

國立中央大學

數學研究所
碩士論文

自動分類軟體在動作電位上的研究

A Study of Spike Sorting by Autoclass

研究生：謝博文

指導教授：單維彰博士

中華民國九十五年六月



國立中央大學圖書館 碩博士論文電子檔授權書

(93年5月最新修正版)

本授權書所授權之論文全文電子檔，為本人於國立中央大學，撰寫之碩/博士學位論文。(以下請擇一勾選)

- ()同意 (立即開放)
()同意 (一年後開放)，原因是：_____
()同意 (二年後開放)，原因是：_____
()不同意，原因是：_____

以非專屬、無償授權國立中央大學圖書館與國家圖書館，基於推動讀者間「資源共享、互惠合作」之理念，於回饋社會與學術研究之目的，得不限地域、時間與次數，以紙本、微縮、光碟及其它各種方法將上列論文收錄、重製、公開陳列、與發行，或再授權他人以各種方法重製與利用，並得將數位化之上列論文與論文電子檔以上載網路方式，提供讀者基於個人非營利性質之線上檢索、閱覽、下載或列印。

研究生簽名：謝博文

論文名稱：自動分類軟體在動作電位上的研究

指導教授姓名：單維彰

系所：數學研究所 博士 碩士班

學號：93221019

日期：民國95年7月3日

備註：

1. 本授權書請填寫並親筆簽名後，裝訂於各紙本論文封面後之次頁（全文電子檔內之授權書簽名，可用電腦打字代替）。
2. 請加印一份單張之授權書，填寫並親筆簽名後，於辦理離校時交圖書館（以統一代轉寄給國家圖書館）。
3. 讀者基於個人非營利性質之線上檢索、閱覽、下載或列印上列論文，應依著作權法相關規定辦理。

國立中央大學碩士班研究生

論文指導教授推薦書

數學 學系/研究所 謝博文 研究生

所提之論文

自動分類軟體在動作電位上的研究

係由本人指導撰述，同意提付審查。

指導教授 單維勤 (簽章)

95 年 6 月 22 日

國立中央大學碩士班研究生
論文口試委員審定書

數學 學系/研究所 謝博文 研究生

所提之論文

自動分類軟體在動作電位上的研究

經本委員會審議，認定符合碩士資格標準。

學位考試委員會召集人

委

員

蕭孟列

蕭嘉璋

單維新

中華民國 95 年 6 月 28 日

論文摘要

為了分析一顆神經細胞的行為，我們需要了解其發生動作電位的頻率變化。通常一根微電極會接收到多個神經細胞所發出的動作電位以及一些環境雜訊，因此對於這些混雜的神經訊號，我們處理的步驟是先過濾雜訊，之後再將不同神經細胞所發出的動作電位分類。

主成分分析法最常被用於處理動作電位的分類問題上，但分類結果需由使用者觀看資料分布圖來決定，所以客觀性不高。Autoclass 是目前網路上公開的一種分類軟體，其分類結果會顯示每筆資料屬於每一群的機率值。以機率值大小將資料分群的結果比用眼睛判斷的分類結果還具客觀性。本論文的目的是使 Autoclass 能有效應用在處理神經細胞動作電位的工作上，同時提出一個先驗標準來判斷一組訊號是否該以 Autoclass 分類，以及一個後驗方法來評估分類的結果是否可以信賴。

關鍵字：動作電位分類，神經動作電位，Autoclass。

Abstract

In order to analyze the actions of a neuron, we need to understand the variations of the frequency of its spikes. An electrode usually receives spikes generated by several neurons as well as some background noises, which become interference in our analysis. Therefore, the task is to first filter out noises from these electrical signals and then to classify the spikes.

In spike sorting, Principal Component Analysis (PCA) is the method scholars tend to adopt most often. However, the classification produced by PCA is not objective for it is based on operators' observation of the graph. An alternative choice, Autoclass, is a popular method for clustering and its classification shows the probability of each data which belongs to each cluster. We can get a more objective classification according to probability than by visual estimation. Hence, a classification produced by Autoclass is more objective than one produced by PCA. The goal of this article is to show how Autoclass makes spike sorting more effective, meanwhile, to provide a prior knowledge to determine whether a data is suitable to be clustered by Autoclass, and to propose a posterior method to confirm the validity of its classification.

Keywords: spike sorting, neural action potential, Autoclass.

目錄

第 1 章 緒論	1
第 2 章 背景知識	2
2.1 動作電位	2
2.2 門檻探測法	3
2.3 主成分分析法	3
第 3 章 Autoclass 使用指引	4
3.1 Autoclass 簡介.....	4
3.2 如何使用 Autoclass	5
3.2.1 搜尋	5
3.2.1.1 Data 檔案	5
3.2.1.2 Header 檔案	7
3.2.1.3 Model 檔案	11
3.2.1.4 Search params 檔案	13
3.2.1.5 執行 Search 指令	20
3.2.2 報告.....	21
3.2.2.1 Report params 檔案	21
3.2.2.2 執行 Report 指令	23
3.2.2.3 influence 檔案、case 檔案、class 檔案	23
第 4 章 Autoclass 測試	24
4.1 參數的測試	24
4.2 測試最佳分類的準確性.....	25
第 5 章 Autoclass 實驗	26
5.1 實驗一.....	26
5.1.1 測試一.....	26
5.1.2 測試二.....	26
5.1.3 測試三.....	27

5.2 實驗二	28
5.3 實驗三	29
5.4 實驗四	30
5.5 實驗五	31
第六章 結論與建議	32
6.1 Autoclass 的優缺點	32
6.2 Autoclass 處理分類問題步驟之建議	33
參考書目	34
附錄	35
附錄一	35
附錄二	36
附錄三	37
附錄四	39
附錄五	40
附錄六	40
附錄七	42
附錄八	42
附錄九	42
附錄十	43
附錄十一	44

