

群聚分析操作介紹

-以SOM和K-means為例

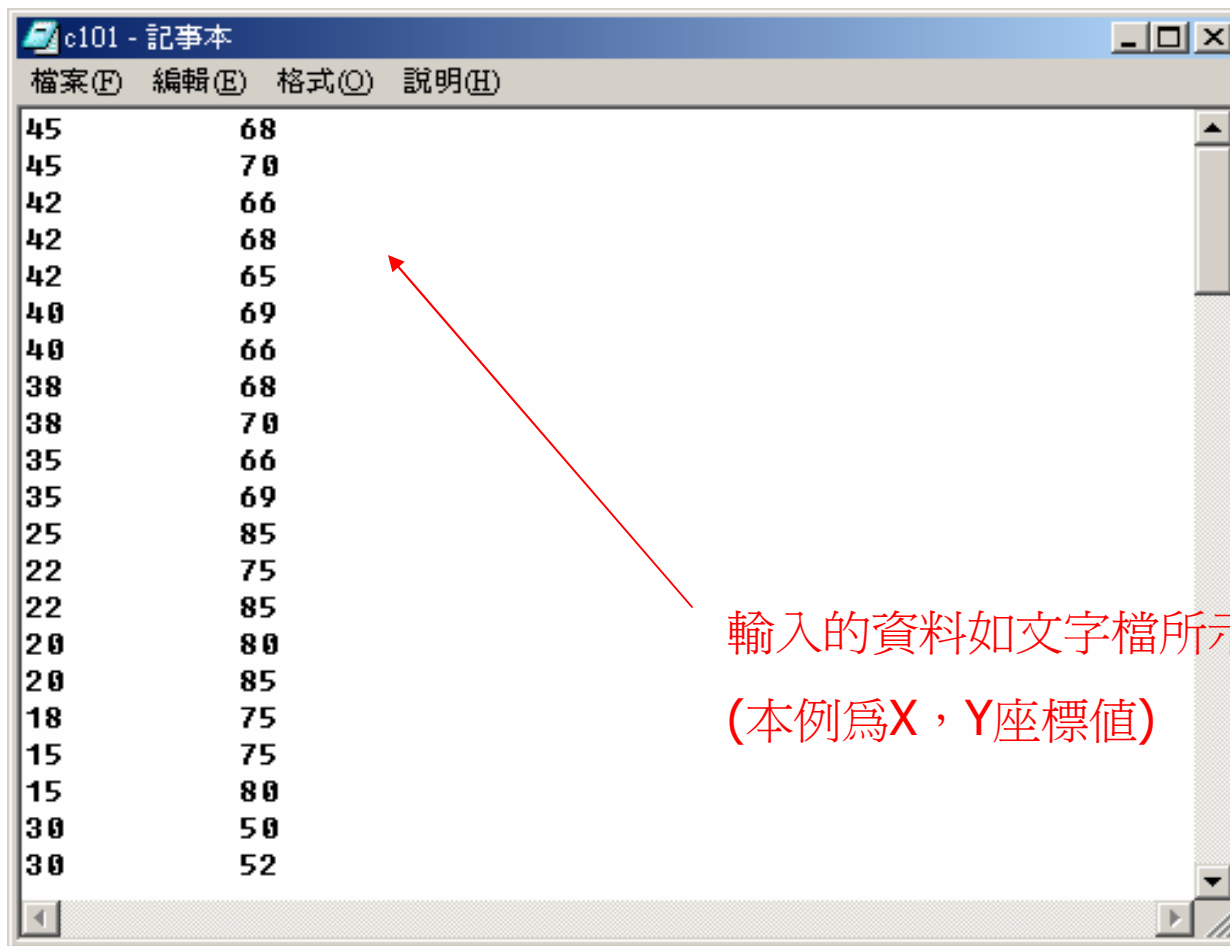
使用工具:**MATLAB 6.1**
SPSS11.01



SOM操作步驟

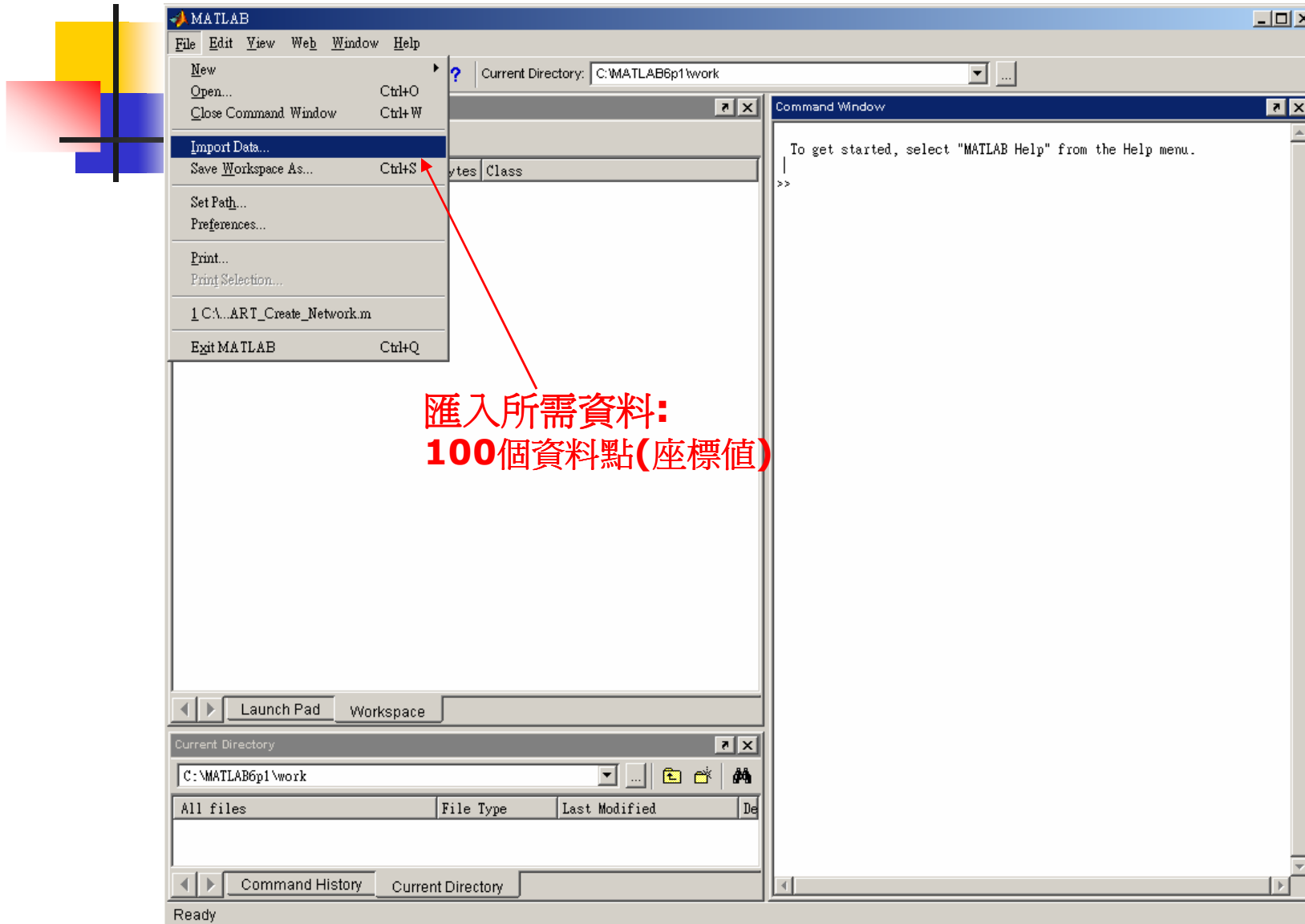
- 資料前置處理
- Matlab操作

資料前置處理

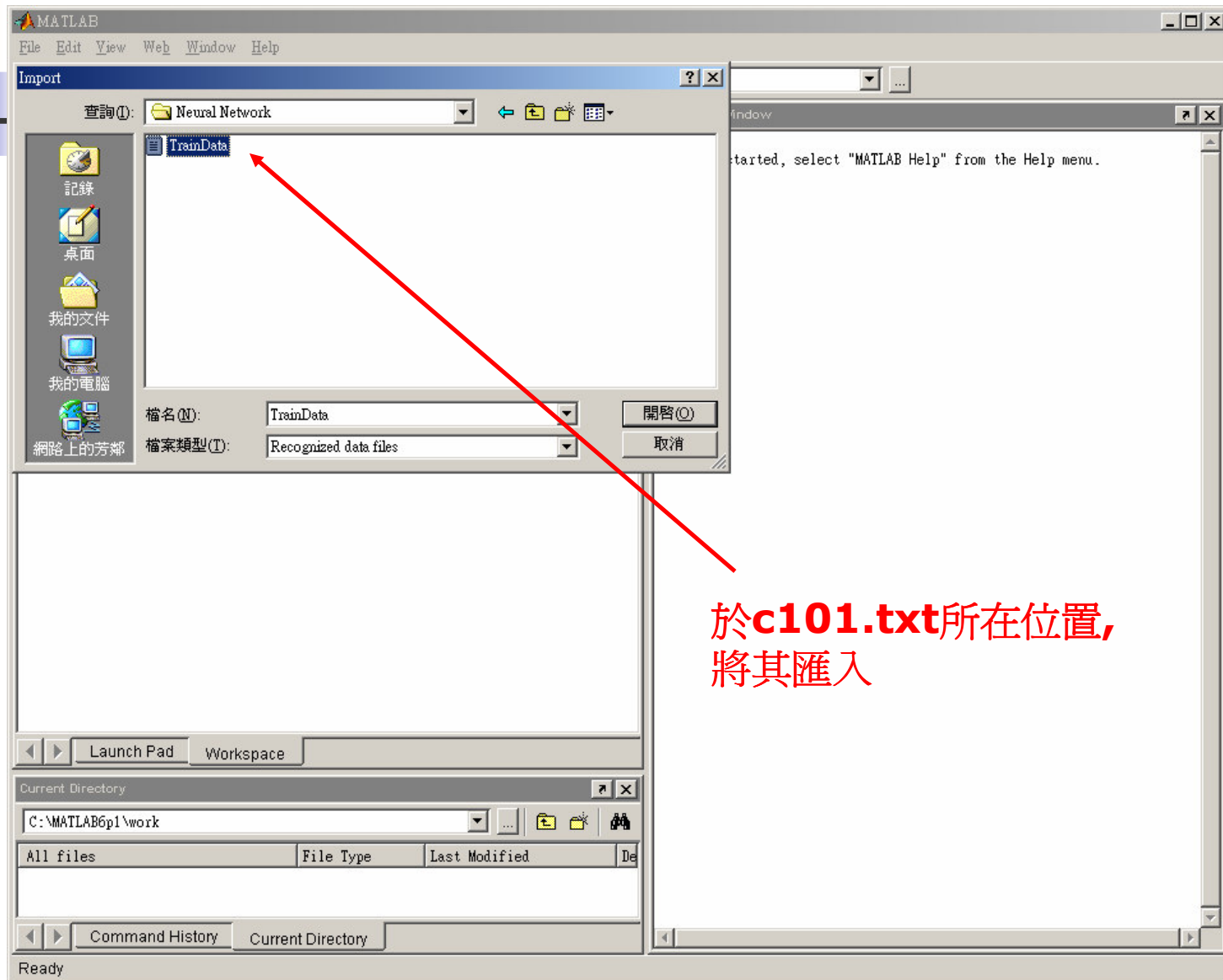


輸入的資料如文字檔所示
(本例為X，Y座標值)

Matlab操作—匯入資料



Matlab操作—匯入資料(續)



於c101.txt所在位置,
將其匯入

Matlab操作—匯入資料(續)

The screenshot shows the MATLAB Import Wizard dialog box for the file 'c101.txt'. The 'Column separator' is set to 'Space' and 'Text header lines' is set to '0'. A preview of the data is shown in a table format.

