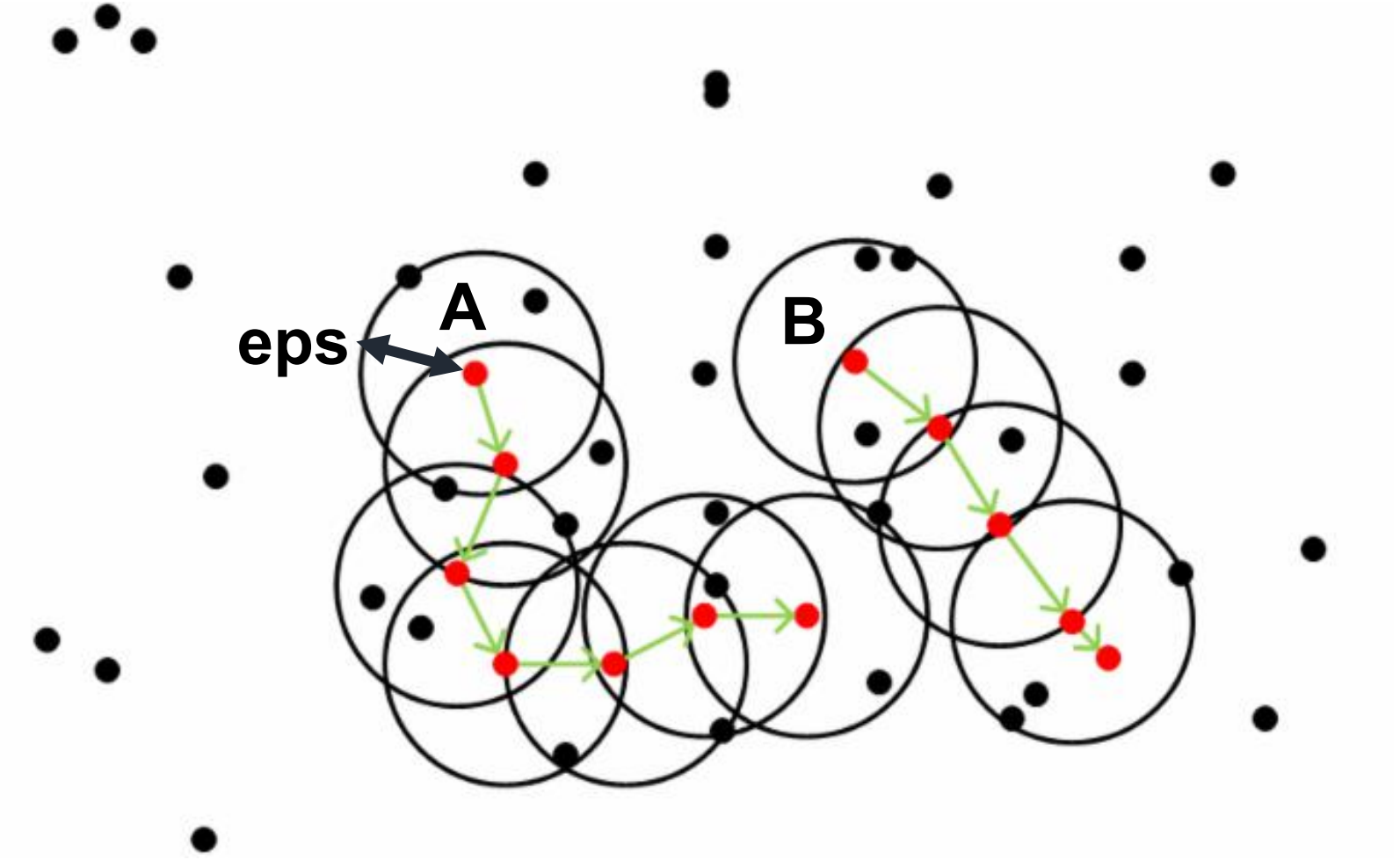
MinPts = 7



https://en.wikipedia.org/wiki/DBSCAN

DBSCAN Algorithm

MinPts = 5



https://en.wikipedia.org/wiki/DBSCAN

Implementing DBSCAN with Scikit-Learn

- sklearn.cluster.DBSCAN

  class sklearn.cluster.DBSCAN(**eps=0.5**, *, **min_samples=5**, metric='euclidean', metric_params=None, algorithm='auto′ , leaf_size=30, p=None, n_jobs=None)

