

07

## 혼합분포 군집분석

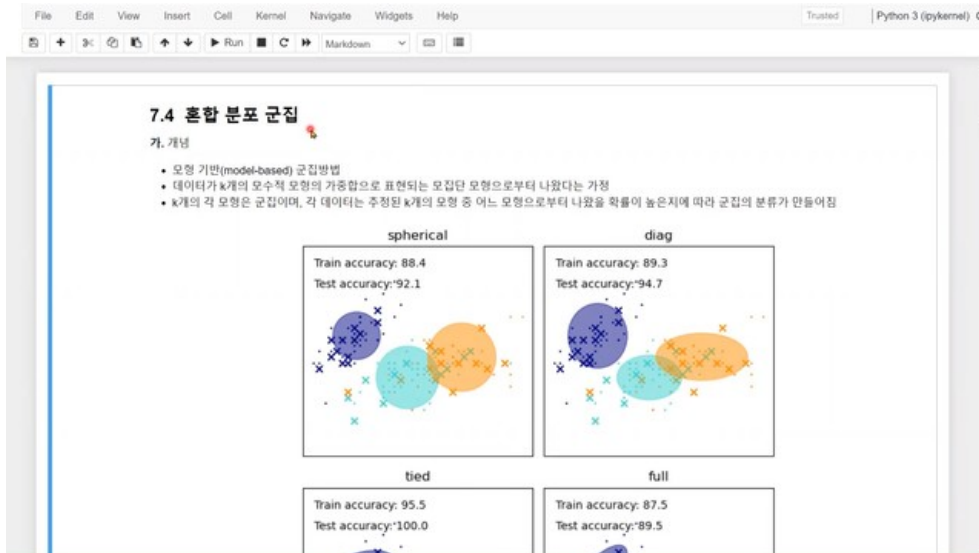
[혼합분포 군집분석 개념]

# ADP는 다양한 알고리즘을 사용할 수 있는지를 물어봅니다.

군집분석 안에도 다양한 알고리즘을 쓸 수 있어야 합니다.

- 혼합분포 군집은 분포기반, 확률 기반 군집분석입니다.

00:00



[혼합분포 군집분석 실습]

# k-means와 같이 군집 수를 지정해 주어야 합니다.

04:18

File Edit View Insert Cell Kernel Navigate Widgets Help Trusted Python 3 (pykernel)

→ 몇 개의 확률분포를 혼합할 것인지는 사용자가 지정함. (그 개수에 맞춰서 모델이 만들어짐)

**[예제]**  
iris 데이터의 Species를 제외하고 혼합분포군집분석을 실시하기

```
In [1]: 1 # 필요한 모듈
2 import pandas as pd
3 import numpy as np
4 import sklearn
5 from sklearn.preprocessing import StandardScaler
6 from sklearn.mixture import GaussianMixture
```

```
In [2]: 1 # 데이터 로드
2 iris = pd.read_csv('../data/iris.csv')
3 df = iris.drop('target', axis=1)
4 df.head()
```

```
Out[2]:
```

	sepal length	sepal width	petal length	petal width
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2

```
In [3]: 1 scaler = StandardScaler()

In [4]: 1 df_scaled = scaler.fit_transform(df)

In [5]: 1 # 가우시안 혼합모델 구축
```

File Edit View Insert Cell Kernel Navigate Widgets Help Trusted Python 3 (pykernel)

→ 몇 개의 확률분포를 혼합할 것인지는 사용자가 지정함. (그 개수에 맞춰서 모델이 만들어짐)

**[예제]**  
iris 데이터의 Species를 제외하고 혼합분포군집분석을 실시하기

```
In [1]: 1 # 필요한 모듈
2 import pandas as pd
3 import numpy as np
4 import sklearn
5 from sklearn.preprocessing import StandardScaler
6 from sklearn.mixture import GaussianMixture
```

```
In [2]: 1 # 데이터 로드
2 iris = pd.read_csv('../data/iris.csv')
3 df = iris.drop('target', axis=1)
4 df.head()
```

```
Out[2]:
```

	sepal length	sepal width	petal length	petal width
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2

```
In [3]: 1 scaler = StandardScaler()

In [4]: 1 df_scaled = scaler.fit_transform(df)

In [5]: 1 # 가우시안 혼합모델 구축
```