	1	2	3	
1	45	68		
2	45	70		
3	42	66		
4	42	68		
5	42	65		
6	40	69		

A red arrow points to the 'Next >' button, with the text '選擇“Next”' (Select 'Next') written in red.

SOM and K-means example

Matlab操作—匯入資料(續)

The screenshot shows the MATLAB Import Wizard dialog box for the file 'c101.txt'. The 'Create variables matching preview' option is selected. The variable 'c101' is checked in the list. A preview table shows the first 10 rows of data.

	1	2	3	4
1	45	68		
2	45	70		
3	42	66		
4	42	68		
5	42	65		
6	40	69		
7	40	66		
8	38	68		

Below the dialog box, a red arrow points to the 'Finish' button with the text '選擇“Finish”' (Select "Finish").

SOM and K-means example

Matlab操作—匯入資料(續)

The image shows the MATLAB software interface. The top menu bar includes File, Edit, View, Web, Window, and Help. The current directory is C:\MATLAB6p5\work. The Workspace window displays a table with the following data:

Name	Size	Bytes	Class
c101	100x2	1600	double array

The Command Window shows the following text:

```
Using Toolbox Path Cache. Type "help toolbox_path_cache"  
To get started, select "MATLAB Help" from the Help menu.  
Import Wizard created variables in the current workspace.  
>> c101'
```

A red arrow points from the text below to the command `c101'` in the Command Window.

將輸入資料由直列轉橫列
c101'(c101為檔案名稱)

Matlab操作—呼叫類神經網路建構工具

於命令列下鍵入“nntool”
以呼叫類神經網路模組

Name	Size	Bytes	Class
ans	2x100	1600	double array
c101	100x2	1600	double array

```
Columns 55 through 63
42 40 40 38 38 35 50 50 50
15 5 15 5 15 5 30 35 40

Columns 64 through 72
48 48 47 47 45 45 95 95 53
30 40 35 40 30 35 30 35 30

Columns 73 through 81
92 53 45 90 88 88 87 85 85
30 35 65 35 30 35 30 25 35

Columns 82 through 90
75 72 70 68 66 65 65 62 60
55 55 58 60 57 56 70 85 55

Columns 91 through 99
60 67 65 65 62 60 60 58 55
60 85 85 82 80 80 85 75 80

Column 100
55
85

>> nntool
```

轉置後得到ans檔

Matlab操作—呼叫類神經網路建構工具(續)

The screenshot shows the MATLAB environment with the Network/Data Manager dialog box open. The dialog box has several sections: Inputs, Networks, Outputs, Targets, Errors, Input Delay States, and Layer Delay States. At the bottom, there are buttons for 'Help', 'New Data...', 'New Network...', 'Import...', 'Export...', 'View', and 'Delete'. A red arrow points to the 'Import...' button. The background shows the MATLAB workspace with variables 'ans' and 'c101', and the Command Window with a grid of numbers.

選擇“Import”，
將資料ans匯入

Matlab操作—呼叫類神經網路建構工具(續)

Import or Load to Network/Data Manager

Source

- Import from MATLAB workspace
- Load from disk file

MAT-file Name

Browse

Select a Variable

(no selection)

ans

c101

Destination

Name

ans

Import As:

- Network
- Inputs
- Targets
- Initial Input States
- Initial Layer States
- Outputs
- Errors

Cancel Import

Workspace

Name	Size
ans	2x100
c101	100x2

Command Window

```
>> mtool  
>>
```

將ans設為Input型式

SOM and K-means example

Matlab操作——呼叫類神經網路建構工具(續)

