

## ANN Homework 2

due 2021/11/29

### 3. 分別以下列式子產生三個群集，每個群集有20個點

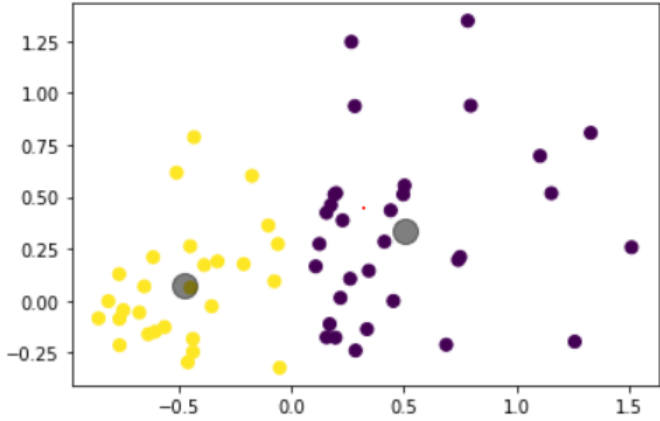
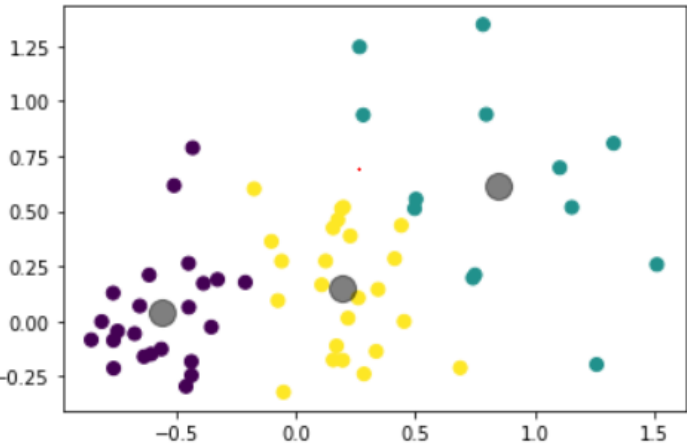
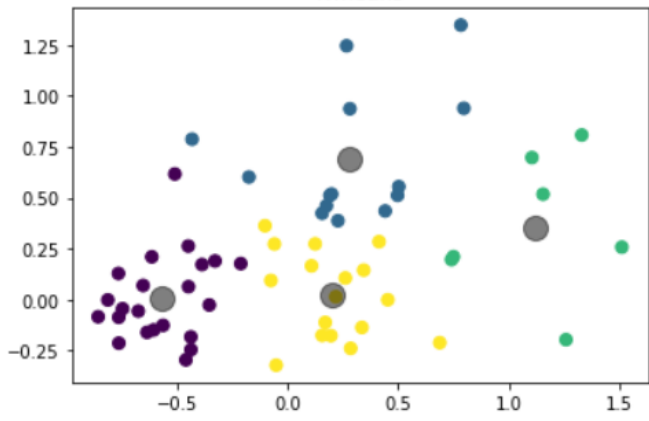
$x_1 = \sin(\theta_1) \cos(\phi_1) + \varepsilon_{11}$	$x_2 = \sin(\theta_2) \cos(\phi_2) + \varepsilon_{21}$	$x_3 = \sin(\theta_3) \cos(\phi_3) + \varepsilon_{31}$
$y_1 = \sin(\theta_1) \sin(\phi_1) + \varepsilon_{12}$	$y_2 = \sin(\theta_2) \sin(\phi_2) + \varepsilon_{22}$	$y_3 = \sin(\theta_3) \sin(\phi_3) + \varepsilon_{32}$
$\theta_1 \in \left[-\frac{\pi}{4}, \frac{\pi}{4}\right]$	$\theta_2 \in \left[\frac{\pi}{4}, \frac{3\pi}{4}\right]$	$\theta_3 \in \left[-\frac{\pi}{2}, \frac{3\pi}{2}\right]$
$\phi_1 \in [0, 2\pi]$	$\phi_2 \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$	$\phi_3 \in \left[-\frac{\pi}{4}, \frac{3\pi}{4}\right]$