https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html

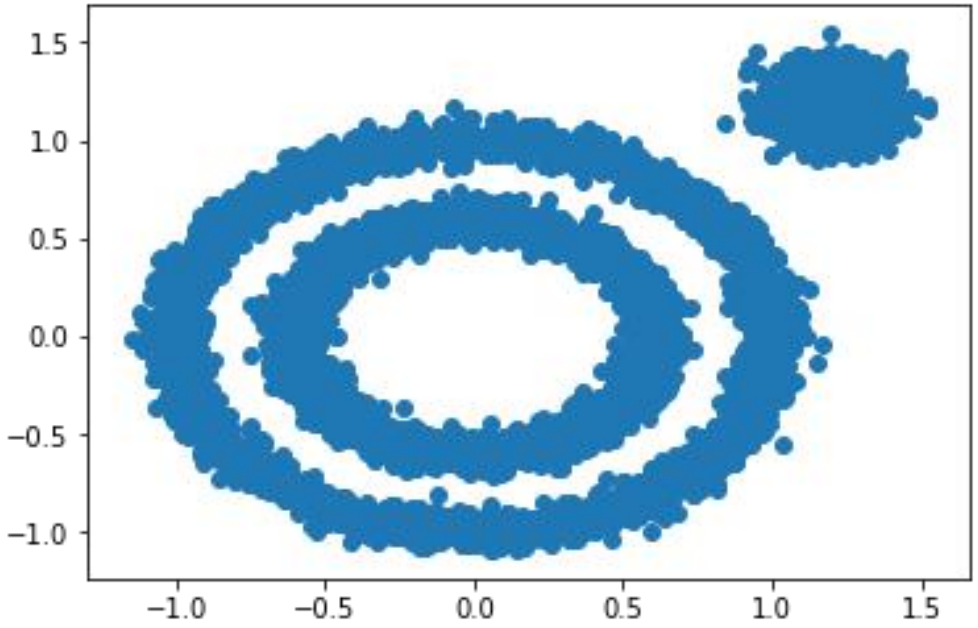
Implementing DBSCAN with Scikit-Learn

- Sample data: DBSCAN_SAMPLE.csv

DBSCAN_SAMPLE.csv

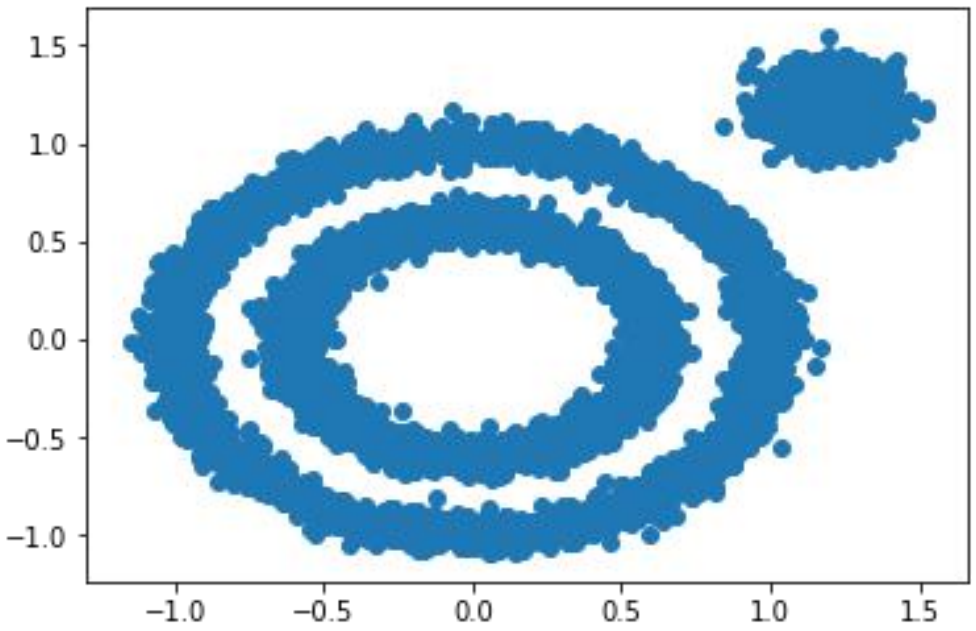
Attribute

X0	X1
0.187219	-0.58478
-0.48401	0.226619
-0.60769	0.27942
0.829826	0.481182
0.479004	0.384539
-0.58032	0.891131
0.02593	-0.98776
0.952103	0.300135
0.877564	0.490907