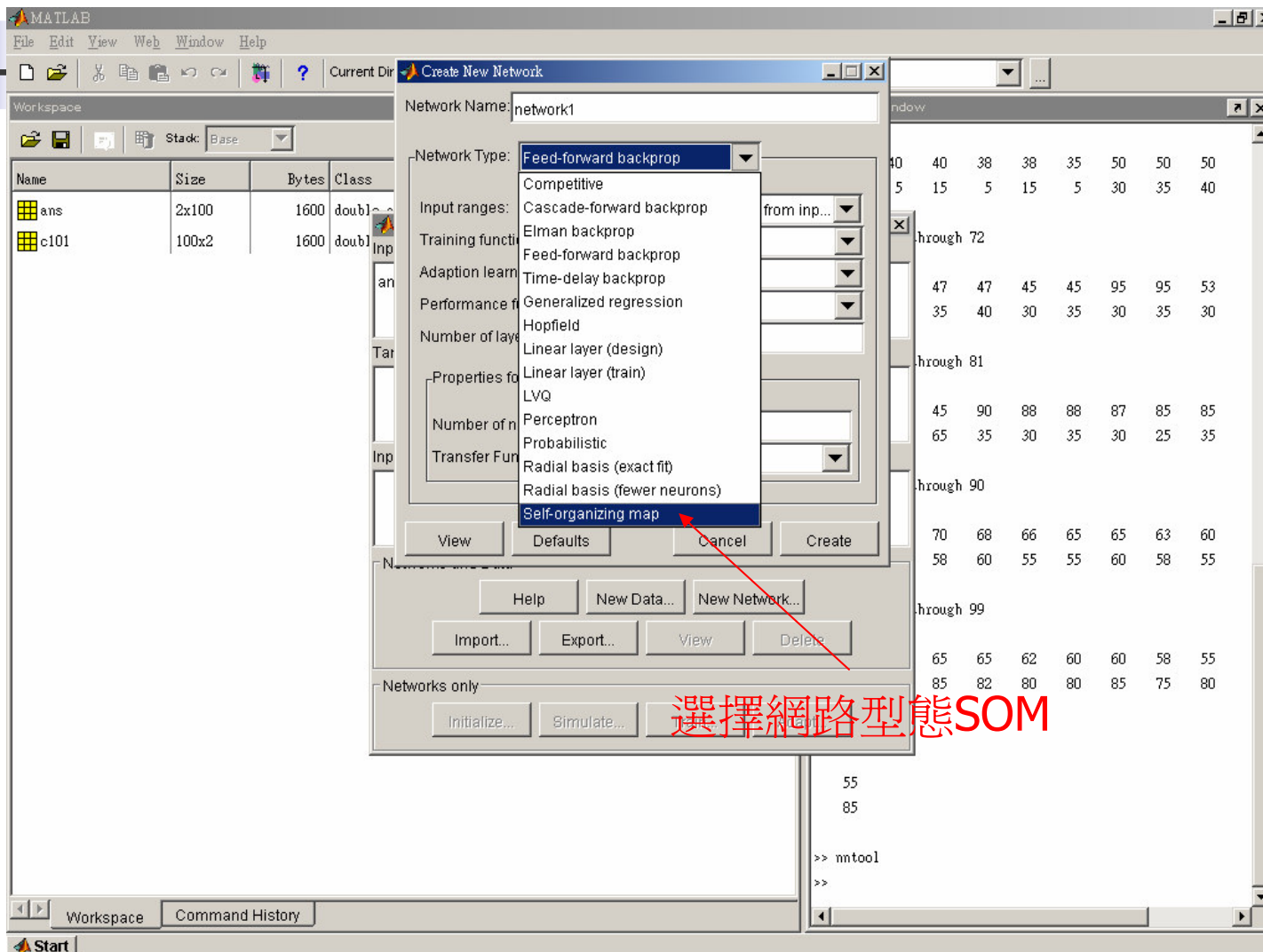
匯入的資料

待所需資料匯入後，即可選擇“**New Network**”，以建構類神經網路模組

Name	Size	Bytes	Class
ans	2x100	1600	doubl
c101	100x2	1600	doubl

42	40	40	38	38	35	50	50	50
15	5	15	5	15	5	30	35	40
through 72								
47	47	45	45	95	95	53		
35	40	30	35	30	35	30		
through 81								
45	90	88	88	87	85	85		
65	35	30	35	30	25	35		
through 90								
70	68	66	65	65	63	60		
58	60	55	55	60	58	55		
through 99								
65	65	62	60	60	58	55		
85	82	80	80	85	75	80		

Matlab操作——呼叫類神經網路建構工具(續)



選擇網路型態SOM

Matlab操作—呼叫類神經網路建構工具(續)

Network Name: network1

Network Type: Self-organizing map

Input ranges: [0 1; -1 1] **Get from i...**
Get from input:
ans

Dimensions of map: [5 8]

Topology function: HEXTOP

Distance function: LINKDIST

Ordering phase learning rate: 0.9

Ordering phase steps: 1000

Tuning phase learning rate: 0.02

Neighborhood distance: 1.0

Buttons: View Defaults Cancel Create

Networks and Data: Help New Data... New Network...
Import... Export... View Delete

Networks only: Initialize... Simulate... Train... Adapt...

Workspace:

Name	Size	Bytes	Class
ans	2x100	1600	doubl
c101	100x2	1600	doubl

Command Window:

```
>> mntool  
>>
```

設定網路的輸入資料來源

Matlab操作—呼叫類神經網路建構工具(續)

Network Name: network1

Network Type: Self-organizing map

Input ranges: [0 95;5 85] Get from i...

Dimensions of map: [5 2]

Topology function: HEXTOP

Distance function: LINKDIST

Ordering phase learning rate: 0.9

Ordering phase steps: 1000

Tuning phase learning rate: 0.02

Neighborhood distance: 1.0

設定map的權重數目 [5 2]即為10群

Workspace:

Name	Size	Bytes	Class
ans	2x100	1600	doubl
c101	100x2	1600	doubl

Command Window:

```
>> mtool  
>>
```

SOM and K-means example

Matlab操作——呼叫類神經網路建構工具(續)

Network Name: network1

Network Type: Self-organizing map

Input ranges: [0 95,5 85] Get from i...

Dimensions of map: [5 2]

Topology function: HEXTOP

Distance function: LINKDIST

Ordering phase learning rate: 0.9

Ordering phase steps: 1000

Tuning phase learning rate: 0.5

Neighborhood distance: 0.5

View Defaults Cancel Create

Networks and Data: Help New Data... New Network... Import... Export... View Delete

Networks only: Initialize... Simulate... Train... Adapt...

Workspace: ans (2x100, double), c101 (100x2, double)

Command Window: >> mtool

設定學習速率和鄰近半徑值

Matlab操作——呼叫類神經網路建構工具(續)

得到設定好的網路

設定網路所需的相關架構參數後，接著選擇“Train...”，將Train Pattern輸入，以便網路進行學習模擬

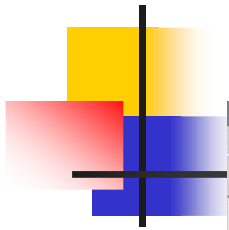
Name	Size	Bytes	Class
ans	2x100	1600	doubl
c101	100x2	1600	doubl

Command Window:

```
>> mntool
```

42	40	40	38	38	35	50	50	50
15	5	15	5	15	5	30	35	40
through 72								
35	40	30	35	30	95	95	53	
through 78								
90	88	9	87	85	85	85	85	
65	35	30	35	30	25	35	35	
through 90								
70	68	66	65	65	63	60	60	
through 99								
58	60	55	55	60	58	55	55	
65	65	62	60	60	58	55	55	
85	82	80	80	85	75	80	80	
55								
85								

Matlab操作—呼叫類神經網路建構工具(續)



選擇Clustering所需的
input pattern

Name	Size
ans	2x100
c101	100x2

Training Data

Inputs: ans

Targets: (zeros)

Init Input Delay States: (zeros)

Init Layer Delay States: (zeros)

Training Results

Outputs: network1_outputs

Errors: network1_errors

Final Input Delay States: network1_inputStates

Final Layer Delay States: network1_layerStates

58 60 55

through 99

55 65 62 60 60 58 55

85 82 80 80 85 75 80

55

85

```
>> mtool
>>
```

SOM and K-means example

Matlab操作—呼叫類神經網路建構工具(續)