圖目

圖一 動作電位門檻圖	45
圖二 測試資料之分布圖	45
圖三 五次的最佳分類圖	45
圖四 測試資料之分布圖	46
圖五 多次的最佳分類圖	46
圖六 第二十六次的最佳分類圖	46
圖七 第二十九次的最佳分類圖	47
圖八 人工動作電位圖	47
圖九 人工動作電位加上真實雜訊	47
圖十 PCA 圖	48
圖十一 PCA 圖	48
圖十二 實驗一之主成分個數與分類正確率關係圖	48
圖十三 PCA 圖	49
圖十四 人工動作電位圖	49
圖十五 人工動作電位加上真實雜訊	49
圖十六 PCA 圖	50
圖十七 PCA 圖	50
圖十八 PCA 圖	50
圖十九 實驗二之主成分個數與分類正確率關係圖	51
圖二十 主成分重要性與分類正確率關係圖	51
圖二十一 人工動作電位圖	52
圖二十二 人工動作電位加上真實雜訊	52
圖二十三 PCA 圖	52
圖二十四 PCA 圖	53
圖二十五 PCA 圖	53
圖二十六 人工動作電位圖	53
圖二十七 人工動作電位圖	54
圖二十八 m_dn_rat 與分類正確率關係圖	54

表 目

表一 年齡、身高、體重數據表.....	5
表二 年齡、身高、體重數據表.....	6
表三 屬性設定對應表.....	10

第 1 章 緒論

關於神經系統網路運作的研究，目前最熱門的方式是影像處理法，利用磁振造影來觀察大腦的運作，但由於解析度的問題，所以無法看到整個神經細胞的運作以及了解傳遞訊號的過程。另一種研究方式是透過一些細的電極來記錄神經元在受刺激下的反應，此方法的解析度可以到達一個神經細胞，且目前捕捉電訊號的技術非常發達，因此我們可以分析網路中單一神經細胞的動作電位放電狀況，甚至是某些區域細胞間的交互作用。然而，以電極記錄活體生物的神經訊號會有一個技術上的問題，由於記錄端置於細胞外面，所以停留的地方附近不只有一顆神經細胞，導致電極可能記錄到許多神經細胞的訊號。

爲了分析一顆神經細胞的行爲，我們需要了解其發生動作電位的頻率變化，因此對於細胞外電極所記錄到混雜的神經訊號，我們需將不同神經細胞所發出的動作電位分類。至於訊號分類是根據以下兩個基本假設：一是不同神經元相對於同一根電極所展現的動作電位波形是不一樣的；二是每一顆神經元所產生的動作電位是不會隨著時間改變的。

Autoclass 是目前網路上所公開的一種分類軟體，本研究的目的是希望找到一個好的程序和參數設定，使其能有效應用在處理神經細胞動作電位的工作上，並且將操作 **Autoclass** 的過程寫成一篇使用指引，希望能提供未來的使用者作參考。

第 2 章 背景知識

2.1 動作電位

神經元是神經系統的構造，主要是傳遞大腦與身體各處間的訊息，當沒有訊息傳遞時，神經元會處於極化狀態，此時細胞膜內外保持一定的電位差，稱為靜止膜電位。通常哺乳類動物的膜外電位會比膜內電位高 70mv 左右，所以若定義膜外電位為 0mv，則膜內電位則是 -70mv。

造成細胞膜內外靜止膜電位差的原因有二，一是細胞內外的離子濃度分布不同，細胞內 k^+ 濃度高於細胞外，細胞內 Na^+ 低於細胞外。二是神經細胞膜內具有帶負電性的蛋白質和有機磷酸化合物，使得細胞膜內的電位較細胞膜外的電位低。當神經元受刺激時，其細胞膜的通透性會發生急劇變化。首先是 Na^+ 的通道打開，因為細胞膜內外電位及 Na^+ 濃度的差異，使得 Na^+ 瞬間大量內流，此過程稱為去極化。當去極化達到高峰時，細胞膜內的電位會高於細胞外，同時 Na^+ 通道會關閉， k^+ 通道開到最大，有利於 k^+ 從細胞膜內流到細胞膜外，此過程稱為再極化。從去極化到再極化結束的膜電位變化就稱為一個動作電位。

動作電位的產生遵循全或無定律，當刺激夠大時，就會產生動作電位，否則就不發生，不會有介於兩者之間的情況，且刺激大到足以發生動作電位時，不論刺激有多大，動作電位的大小都不會產生變化，只不過較強烈的刺激會產生較多次的動作電位，也就是動作電位的發生頻率會比較大。

神經細胞所發生的動作電位，可以被微電極記錄下來，不過隨著微電極與同一神經元的相對位置不同，所探測到的動作電位波形也有所不同。因此神經元發出的動作電位，實際上是指微電極相對於神經元位置所收到的電位變化。由於每個神經元相對電極的位置是有差異的，故我們假設微電極所記錄到的不同神經元動作電位是不相同的。

2.2 門檻探測法

生物學家在探測生物訊號時常會受到雜訊的干擾，其干擾分為兩種，一是環境雜訊，二是生物體本身產生的干擾。生物學家慣用門檻探測法來搜尋動作電位。

Thresholding 的功用是從原始生物訊號中過濾出可以使用的動作電位。此方法步驟是先根據探測到的電訊號來決定一個門檻值，之後以這個門檻值來判別一個波形能不能視為一個動作電位。以圖一為例，假設我們認為振幅要超過 $40\mu\text{V}$ 才是動作電位，則圖一中只有 4ms 附近有一個動作電位，其餘的都被視為雜訊[3]。

通常微電極附近會有多個神經元，如果有多數神經元同時產生動作電位，則這些動作電位會以一個合成波的形式被記錄到，但在本論文中不討論此種情況，我們假設每一個被探測到的動作電位，是屬於某一個神經細胞所發出來的訊號。