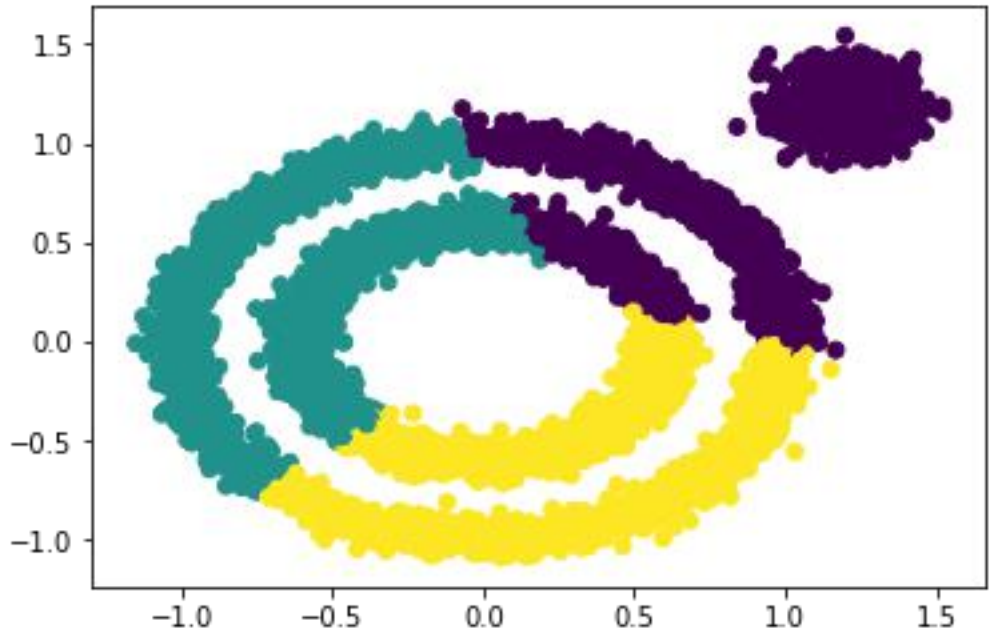


Implementing DBSCAN with Scikit-Learn

- Sample data: DBSCAN_SAMPLE.csv



K-MEANS



K=3

Implementing DBSCAN with Scikit-Learn

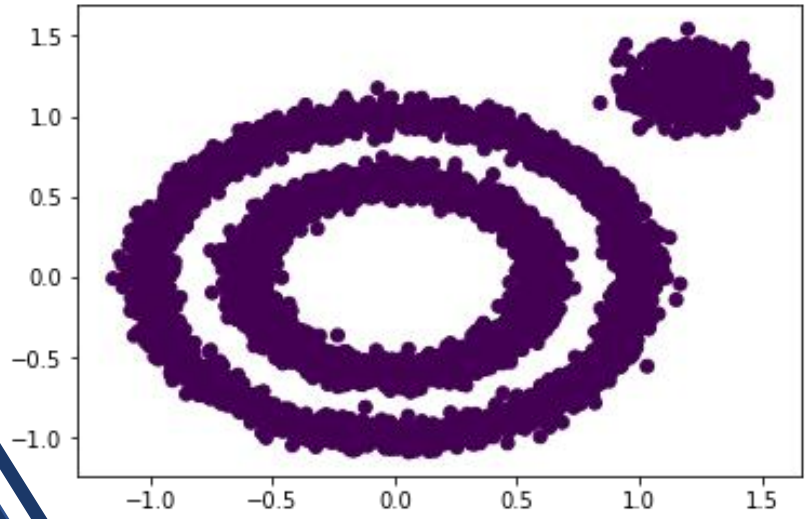
1. Load data from csv files.
2. Data cleaning.
3. Get features.
4. Build a DBSCAN model with scikit-learn.

```
from sklearn.cluster import DBSCAN
dbscan = DBSCAN()
#dbscan = DBSCAN(eps = 0.1)
#dbscan = DBSCAN(eps = 0.1, min_samples = 10)
```