設定欲學習次數

待設定完各相關參數後,
即可選擇“Train Network”
進行網路的Training

35	50	50	50
5	30	35	40
45	95	95	53
35	30	35	30
88	87	85	85
35	30	25	35
65	65	63	60
55	60	58	55
65	65	62	60
85	82	80	80
60	60	58	55
85	75	80	

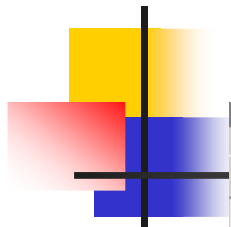
Matlab操作—呼叫類神經網路建構工具(續)

按“Export”將結果匯出

練訓完後得到網路輸出結果

SOM and K-means example

Matlab操作—呼叫類神經網路建構工具(續)



將網路和輸出結果Export

Export or Save from Network/Data Manager

Select Variables:

- network1
- ans
- network1_outputs

Select one or more variables. Then [Export] the variables to the MATLAB workspace or [Save] them to a disk file

Select All Select None Cancel Export Save

Networks and Data

Help New Data... New Network... Import... Export... View Delete

Networks only

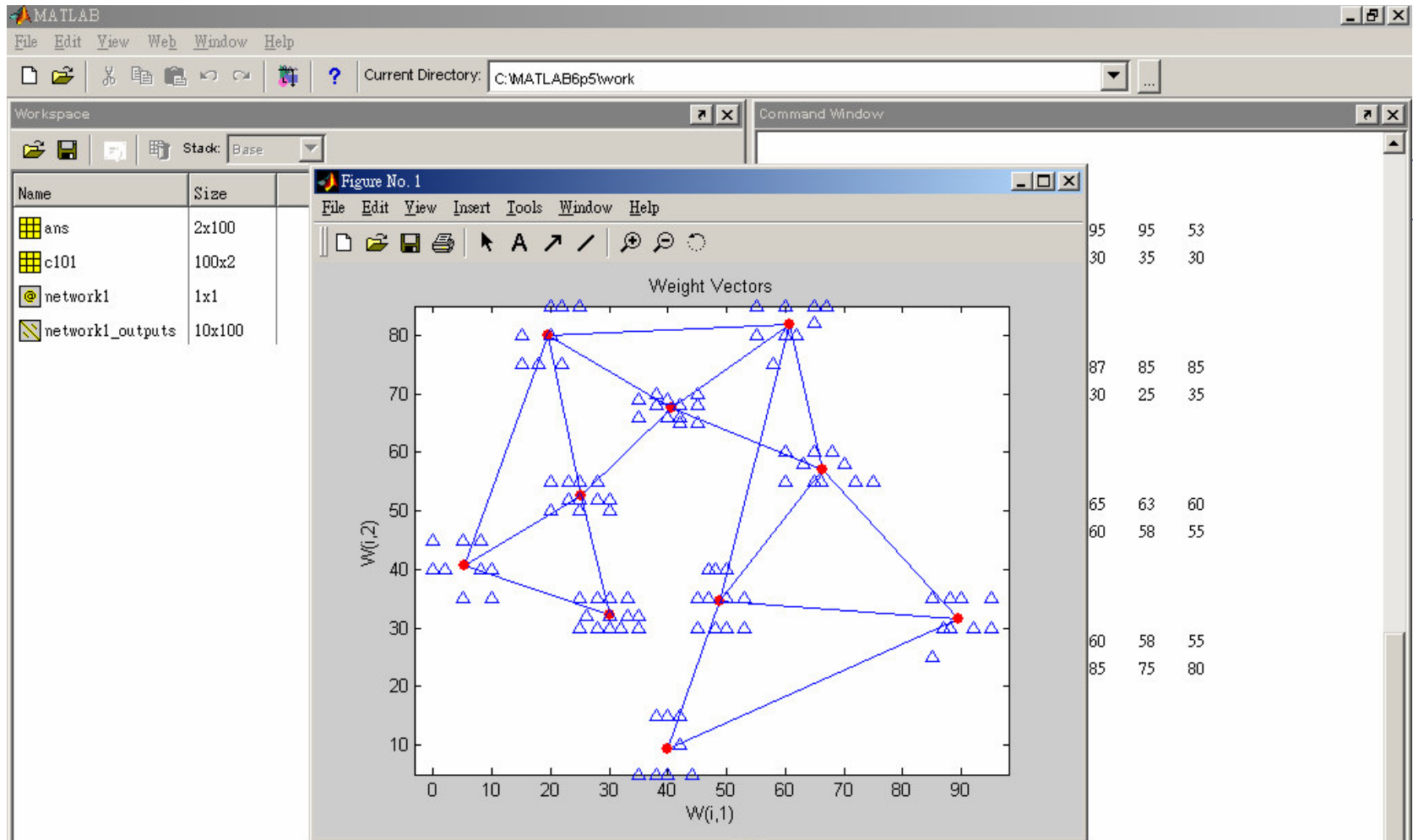
Initialize... Simulate... Train... Adapt...

Workspace

Name	Size	Bytes	Class
ans	2x100	1600	double
c101	100x2	1600	double

```
>> mtool
>>
```

確定後,選擇“Export”



在命令列輸入

plotsom(network1.iw{1,1},network1.layers{1}.distances)

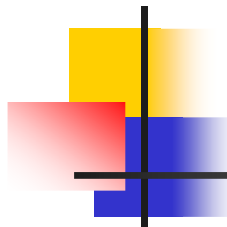
(輸入繪圖指令將群聚中心點標示出來)

hold on(保留圖中的群聚中心點)

plot(ans(1,:),ans(2,:),'^b')

(輸入繪圖指令將資料點標示出來)

Matlab操作—呼叫類神經網路建構工具(續)



Name	Size	Bytes	Class
ans	2x100	1600	double array
c101	100x2	1600	double array
network1	1x1	18685	network object
network1_outputs	10x100	1604	double array (sparse)