2.3 主成分分析法

經由門檻探測法得到一些動作電位之後，接著我們要將這些訊號加以分析。最常見的分析資料方法就是主成分分析法（Principal Component Analysis），簡稱 PCA，其目的是希望用較少的變數來分辨原始資料間的差異。假設我們手上有 n 筆資料，稱為 v_1, v_2, \dots, v_n ，對任意一個 i ， v_i 都是 m 維向量。在我們分析 $\{v_i\}_{i=1}^n$ 時， m 越小，就越容易分析。由於 m 的大小是不能控制的，所以為了分析方便，我們通常會選擇一個程序 Γ ，使得 Γ 會從每一個 v_i 得到一個 s 維的向量 w_i 。例如我們可以選擇 Γ 是求平均值，則 $s = 1$ ，而對於每一個 i ， w_i 是 v_i 的平均值。一般來說， $\{w_i\}_{i=1}^n$ 包含的資訊量會比 $\{v_i\}_{i=1}^n$ 少，但是前者通常可以表現出一甚至於突顯出一後者的某些特徵。我們希望 Γ 能夠表現出足夠的特徵，使得在只考慮這些特徵的情況下，對 $\{w_i\}_{i=1}^n$ 的分析等同於對 $\{v_i\}_{i=1}^n$ 的分析。求平均值雖然是一個簡單且普遍的程序，但是他表現的特徵並不足以應付大多數的需求。在多變量統計領域裡，有許多的程序供我們選擇，PCA 即是其中一種。詳細內容請參考李易霖[1]。

第 3 章 Autoclass 使用指引

Autoclass 程式可從下列網址下載：

<http://ic.arc.nasa.gov/ic/projects/bayes-group/autoclass/>

其操作的相關說明可從程式附帶之文件檔中取得，而有關數學理論的部份可見於 Bayesian classification (autoclass) [2]。本章節的內容是根據長時間使用 Autoclass 的心得所編寫而成的，希望能提供使用者參考。

3.1 Autoclass 簡介

早期，貝斯小組在資料分析技術上，發展出一套基本的理論和相關程式。他們將這些理論和技術應用在自動分類的問題上，而設計出 Autoclass 自動化的分類程式。

Autoclass 具有以下的主要特色：

- Autoclass 是一種 **unsupervised** 的分類法，此分類法具有與一般分類法不同的特色。一般分類法在執行時，大多只能先假設資料分成 N 群，而且每一筆資料在分類後只能屬於這 N 群中的某一群，但是 Autoclass 並非如此，使用者除了可以一次設定好幾種群數讓 Autoclass 搜尋之外，在搜尋結束後，還能得到每一筆資料分別屬於每一群的機率，所以使用者能根據自己的專業，決定出比較好的分類結果。
- 檔案重複使用的便利性。在 Autoclass 中，每一筆資料都以一個向量來表示，其中每個分量都分別代表不同的屬性，這些屬性數據可以是連續型或是離散型，Autoclass 要求使用者將資料存成 Data File (存放資料) 與 Header File (描述資料屬性) 兩部分，如此可以讓使用者自由搭配 Data File 和 Header File 而節省輸入資料的時間。例如有 Data1、Data2、Header1 和 Header2 四個檔案，在格式符合條件下，可以直接使用 Data1 Header2 來執行。若是原本 Data File 與 Header File 是儲存在同一個檔案中，這樣資料在使用上就不怎麼方便。
- 可以處理 missing value。當一筆資料中的某些屬性值有缺漏時，Autoclass 仍可將此筆資料分類。

3.2 如何使用 Autoclass

Autoclass 主要有兩個執行指令，分別是：Search 及 Report。其中 Report 必須依據 Search 的結果才能執行，所以我們要先了解如何進行 Search 這個部份。

3.2.1 搜尋

在進行 Search 指令前，使用者必須先準備好四個檔案：Data file、Header file、Model file 和 Search params file，以下詳述個別的內容及儲存格式。

3.2.1.1 Data 檔案

Data file 簡單來說就是存放資料的地方。Autoclass 會將此檔案內的每一列視為一筆資料，而每一行視為一種屬性；因此代表一筆資料的屬性值必須寫在同一列裡，同時每一種屬性值都是以一個空白隔開（預設）。我們用例子來說明比較容易了解，假如有 4 個人的年齡、身高和體重等數據分別如表一：

編號	年齡	身高	體重
0	25	170	60
1	25	172	55
2	16	165	53
3	16	178	80

表一 年齡、身高、體重數據表

則 Data file 對照輸入的標準格式為

```
25 170 60
```

```
25 172 55
```

```
16 165 53
```

```
35 178 80
```

或是

```
;表一 Data file
```

```
0 25 170 60
```

```
1 25 172 55
```

```
2 16 165 53
```

```
3 16 178 80
```

這裡請使用者注意，同一行的數值必須是相同的屬性，而且若是某一列的開頭出現「!」、「#」、「;」、「 」或是「\n」(empty line)，則讀取資料時會忽略這一行，因此這些符號可以用在註解上。再者，我們建議使用者以第 2 種格式來輸入資料，因為在處理龐大資料的時候，最左邊的列號可以使我們一目了然總共有多少筆資料。最後，當資料輸入完畢，儲存副檔名為 db2 即完成 Data file 的準備工作。

附註 1

若是一筆資料有 missing value 的時候（缺少某些屬性值），使用者請在該欄位輸入「？」，因為 Autoclass 要求每一欄位都必須有值，如此才可以避免在讀取資料時，產生數值與屬性不對應的問題。不過在進行分類的時候，這些「？」會是影響分類結果的因素之一，因為 Autoclass 會將同一行的「？」視為某一個正確值來進行分類。故當資料沒有 missing value 時，是依據資料的特性所以產生出來的分類結果，相反的，若資料具有大量的 missing value，分類結果顯示的是資料蒐集程序上的問題。附帶一提，有很強的先驗知識，才能根據分類結果找出資料蒐集上的問題點。

表二是具有 missing value 的資料：

編號	年齡	身高	體重
0		170	60
1	25	172	55
2	16	165	53
3		178	

表二 年齡、身高、體重數據表

則 Data file 對照輸入的標準格式為

;**表二 Data file**

0 ? 170 60

1 25 172 55

2 16 165 53

3 ? 178 ?