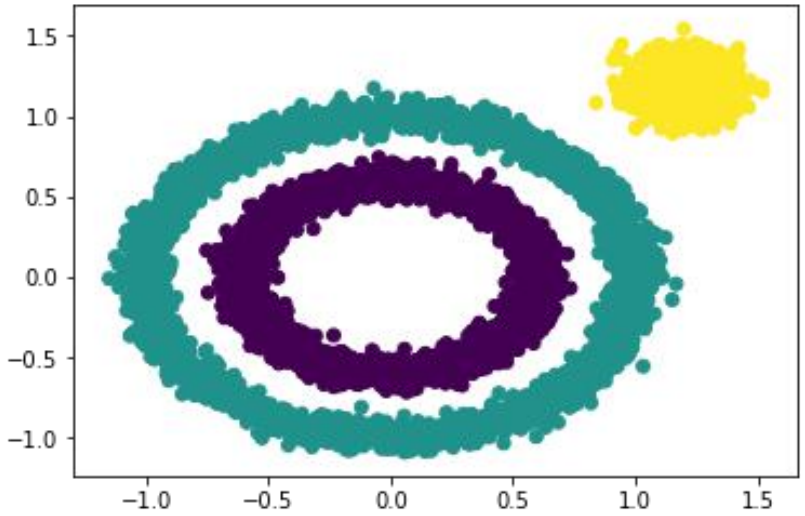
Implementing DBSCAN with Scikit-Learn

5. Visualize the cluster result.

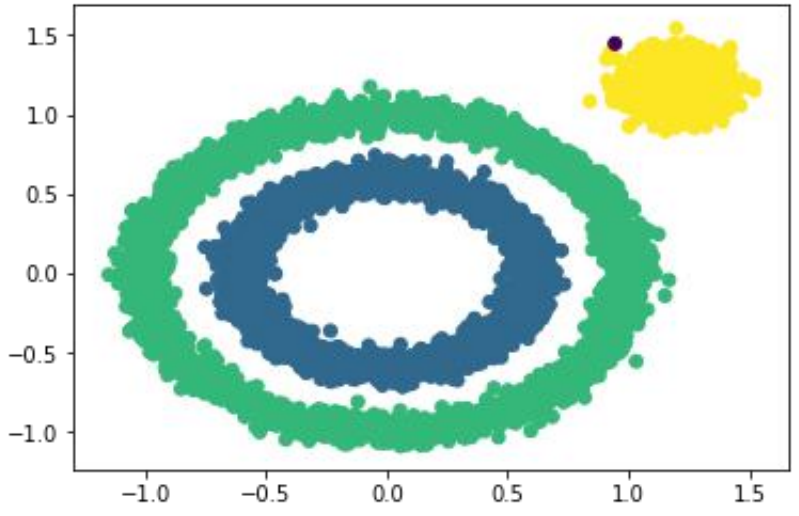
```
y_pred  = dbscan.fit_predict(X)
plt.scatter(X[:, 0], X[:, 1], c=y_pred)
plt.show()
```



default parameters



eps=0.1



eps=0.1, min_samples=10

Task1:

- Implementing DBSCAN based on the given dataset.

DBSCAN_TASK 1.csv

End of Lab9