$\varepsilon_{ij}$  為平均值 = 0，標準偏差 = 0.6 的隨機亂數

試利用 K-Means、Fuzzy C-Means 及減法聚類法來進行資料分類。

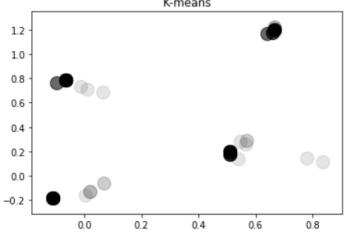
(提示：K-Means 與 Fuzzy C-Means 試使用 2~4 類進行分類)

Q3

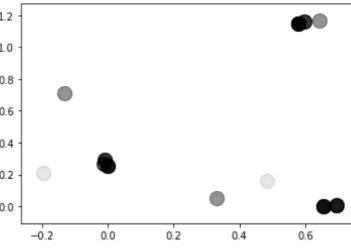
	K-MEANS	center
K=2		<pre>[ 0.504,  0.334] [-0.476,  0.072]</pre>
K=3		<pre>[-0.560,  0.043] [ 0.844,  0.617] [ 0.195 , 0.1536]</pre>
K=4		<pre>[-0.566 ,  0.009] [ 0.279 ,  0.690] [ 1.121 ,  0.356] [ 0.204 ,  0.020]</pre>

- 重複做 500 次
- 計算-各類每點到該類中心的距離平均 (書上的  $J_i$  的平均)
- 繪製每次各群中心位置，觀察 500 次的變動

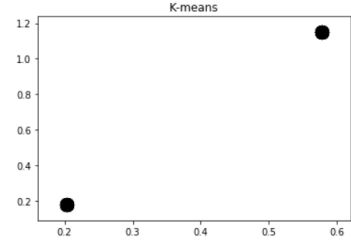
#### Kmeans-4 群

各類每點到該類中心的距離平方平均	C1_d	C2_d	C3_d	C4_d	
只做 1 次	0.96	0.77	1.43	0.89	
Iter =500(500 次的平均)	1.085	1.074	1.028	0.951	
Iter =500(500 次變異數)	0.058	0.051	0.0636	0.050	

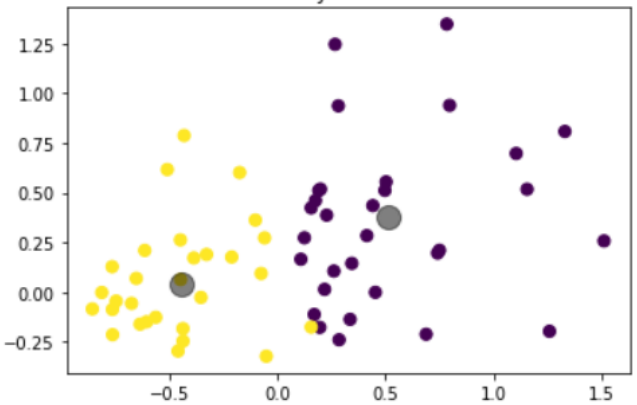
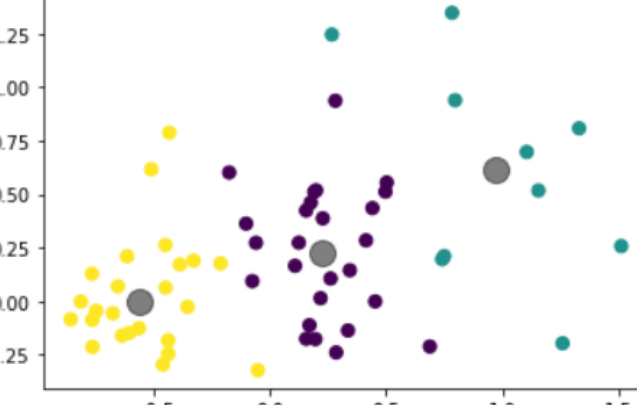
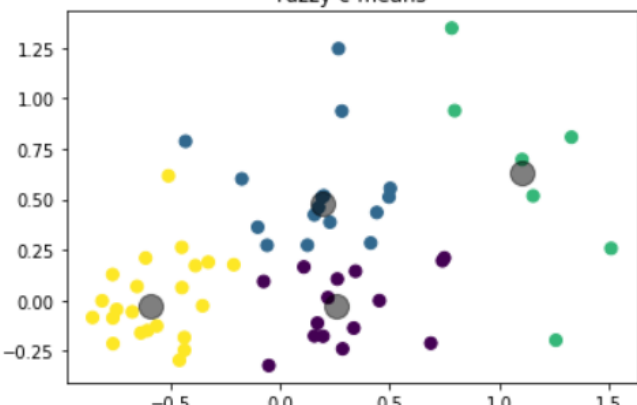
#### Kmeans-3 群