附註 2

雖然 Data file 沒有限制資料總筆數 (m) 和屬性總個數 (n) 的上限，但是當 $m*n$ 大於 10^5 時，會花費大量的時間來進行搜尋，所以先將 Data file 轉成 binary 的形式儲存，在執行上就可以節省時間且減少當機的機率。

附註 3

在表一 Data file 的第二種輸入格式中，Autoclass 仍會將第一行視為一種屬性，但是之後在設定 Header file 時，能使 Autoclass 忽略此行而進行分類，所以分類結果仍與第一種輸入格式的分類結果相同。

3.2.1.2 Header 檔案

Header file 主要是在描述 Data file 中的資料屬性，其檔案內容分為宣告及屬性設定兩個部份。我們承接表一的 Data file 來說明如何設定 Header file，為了方便講解，之後的每一個檔案都標有列數；當然列號不是檔案內容的一部分。類似於 Data file，若是某一列的開頭出現「!」、「#」、「;」、「」或是「\n」等符號，表示這一系列的其餘部分為註解。

```
1 ;表一 Header file
2 num_db2_format_defs 1
3 number_of_attributes 4
4 ;separator_char ';'
5 ;comment_char '%'
6 ;unknown_token '?'
7 ;<zero-based att#> <att_type> <att_sub_type> <att_description> <att_param_pairs>
8 0 dummy nil "case number"
9 1 discrete nominal 2
10 2 real scalar "h e i g h t" zero_point 0.0 rel_error 0.01
11 3 real scalar weight zero_point 0.0 rel_error 0.01
```

首先說明 Header file 的宣告部分。Header file 一開始必須有次序的宣告底下兩件事：

num_db2_format_defs n

number_of_attributes m

其中 n 是指之後還有幾列為宣告式，在表一 Header file 中，我們很容易看出第 2 列之後只有第 3 列是宣告式，所以 n 輸入 1。m 是指 Data file 中的行數，所以對照表一 Data file，m 輸入 4。除此之外，第 4、5、6 列也是 Header file 所允許的宣告式；第 4 列是宣告在 Data file 中，一筆資料的屬性值，可以使用引號內的符號來分開，上面採用的是逗號。第 5 列是宣告在 Data file 中，除了預設的註解符號外，新增引號內的符號為註解符號，上面新增的是百分比符號。第 6 列則是宣告引號內的符號來代表 Data file 中的 missing value，因此當資料有 missing value 時，記得做此宣告，上面選擇的是問號。

第二個要說明的是屬性的設定部份。觀看表一 Header file 的第 7 列，此註解列出每個屬性必須依序輸入的 5 個設定，且設定之間都以空白隔開。底下我們詳述其內容。

1. zero-based att：屬性的編號，從 0 開始，依序編號至 $m-1$ ，其中編號為 0 的屬性對應 Data file 的第一行，編號為 1 的屬性對應 Data file 的第二行，依此類推。而根據表一 Data file，屬性編號應由 0 至 3。
2. att_type：屬性的型態，分為 real、discrete、dummy 三種。選擇 dummy 的屬性，Autoclass 會將此屬性忽略，也就是分類的結果不受其影響。選擇 discrete 的屬性，通常為離散型態，其屬性值只能輸入符號、整數及字串。選擇 real 的屬性，通常為連續型態，其屬性值只能輸入數字，包含整數與浮點數。根據表一 Data file 中的屬性，第一行至第四行的屬性型態應依序為 dummy discrete real real，其中忽略第一行屬性的理由是因為第一行是我們自己加入的編號。

3. `att_sub_type`：屬性的子型態，分為 `none`、`nominal`、`location`、`scalar` 四種，使用者必須根據 `att_type` 來選擇子型態的形式。`dummy` 的子型態能設定 `none` 或 `nil`，其結果沒有什麼不同；`discrete` 的子型態只能設定 `nominal`；`real` 的子型態能設定 `location` 或 `scalar`，然而在選取上必須依據 `att_param_pairs` 決定。
4. `att_description`：屬性的註解區，使用者可以用英文對此屬性稍作敘述，以利於觀看分類的結果。此註解區的輸入不得超過 40 個字元，而且當敘述需要用到空白字元時，請在敘述的頭尾加上雙引號。
5. `att_param_pairs`：屬性的參數對，包含參數名與參數值的設定，輸入時需以空白隔開。屬性的參數名分為 `range`、`error`、`zero_point`、`rel_error` 三種，當 `att_type` 為 `dummy` 時，此欄位就不需要做任何設定；當 `att_type` 為 `discrete` 時，參數名就為 `range`，而參數值的決定，必須觀看屬性值種類個數，舉例來說，假設該屬性的屬性值有 2、4、3、?、?、6、2，總共包含 2、4、3、?、6 五種數據，因此參數值設定為 5；當 `att_type` 為 `real` 時，參數名可以選擇 `error` 或 `zero_point rel_error`，如果選定 `error`，則對應的 `att_sub_type` 為 `location`，且輸入的參數值代表一個絕對誤差範圍，相反的，如果選擇 `zero_point rel_error`，所對應的 `att_sub_type` 為 `scalar`，且輸入的參數值，一個代表新的原點，使得此屬性值相對平移後都不小於零，而另一個參數值代表相對誤差範圍。接下來說明的細節，可以讓使用者更了解這一部份的設定。