按滑鼠左鍵二下,
以觀看其群聚結果

(8,1)代表資料點1分在第8群

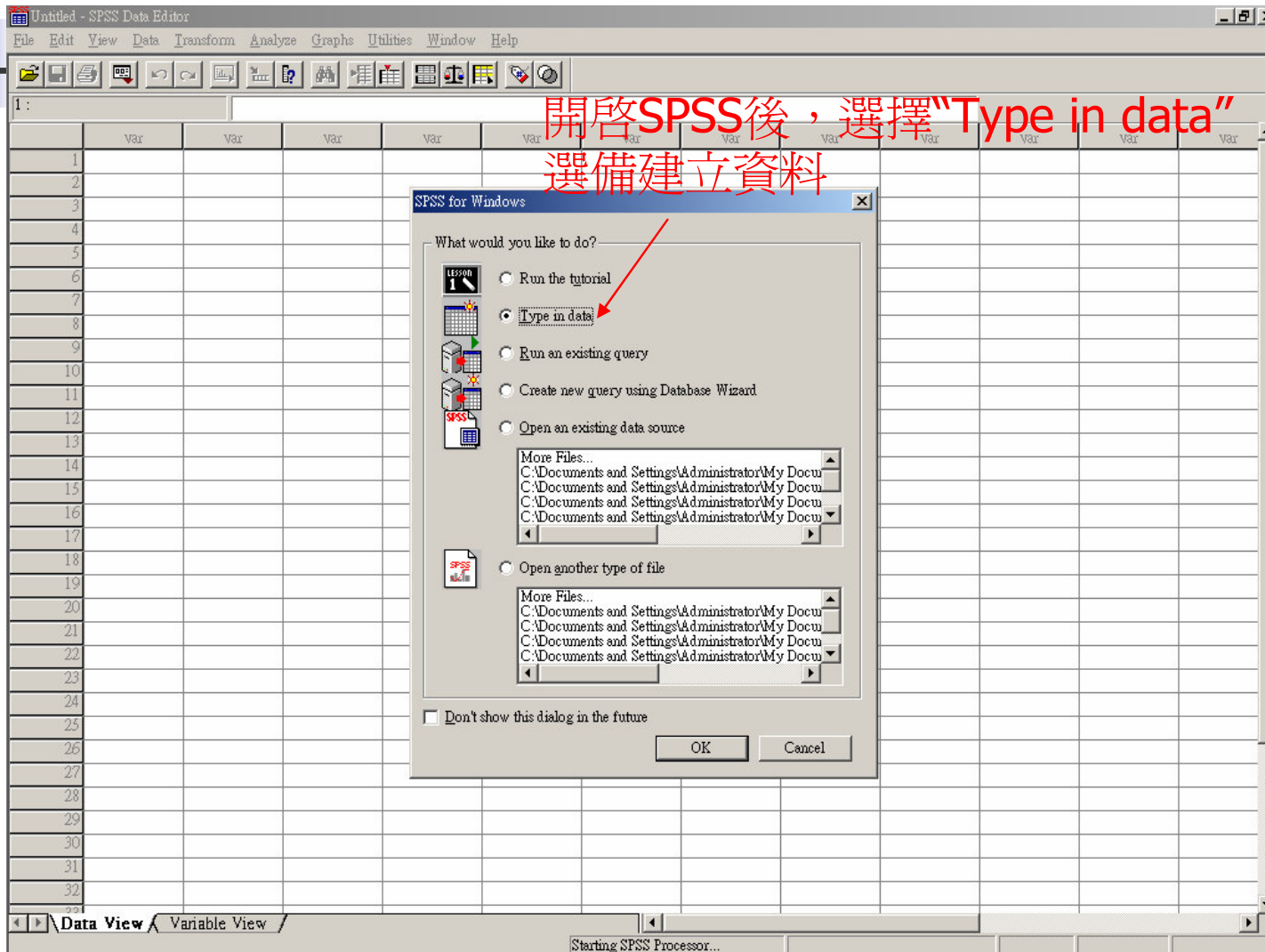
```
>> nntool
>> plotsom(network1.iw{1,1},network1.layers{1}.distances)
>> hold on
>> plot(ans(1,:),ans(2,:), 'Ab')
>>
```



