In [1]:

```
import numpy as np
import pandas as pd
```

(i)

a1, a2: Aに属する点
b1, b2, b3: Bに属する点

a1, a2: Elements of A
b1, b2, b3: Elements of B

In [2]:

```
a1=np.asarray((0,0))
a2=np.asarray((0,1))

b1=np.asarray((1,1))
b2=np.asarray((1,2))
b3=np.asarray((2,2))
```

In [5]:

```
# 重心を求める
# computing center of mass

com_a=(a1+a2)/2
com_b=(b1+b2+b3)/3

com_tot=(a1+a2+b1+b2+b3)/5
```

In [6]:

```
print(com_a)
print(com_b)
print(com_tot)
```

```
[0.  0.5]
[1.33333333 1.66666667]
[0.8 1.2]
```

In [7]:

```
# 重心からの2乗距離の和
# sum of squared norm from the center of mass

# set A
dist_a=np.dot(a1-com_a, a1-com_a)+np.dot(a2-com_a, a2-com_a)

# set B
dist_b=np.dot(b1-com_b, b1-com_b)+np.dot(b2-com_b, b2-com_b)+np.dot(b3-com_b, b3-com_b)
```

In [8]:

```
# A and B
dist_tot=np.dot(a1-com_tot, a1-com_tot)+np.dot(a2-com_tot, a2-com_tot)+¥
np.dot(b1-com_tot, b1-com_tot)+np.dot(b2-com_tot, b2-com_tot)+np.dot(b3-com_tot, b3-com_tot)
```

In [9]:

```
print(dist_tot, dist_a, dist_b)
```

5.6000000000000005 0.5 1.3333333333333333

In [10]:

```
# Ward法による距離
# Dinstance by ward method
dist_tot-(dist_a+dist_b)
```

Out[10]:

3.7666666666666675

Ans.

3.77

(ii)

In [11]:

```
from scipy.cluster.hierarchy import linkage,  fcluster
```

In [12]:

```
df=pd.read_csv('ai_mid4.csv', header=0)
```

In [13]:

```
df.head()
```

Out[13]:

	alcohol	malic_acid	ash	class
0	1.518613	-0.562250	0.232053	A
1	0.246290	-0.499413	-0.827996	A
2	0.196879	0.021231	1.109334	A
3	1.691550	-0.346811	0.487926	A
4	0.295700	0.227694	1.840403	A

階層型クラスタリングの実行
Hierarchical clustering

In [14]:

```
Z = linkage(df[['alcohol', 'malic_acid', 'ash']], method='ward', metric='euclidean')
```

クラスタ数3の場合
case with # of clusters = 3

In [15]:

```
cluster = pd.Series( fcluster(Z, 3, criterion='maxclust') )
```

class 列とクロス集計表を作成
making cross table with the class column

In [16]:

```
pd.crosstab(df['class'], cluster)
```

Out[16]:

col_0	1	2	3
class			
A	6	46	7
B	52	4	15
C	3	4	41

Ans

6, 46, 7