一般來說，選擇 `real` 的連續型屬性，其數據都有測量上的誤差，例如 `truncation error` 或是 `background noise` 造成的，所以使用者需提供 `Autoclass` 誤差範圍。若誤差範圍太大，可能將應屬於不同一群的資料歸類成同一群；若誤差範圍太小，可能將應屬於同一群的資料分成不同群。明顯的，選取的誤差範圍不夠精確，則產生的分類結果會不盡理想，所以使用者應該根據自己所知的知識，設定準確的誤差範圍。`Autoclass` 提供相對誤差與絕對誤差兩種輸入方式，輸入相對誤差時，`Autoclass` 會將此屬性的數據以 `zero_point` 為新原點做平移，之後把新的數據值取

log 轉換，同時將選擇的 model 改以 log-normal 分佈的型態來進行分類。輸入絕對誤差時，除了將絕對誤差除以屬性值平均來轉成相對誤差之外，在分類上就沒有任何轉換動作。至於使用者沒有誤差的相關資訊時，可以取 error 為數據彼此間的最小差距的一半。最後，儲存副檔名為 hd2 即完成 Header file 的準備工作。

附註 4

為了讓使用者方便設定，我們將 att_type、att_sub_type、att_param_pairs 的對應關係整合成表三，同時請使用者注意，此部份必須以小寫來輸入，否則執行時會產生類似底下的錯誤訊息。

att_type	att_sub_type	att_param_pairs
dummy	none/nil	
discrete	nominal	range
real	location	error
real	scalar	zero_point rel_error

表三 屬性設定對應表

ERROR[2]: unknown type/sub_type = **Real**/location
for attribute #3: "X_coordinate"

ERROR[2]: unknown type/sub_type = **Discrete**/nomial
for attribute #2: "Y_coordinate "

ERROR[2]: unknown type/sub_type = real/**Location**
for attribute #3: "X_coordinate"

ERROR[2]: unknown type/sub_type = discrete/**Nomial**
for attribute #2: "Y_coordinate "

附註 5

表二的 Header file 設定如下：

```

1 ;表二 Header file
2 num_db2_format_defs 2
3 number_of_attributes 4
4 ; separator_char ' '
5 ; comment_char ';'
6 unknown_token '?'
7 ; <zero-based att#> <att_type> <att_sub_type> <att_description> <att_param_pairs>

```

```
8 0 dummy nil "case number"
9 1 discrete nomial 2
10 2 real scalar "h e i g h t" zero_point 0.0 rel_error 0.01
11 3 real scalar weight zero_point 0.0 rel_error 0.01
```

3.2.1.3 Model 檔案

每個屬性都有各自的機率分佈範圍，因此對於每一個屬性，我們需給予一個合適的機率模型。以下我們就先介紹 Autoclass 所支援的五個機率模型，之後再詳述如何建構 Model File。

1. Single Multinomial Model

此 model 適用於屬性具有下列特徵：

- 型態為 discrete
- 每個屬性是 conditionally independent
- 具有 missing value

2. Single Normal CN Model

此 model 適用於屬性具有下列特徵：

- 型態為 real
- 每個屬性是 conditionally independent
- 沒有 missing value
- Gaussian normal 分佈

3. Single Normal CM Model

此 model 適用於屬性具有下列特徵：

- 型態為 real
- 每個屬性是 conditionally independent
- 具有 missing value
- Gaussian normal 分佈

4. Multi Normal CN Model

此 model 適用於屬性具有下列特徵：

- 型態為 real
- 每個屬性是 conditionally independent
- 沒有 missing value
- covariant normal 分佈

5. Ignore Model

選擇此 model 的屬性會被 Autoclass 忽略，也就是分類結果不受其影響，所以使用者可以將之作爲註解之功用，例如編號。

接下來我們承接表一的 Header file 來說明如何設定 Model file。

```
1 ;表一 model file
2 model_index 0 3
3 ingnore 0
4 single_normal_cn 2 3
5 single_multinomial_cn 1
6 ;single_normal_cm
7 ;multi_normal_cn
```

Model file 一開始必須宣告：

```
model_index 0 n
```

其中 n 是指之後還有幾列爲定義線，在上面例子當中，我們很容易看出第 2 列之後有第 3、4、5 列是定義線，所以 n 輸入 3。而每個機率模型後面的數字，則是代表 Header file 中的屬性編號。我們觀看第 4 列，意思是說，編號 2、3 的屬性都使用 `single_normal_cn` 爲機率模型，更進一步說明，不同屬性使用同一種機率模型時，只需將屬性編號以空白格開。最後，儲存副檔名爲 `model` 即完成 Model file 的準備工作。

3.2.1.4 Search params 檔案

這個檔案主要是有關 Autoclass-Search 的參數設定，使用者在這部份必須有相當的了解，底下我們敘述每個參數的功用。

start_j_list = 2, 3, 5, 7, 10, 15, 25

此參數預設為 2, 3, 5, 7, 10, 15, 25，其數字代表群數。Autoclass 在進行搜尋時，會先根據此部份所設定的群數依序分類，但若是欲分類群數與分類結果的群數差異很大時，Autoclass 則會隨機製造合適的群數來分類，而不會繼續之後的群數來分類。因此使用者應注意，在此輸入的群數會影響分類結果，所以使用者必須自己估計資料可能被分為的群數。假設我們相信資料可以被分為 75 群，那麼此設定可以輸入 50, 60, 70, 80, 90, 100。

n_classes_fn_type = "random_ln_normal"