K-means操作步驟

- 資料的輸入
- SPSS操作

SPSS操作-資料的輸入



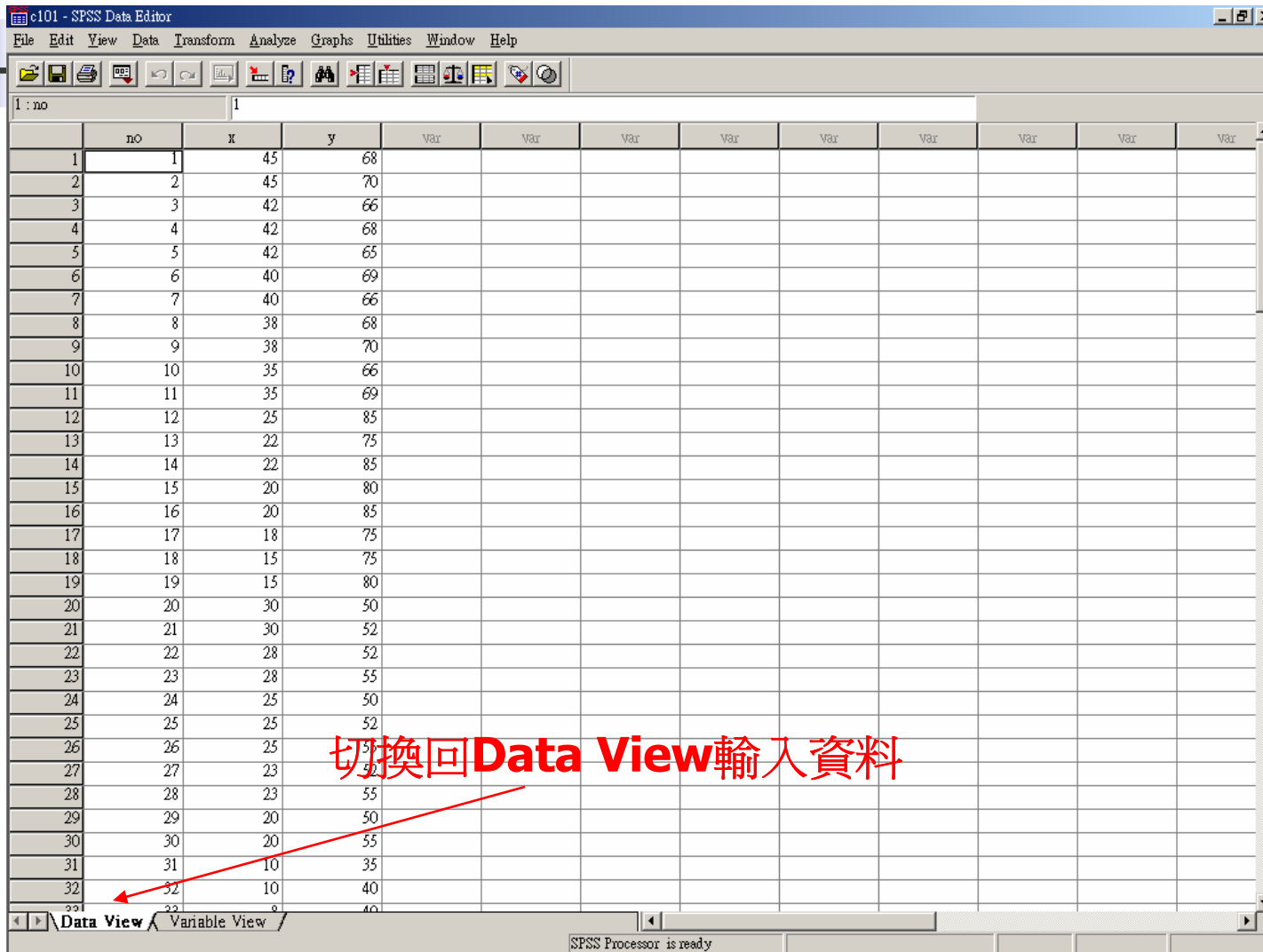
SPSS操作-資料的輸入(續)

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	no	Numeric	3	0		None	None	8	Right	Scale
2	x	Numeric	3	0		None	None	8	Right	Scale
3	y	Numeric	3	0		None	None	8	Right	Scale
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										
31										
32										
33										
34										

切換至“variable view”
鍵入所需變數名稱

SOM and K-means example

SPSS操作-資料的輸入(續)



SPSS Data Editor window showing a data table with the following columns: no, x, y, and several empty columns labeled 'Var'. The data is as follows:

	no	x	y	Var	Var	Var	Var	Var	Var	Var	Var	Var
1	1	45	68									
2	2	45	70									
3	3	42	66									
4	4	42	68									
5	5	42	65									
6	6	40	69									
7	7	40	66									
8	8	38	68									
9	9	38	70									
10	10	35	66									
11	11	35	69									
12	12	25	85									
13	13	22	75									
14	14	22	85									
15	15	20	80									
16	16	20	85									
17	17	18	75									
18	18	15	75									
19	19	15	80									
20	20	30	50									
21	21	30	52									
22	22	28	52									
23	23	28	55									
24	24	25	50									
25	25	25	52									
26	26	25	55									
27	27	23	52									
28	28	23	55									
29	29	20	50									
30	30	20	55									
31	31	10	35									
32	32	10	40									
33	33	0	40									

SOM and K-means example

SPSS操作-群聚分析

The screenshot shows the SPSS Data Editor interface. The 'Analyze' menu is open, and the 'Classify' sub-menu is selected. Within 'Classify', the 'K-Means Cluster...' option is highlighted. A red arrow points from the Chinese text below to this option.

no	x
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	38
9	38
10	35
11	35
12	25
13	22
14	22
15	20
16	20
17	18
18	15
19	15
20	30
21	30
22	28
23	28
24	25
25	25
26	25
27	23
28	23
29	20
30	20
31	10
32	10

選擇群聚分析工具下的“K-means Cluster”

SPSS操作-群聚分析(續)

1. 將x和y座標選為變數，no選為標籤項

2. 群數的設定(可由系統隨機選取起始群心)，亦可利用檔案匯入自設的群心

no	x	y
1	45	68
2	45	70
3	42	66
4	42	68
5	42	65
6	40	69
7	40	66
8	38	68
9	38	70
10	35	66
11	35	69
12	25	85
13	22	75
14	22	85
15	20	80
16	20	85
17	18	75
18	15	75
19	15	80
20	30	50
21	30	52
22	28	52
23	28	55
24	25	50
25	25	52
26	25	55
27	23	52
28	23	55
29	20	50
30	20	55
31	10	35
32	10	40

SPSS操作-群聚分析(續)

The screenshot shows the SPSS Data Editor window with a data grid. Overlaid on the grid are two dialog boxes. The 'K-Means Cluster Analysis' dialog box has 'Variables' set to 'x' and 'y', 'Label Cases by' set to 'no', 'Number of Clusters' set to 10, and 'Method' set to 'Iterate and classify'. The 'Cluster Centers' section has 'Read initial from' checked. The 'K-Means Cluster Analysis: Read from File' dialog box shows the search location 'clustering example' and the file 'center' selected. A red arrow points from the 'Read initial from' checkbox to the 'center' file in the second dialog box.

no	x	y	Var	Var	Var	Var	Var	Var	Var
1	1	45	68						
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18	18	15	75						
19	19	15	80						
20	20	30	50						
21	21	30	52						
22	22	28	52						
23	23	28	55						
24	24	25	50						
25	25	25	52						
26	26	25	55						
27	27	23	52						
28	28	23	55						
29	29	20	50						
30	30	20	55						
31	31	10	35						
32	32	10	40						
33									
34									

勾選“匯入檔案”，選擇所設群心的檔案
(群心檔如右所示)

The screenshot shows the 'center' file data in the SPSS Data Editor. The data grid has columns 'cluster_', 'x', 'y', 'Var', 'Var', and 'Var'. The data is as follows:

cluster_	x	y	Var	Var	Var
1	60.0000	80.0000			
2	42.0000	63.0000			
3	18.0000	79.0000			
4	27.0000	51.0000			
5	5.0000	43.0000			
6	26.0000	27.0000			
7	11.0000	13.0000			
8	40.0000	9.0000			
9	61.0000	18.0000			
10	54.0000	42.0000			
11					
12					

SOM and K-means ex

SPSS操作-群聚分析(續)

The screenshot shows the SPSS Data Editor interface with a data table and two dialog boxes. The data table has columns 'no', 'x', and 'y' with 32 rows of data. The K-Means Cluster Analysis dialog box is open, showing the following settings:

- Variables: x, y
- Label Cases by: no
- Number of Clusters: 10
- Method: Iterate and classify, Classify only
- Cluster Centers: Read initial from File... C:\...center.sav, Write final as File...