各類每點到該類中心的距離平方平均	C1_d	C2_d	C3_d	
只做 1 次	2.079	1.103	1.017	
Iter =500(500 次的平均)	1.356	1.452	1.326	
Iter =500(500 次變異數)	0.110	0.1027	0.1095	

#### Kmeans-2 群

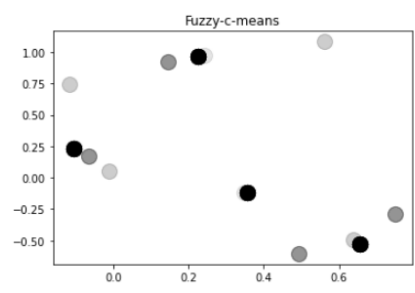
各類每點到該類中心的距離平方平均	C1_d	C2_d	
只做 1 次	2.595	1.278	
Iter =500 (500 次的平均)	1.963	1.910	
Iter =500 (500 次變異數)	0.4332	0.4332	

-從上表觀察發現，重複做 kmeans 500 次，也就是隨機選擇初始位置 500 次。群集數目越少，其分群結果，各類每點到該類中心的距離平方平均 500 次的變異數會越大。

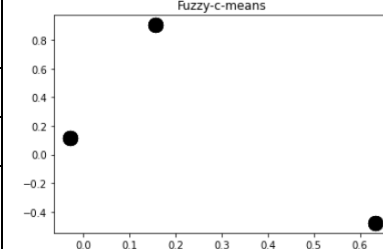
c	Fuzzy-c-means	
2		<p>[ 0.513,0.381], [-0.445,0.037]</p>
3		<p>[ 0.222,0.223], [ 0.9699,0.6189], [-0.558, -0.0045]</p>
4		<p>[0.259, -0.0237], [ 0.196,0.4768], [ 1.104,0.6281], [-0.594,-0.02451]</p>

- 重複做 500 次
  - 計算-各類每點到該類中心的距離平方平均
  - 繪製每次各群中心位置，觀察 500 次的變動

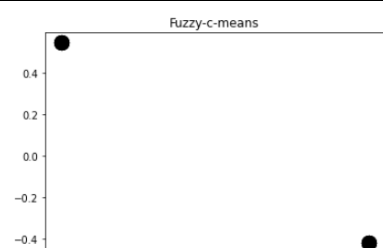
#### Fuzzy-c-means -4 群

各類每點到該類中心的距離平方平均	C1_d	C2_d	C3_d	C4_d	
只做 1 次	1.175	1.313	1.419	1.943	
Iter =500(500 次的平均)	1.182	1.171	1.193	1.180	
Iter =500(500 次變異數)	0.049	0.051	0.050	0.049	