此參數只能設定 "random_ln_normal"。當 Autoclass 根據 start_j_list 參數執行一陣子之後，可得 10 個最佳的分類結果，此參數的作用就是從這 10 個欲分類群數中隨機取一個當作下一次 try 的欲分類群數。

randomize_random_p = true

設定為 true 時，亂數產生器是以任意的一個時間點開始取亂數；若設定為 false，則亂數產生器每次固定從 1 開始取亂數。

min_report_period = k

此參數設定 k 時，表示兩次回傳結果的時間，最少間隔 k 秒。若使用者希望每次搜尋後的結果都相同，則必須將 k 值設得夠大，使得每一次回傳的報告能夠完整。此參數預設為 30 秒。

fixed_j = k

當 $k > 0$ 時，Autoclass 會忽略 start_j_list 及 n_classes_fn_type 的設定，同時只會將資料以 k 群來分類。

interactive_p = true

設定為 true 時，表示在執行 Autoclass-Search 的過程中，按下 Q 鍵可以停止收尋；設定為 false 時則否。

max_duration = k

k 為 0 時，表示搜尋不受時間的影響；k 為正數時，表示執行 k 秒後，若搜尋尚未結束，則強制停止搜尋。

max_n_tries = k

Autoclass 在執行 search 的時候，會製造一個粗糙的分類，之後透過局部的變動，使得變動後的分類結果有較高的機率值，如此持續到機率值不再改變，也就是收斂至某個“local maximum”。我們稱這一個過程為一個“try”，且每次局部的變動稱為一個 converge cycle。此參數設定 k 為 0 時，表示搜尋不受 try 的次數影響；k 為正整數時，表示搜尋 k 個 try 之後則停止搜尋。

rel_error = k

當 clsf-DS-% 的值 $\leq k$ ，則我們認定一個變動前與變動後的分類結果是完全一樣的。此參數預設為 0.01。

start_fn_type = "random"

決定一個 try 是以何種方式創造初始分類。設定為 random 時，Autoclass 隨機從資料中取樣，之後將這些取樣出來的資料視為不同群的中心，加上適當的變異量後再把其他資料分類；設定為 block 時，Autoclass 將資料分割成大小差不多的區塊，使其可以產生不變的分類結果。

try_fn_type = "converge_search_3"

決定機率值收斂至 local maximum 的標準，此參數提供三種收斂法則，分別是 converge_search_3、converge_search_4、converge。一般來說，converge_search_3 能處理的問題很多，尤其當資料很多且需要花費大量時間來進行分類時，結果都

會比其餘兩者來的好。

rel_delta_range = k

針對 try_fn_type 選擇 converge_search_3 時的進一步設定，預設為 0.0025。當 try_fn_type 設定為 converge_search_3 時，Autoclass 會對兩個連續 converge cycles 中的每一群做計算，個別得到一個比值，之後將前一個 cycle 中每一群所得到的比值，和後一個 cycle 中對應的群的比值做相減，當每一個差值都不超過 k，我們稱做是一次成功的收斂。設定的 k 愈大，表示放寬了收斂法則，同時 converge cycles 次數會減少，相反的，設定的 k 愈小，收斂法則較為嚴格，同時 converge cycles 次數會增加。

n_average = k

針對 try_fn_type 選擇 converge_search_3 或 converage 時的進一步設定，預設為 3。k 表示在進行一個 try 時，至少連續出現 k 次成功的收斂，才能終止這一個 try。

cs4_delta_range = k

針對 try_fn_type 選擇 converge_search_4 時的進一步設定，預設為 0.0025。當 try_fn_type 設定為 converge_search_4 時，Autoclass 會根據 sigma_beta_n_value 參數得到一個 cycle 的個數。假設將資料分為 n 群，且設定的 cycle 個數為 m，Autoclass 會先對 m 個 convergence cycles 中的每一群計算出個別的比值 x_{m1} 、

$x_{m2} \cdots x_{mm}$ ，之後令 $v_j = \frac{\sum_{i=1}^m x_{ij}}{m}$ $j = 1, 2, \dots, n$ ， $V = \max\{|v_k - v_l| : k \neq l\}$ ，若 V 不超過 k，則終止這一個 try。設定的 k 愈大，表示放寬了收斂法則，同時 converge cycles 次數會減少，相反的，設定的 k 愈小，表示收斂法則較為嚴格，同時 converge cycles 次數會增加。

sigma_beta_n_values = k

針對 try_fn_type 選擇 converge_search_4 時的進一步設定，預設為 6。k 是用來計算 σ^2 (noise) 和 β^2 (signal) 的數。

halt_range = k ; halt_factor = m

針對 try_fn_type 選擇 converge 時的進一步設定，halt_range 預設為 0.5，halt_factor 預設為 0.0001。當 try_fn_type 設定為 converge 時，Autoclass 將兩個連續 converge cycles 的 log_marginal classification 做相減，當相差值不超過 k 或是 $m \times \text{current_log_marginal}$ 時，我們稱做是一次成功的收斂。設定的 k 或 m 愈大，表示放寬了收斂法則，同時 converge cycles 次數會減少，相反的，設定的 k 愈小，表示收斂法則較為嚴格，同時 converge cycles 次數會增加。

log_file_p = true

設定為 true 時，則執行 Autoclass-Search 的過程中，在螢幕上所顯示的訊息都會被寫進 log 檔裡，包含執行的日期、錯誤訊息以及預設的 search params 等等；設定 false 則否。

search_file_p = true

設定為 true 時，則執行 Autoclass-Search 之後會產生“.search”檔案；設定為 false 則否。

results_file_p = true

設定為 true 時，則執行 Autoclass-Search 之後，會產生最佳分類結果的 results 檔案（“.results-bin”或是“.results”）；設定為 false，則不會產生 results 檔案。

n_save = k

k 表示 Autoclass 搜尋完之後儲存最佳 k 筆的結果。k 設定為 0 時，即不產生“.search”、“.results[-bin]”檔案，此參數預設為 2。

checkpoint_p = false

設定為 true 時，則執行 Autoclass-Search 的過程中，每過一段時間，Autoclass 會自動將目前的分類結果做一個備份，產生 chkpt 檔案（“.chkpt-bin”或是“.chkpt”）；設定為 false，則不會有備份動作。每過多少時間做一次備份，由 min_checkpoint_period 參數決定。

min_checkpoint_period = k

表示每過 k 秒後做一次備份，並且在螢幕上會輸出“.”作為提示訊息，原本 k 預設為 10800 秒。

reconverge_type = "chkpt"