The Save New Variables dialog box is also open, showing the following settings:

- Cluster membership
- Distance from cluster center

A red arrow points from the 'Save...' button in the K-Means dialog box to the 'Continue' button in the Save New Variables dialog box.

按下此鈕可選擇是否在表格中另存群聚後的新資料，
選擇後，按continue繼續

SOM and K-means example

SPSS操作-群聚分析(續)

The screenshot shows the SPSS Data Editor interface with a data table and two dialog boxes. The data table has columns 'no', 'x', and 'y'. The 'K-Means Cluster Analysis' dialog box is open, showing 'Variables' (x, y), 'Label Cases by' (no), 'Number of Clusters' (10), and 'Method' (Iterate and classify). The 'Options' dialog box is also open, showing 'Statistics' (Initial cluster centers, ANOVA table, Cluster information for each case) and 'Missing Values' (Exclude cases listwise, Exclude cases pairwise). Red arrows point to the 'OK' button in the main dialog and the 'Continue' button in the options dialog.

no	x	y
1	45	68
2	45	70
3	42	66
4	42	68
5	42	65
6	40	69
7	40	66
8	38	68
9	38	70
10	35	66
11	35	69
12	25	85
13	22	75
14	22	85
15	20	80
16	20	85
17	18	75
18	15	75
19	15	80
20	30	50
21	30	52
22	28	52
23	28	55
24	25	50
25	25	52
26	25	55
27	23	52
28	23	55
29	20	50
30	15	55
31	10	35
32	10	40

最後按下ok進行群聚動作

按下此鈕可選擇在“輸出報表”中要顯示的資訊，選擇後，按continue繼續

SPSS操作-輸出報表資訊

Quick Cluster

Initial Cluster Centers

	Cluster								
	1	2	3	4	5	6	7	8	9
X	60	42	18	27	5	26	11	40	61
Y	80	63	79	51	43	27	13	9	18

Input from FILE S subcommand

Iteration History^a

Iteration	Change in Cluster Centers								
	1	2	3	4	5	6	7	8	9
1	2.771	3.924	1.908	2.386	2.467	6.524	.	.395	31.557
2	.000	1.601	.000	.000	.000	1.082	.	.000	.000
3	1.819	.000	.000	.000	.000	.950	.	.000	.000
4	2.305	.000	.000	.000	.000	1.824	.	.000	.000
5	.000	.000	.000	.000	.000	2.578	.	.000	.000
6	.000	.000	.000	.000	.000	2.445	.	.000	.000
7	.000	.000	.000	.000	.000	.000	.	.000	.000

a. Convergence achieved due to no or small distance change. The maximum distance by which any center has changed is .000. The current iteration is 7. The minimum distance between initial centers is 19.209.

Cluster Membership

Case Number	NO	Cluster	Distance
1	1	2	4.445
2	2	2	5.075
3	3	2	2.063
4	4	2	1.502
5	5	2	2.873
6	6	2	1.609
7	7	2	1.600

SPSS Processor is ready

SPSS操作-群聚結果製圖

The screenshot shows the SPSS Output Viewer window. On the left, the 'Output' tree is visible with 'Scatter...' selected. A 'Scatterplot' dialog box is open, showing 'Simple' and 'Overlay' options. A red arrow points from the 'Scatter...' menu item to the dialog box. Another red arrow points from the 'Simple' option to the 'Define' button. The main area of the viewer displays several tables:

	1	2	3	4	5	6	7	8	9
8.853	5.800	6.957	6.550	4.314	6.942	4.867	5.734	7.341	
.000	.000	.000	.000	2.181	.000	.933	.000	.000	
.000	.000	.000	.000	.000	.000	.000	.000	.000	

a. Convergence achieved due to no or small distance change. The maximum distance by which any center has changed is .000. The current iteration is 12. The minimum distance between initial centers is 20.616.

	Cluster								
	1	2	3	4	5	6	7	8	9
X	66	89	61	40	49	5	30	25	20
Y	57	32	82	9	35	41	32	53	80

Cluster	Number of Cases
1	10.000
2	9.000
3	9.000
4	8.000

1.選擇"scatter plot"散佈圖

2.選擇"Simple"後按"Define"

SPSS操作-群聚結果製圖(續)

Output2 - SPSS Viewer

File Edit View Insert Format Analyze Graphs Utilities Window Help

Output

- Quick Cluster
 - Title
 - Notes
 - Initial Cluster Centers
 - Iteration History
 - Final Cluster Centers
 - Number of Cases in each Cluster

→ Quick Cluster

Initial Cluster Centers

Simple Scatterplot

Y Axis: y

X Axis: x

Set Markers by: Cluster Number of Case

Label Cases by: no

Template

Use chart specifications from:

File... Titles... Options...

Cluster	1	2	3	4	5	6	7	8	9
X	66	89	61	40	49	5	30	25	20
Y	57	32	82	9	35	41	32	53	80

Number of Cases in each Cluster

Cluster	1	2	3	4
Cluster	1	10.000		
	2	9.000		
	3	9.000		
	4	8.000		

SPSS Processor is ready

將變數分別匯入後，按“ok”

SPSS操作-群聚結果製圖(續)