#### Fuzzy-c-means -3 群

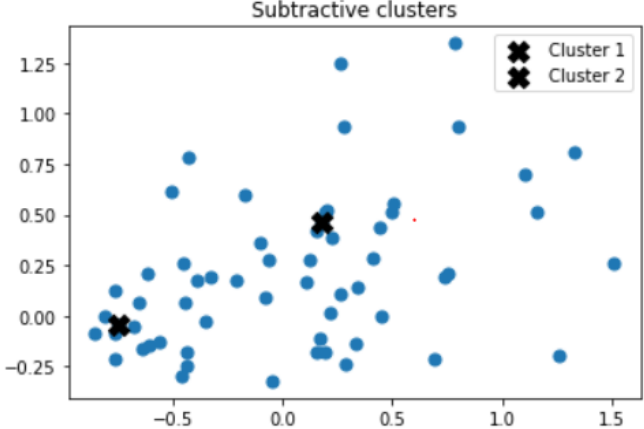
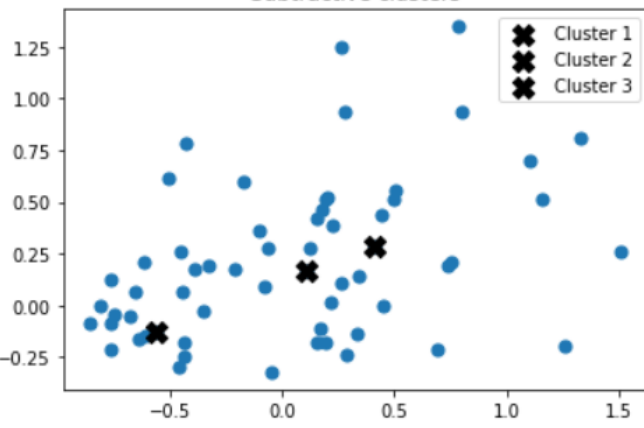
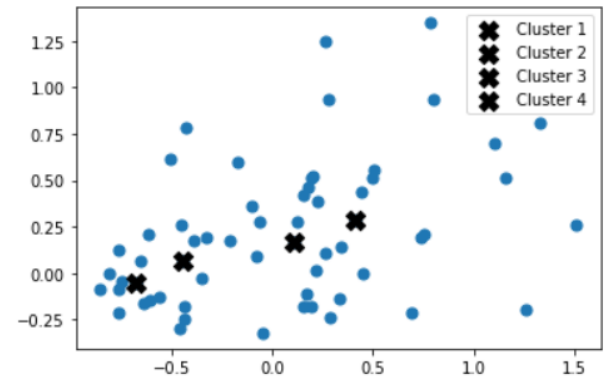
各類每點到該類中心的距離平方平均	C1_d	C2_d	C3_d	
只做 1 次	1.551	1.501	1.135	
Iter =500(500 次的平均)	1.400	1.388	1.399	
Iter =500(500 次變異數)	0.033	0.034	0.035	

#### Fuzzy-c-means -2 群

各類每點到該類中心的距離平方平均	C1_d	C2_d	
只做 1 次	2.337	1.979	
Iter =500 (500 次的平均)	2.151	2.165	
Iter =500 (500 次變異數)	0.031	0.031	

- 比較 kmeas 和 Fuzzy-c-means，Fuzzy-c-means 重複 500 次的變異數，比較不會隨著群集數目減少而變大。Fuzzy-c-means 的分群結果似乎比較穩定。其中觀察兩者重複 500 次，各類中心點變動的情形，比較 Fuzzy-c-means，黑色點的變動幅度也比較小，也就是執行多次的分群結果會比較一致。

依據中心點影響半徑決定群集數目，

r_a(影響半徑)->群集數目		
0.45- >2		<p>[-0.745, -0.0446], [0.1762, 0.4609]</p>
1.9->3		<p>[0.108, 0.1658], [-0.562, -0.126], [0.4139, 0.2847]</p>
2->4		<p>[0.10881, 0.1658], [-0.4474, 0.0638], [-0.6742, -0.056], [0.4139, 0.2847]</p>

Subtractive clusters 是透過影響半徑來決定最後可能會被分成幾個群集，也改進了隨機選擇起始點的問題。在群集數目為 2 時，Subtractive clusters 的中心點位置和前兩者方法大致是一樣的，但是群集數目增加為 3、4 時就很不一樣了。