引號內可以設定 chkpt 或 results。當設定為 chkpt 且 checkpoint_p = true 時，Autoclass 會繼續執行之前未完成分類的檔案；當設定為 results 且 force_new_search_p = false 時，Autoclass 會將已經完成分類的檔案繼續分類。附帶一提，我們是可以更改 try_fn_type 的設定。

save_compact_p = true

設定為 true 時，則執行 Autoclass-Search 之後會產生 binary 形式的 results 及 chkpt 檔案（.results-bin & .chkpt-bin）；設定為 false，則是 ASCII 形式（.results & .chkpt）。

read_compact_p = true

設定為 true 時，Autoclass 讀取 binary 形式的 results 及 chkpt 檔案（.results-bin & .chkpt-bin）；設定為 false，則是 ASCII 形式（.results & .chkpt）。

min_save_period = k

預設為 1800 秒。此參數設定 k，表示每 k 秒儲存目前搜尋的結果。

n_final_summary = k

此參數預設為 10。k 表示 Autoclass 在搜尋完之後，顯示前 k 個最佳結果。

initial_cycles_p = true

設定為 true 時，Autoclass 會執行在 initialize_parameters 中的 base_cycle；通常在測試時才需設定為 false。

n_data = k

如果 k=0，則 Data file 中的全部資料都被讀取；如果 k>0，則只有 k 筆資料被讀取。此參數通常用於資料很多且使用者只想驗證 Header file、Model file 內容的時候。

max_cycles = k

一個 try 所能達到的最大 converge cycle 個數，也就是 converge cycle 達到 k 時必須終止一個 try，參數預設為 200。

converge_print_p = false

當設定為 true 時，會在螢幕上輸出每一個 cycle 中的資訊，此資訊可以讓使用者修正有關 try_fn_type 的參數預設值。

force_new_search_p = true

設定為 true，則執行 Autoclass-Search 時，Autoclass 會忽略先前已存在的 search 和 results 檔案重新搜尋，且詢問是否刪除這兩個檔案；設定為 false，則會沿用之前的 search 及 results 檔案繼續搜尋。

screen_output_p = true

若設定為 false，則搜尋過程及搜尋結果的任何訊息皆不會在螢幕上輸出。

break_on_warnings_p = true

設定為 true 時，若搜尋發生錯誤，則螢幕會輸出使用者是否繼續的訊息。設定為 false 時，Autoclass 會忽略錯誤且繼續搜尋，直到搜尋結束後才會在螢幕輸出錯誤訊息，同時此訊息也會記錄在 log file 中。

free_storage_p = true

此參數設定為 true 時，表示 Autoclass 在每次搜尋完之後，會把搜尋過程中的暫存檔案刪除；設定為 false 時則否。

max_n_store = k

此參數設定 k，表示記憶體最多能儲存 k 個分類結果。參數預設為 10。

最後，儲存副檔名為 s-params 即完成 Search params file 的準備工作。

附註 6

控制 Autoclass-Search 何時停止搜尋的參數有 interactive_p、max_duration、max_n_tries 三種，此三個參數必須設定其中一個，否則 Autoclass-Search 會永無止盡的持續搜尋。我們知道一個 try 含有隨機的成份，搜尋時間或次數越多，就可能找到更佳的分類結果，不過這總是得花費大量的時間，因此 Autoclass 提供使用者一個搜尋停止的參考點，標記為“dup”。我們用以下的例子說明。

欲分類群數	try_fn_type	cycles 次數	應該分類群數	分類結果	第 n 個 try
[j_in=5]	cs-3	22	5	4	1
[j_in=10]	cs-3	41	10	7	2
[j_in=15]	cs-3	28	15	9	3
[j_in=20]	cs-3	52	20	10	4
[j_in=40]	cs-3	55	40	19	5
[j_in=43]	cs-3	51	43	18	6
[j_in=33]	cs-3	40	33	17	7
[j_in=13]	cs-3	31	13	9	8
[j_in=11]	cs-3	78	11	8	9
[j_in=12]	cs-3	31	12	9	10
[j_in=11]	cs-3	78	dup11	8	11
[j_in=12]	cs-3	31	dup12	9	12
[j_in=11]	cs-3	78	dup11	8	13
[j_in=12]	cs-3	31	dup12	9	14
[j_in=11]	cs-3	78	dup11	8	15
[j_in=12]	cs-3	31	dup12	9	16
[j_in=11]	cs-3	78	dup11	8	17

[j_in=12]	cs-3	40	12	10	18
[j_in=23]	cs-3	69	23	15	19
[j_in=11]	cs-3	39	11	9	20
[j_in=11]	cs-3	39	dup11	9	21
[j_in=11]	cs-3	39	dup11	9	22
[j_in=11]	cs-3	39	dup11	9	23
[j_in=11]	cs-3	39	dup11	9	24
[j_in=11]	cs-3	39	dup11	9	25
[j_in=11]	cs-3	39	dup11	9	26
[j_in=11]	cs-3	39	dup11	9	27

----- SUMMARY OF 10 BEST RESULTS -----					
PROBABILITY	exp(-1946.517)	N_CLASSES	4 FOUND ON TRY	1 *SAVED*	
PROBABILITY	exp(-1955.692)	N_CLASSES	7 FOUND ON TRY	2 *SAVED*	
PROBABILITY	exp(-1961.277)	N_CLASSES	8 FOUND ON TRY	9 DUPS	4
PROBABILITY	exp(-1967.216)	N_CLASSES	10 FOUND ON TRY	4	
PROBABILITY	exp(-1969.740)	N_CLASSES	9 FOUND ON TRY	8	
PROBABILITY	exp(-1969.765)	N_CLASSES	9 FOUND ON TRY	3	
PROBABILITY	exp(-1969.832)	N_CLASSES	9 FOUND ON TRY	10 DUPS	3
PROBABILITY	exp(-1971.207)	N_CLASSES	9 FOUND ON TRY	20 DUPS	10
PROBABILITY	exp(-1972.801)	N_CLASSES	10 FOUND ON TRY	18	
PROBABILITY	exp(-2000.396)	N_CLASSES	15 FOUND ON TRY	19	

對照上面兩張圖表，使用者可以清楚知道，當不同 **try** 的欲分類群數及分類結果相同時，則會被標記成 **dup**。同時在最佳分類中，可能重複顯示相同群數的分類結果，其差別在於欲分類群數的不同。

3.2.1.5 執行 Search 指令

準備好 Data file、Header file、Model file 和 Search params file 四個檔案之後，直接在;autoclass-c;目錄下輸入：

```
./autoclass -search <.db2 file path> <.hd2 file path> <.model file path> <.s-params file path>
```

則 Autoclass 便會開始進行搜尋，最後產生 search 及 result 檔。

3.2.2 報告

在進行 Report 指令前，使用者必須先準備好三個檔案：search file、result file 和 report params file，其中 search file 及 result file 會在執行 Search 指令之後自動產生，因此使用者只需了解 report params file 的設定。

3.2.2.1 Report params 檔案

這個檔案主要是有關 Autoclass-Reports 的參數設定，使用者可以根據自己的需求對此部份進行設定，底下我們敘述每個參數的功用。

n_clsfs = k

此參數設定為 k 時，表示執行 Report 指令，Autoclass 會對 result file 中的前 k 個最佳分類結果，產生 report file，其檔案名稱依序為 ".*-*-1" 至 ".*-*-k"。參數預設為 1。

clsf_n_list = k

此參數設定為 k 時，表示執行 Report 指令，Autoclass 會對 result file 中的第 k 個最佳分類結果產生 report file。

report_type = "all"

此參數可以設定為 all、influence_values、xref_case 或是 xref_class。設定 influence_values 表示執行 Report 指令會產生 influence file；設定 xref_case 表示執行 Report 指令會產生 case file；設定 xref_class 表示執行 Report 指令會產生 class file；設定 all 表示執行 Report 指令會產生 case、class、influence file。

report_mode = "text"

此參數可以設定為 text 或是 data。設定 text 表示產生的 report file 皆為文字檔，同時產生的檔名為".*-text-n"；設定 data 表示產生的 report file 在內容上包含更多資訊，同時產生的檔名為".*-data-n"。

comment_data_headers_p = false

此參數是當 report_mode 設定為 data 時所做的進一步設定。設定 true 表示執行 Report 指令所產生的 report file，其開頭敘述的部份會被標上註解符號。設定 false 則否。

order_attributes_by_influence_p = true

此參數設定為 true 時，在 influence file 中，屬性的排列順序是根據影響力的大小決定，同時產生的檔案名稱為 ".influ-o-text-n"；若設定為 false，屬性的排列順序則是按照屬性編號決定，同時產生的檔案名稱為 ".influ-no-text-n"

num_atts_to_list = k

設定 k 表示執行 Report 指令所產生的 influence file，其內容會顯示每一個 try 中的前 k 個屬性資訊，例如標準差。若不做設定則列出所有屬性資訊。

xref_class_report_att_list = k

此參數輸入的數字代表屬性編號。若使用者希望 class file 中顯示屬性編號為 k 和 m 的屬性值，則可以輸入 k, m，依此類推。若不做任何輸入，則不會顯示屬性值。

break_on_warnings_p = true

此參數設定為 true 時，若執行 Report 指令發生錯誤，則會在螢幕輸出是否繼續執行的訊息；設定為 false 時，則 Autoclass 會忽略錯誤訊息而繼續執行。

free_storage_p = true

此參數設定為 true 時，表示 Autoclass 在每次報告完之後，會把報告過程中的暫存檔案刪除；設定為 false 則否。

max_num_xref_class_probs = k

設定 k 表示在 class file 及 case file 中，每筆資料最多只顯示 k 個屬於某一群的機率，不過此參數設定不影響".case-text-n"檔案，其內容只顯示最大的分屬某群之機率值。

sigma_contours_att_list = k

此參數輸入的數字代表屬性編號。若使用者希望 influence file 中顯示屬性編號為 k 和 m 的 sigma class contour values，則先設定 report_mode = "data"，之後在此參數輸入 k, m 即可，依此類推。若不做任何輸入，則不會顯示 sigma class contour values。

最後，儲存副檔名為 r-params 即完成 Report params file 的準備工作。

3.2.2.2 執行 Report 指令

準備好 result file、search file 和 Report params file 三個檔案之後，直接在 ;autoclass-c;目錄下輸入：

```
./autoclass -reports <.result-bin file path> <.search file path> <.r-params file path>
```

則 Autoclass 便會開始進行 report，之後根據 Report params file 的設定產生 report 檔案。

3.2.2.3 influence 檔案、case 檔案和 class 檔案

在標準設定下，執行 Autoclass-Report 會產生三個檔案，其內容分述如下：

- Influence file：顯示搜尋的結果、每筆資料對於每一群的貢獻，以及屬性造成分類結果的影響力。
- case file：列出每筆資料最大的分屬某群之機率值。
- class file：以每一群為單位，列出其包含的所有資料。

第 4 章 Autoclass 測試

本章節的目的是要測試 Autoclass 在何種參數設定下，對相同的資料進行搜尋會有不變的分類結果，以及 Autoclass 是否能處理較簡單的分類問題。

4.1 參數的測試

本論文以 100 筆資料作為測試，前 50 筆為 (0, 0)，後 50 筆為 (1, 1)，見圖二。在參數設定相同下，其中 `start_fn_type = "random"`、`randomize_random_p = true`，Autoclass 對此 Data file 重複 5 次搜尋，卻產生不同的最佳分類結果（附錄一）。在參數設定相同下，其中 `start_fn_type = "random"`、`randomize_random_p = false`，Autoclass 對此 Data file 重複 5 次搜尋，亦產生不同的最佳分類結果（附錄二）。在參數設定相同下，其中 `start_fn_type = "block"`、`randomize_random_p = true`，Autoclass 對此 Data file 重複 5 次搜尋，亦產生不同的最佳分類結果（附錄三）。在參數設定相同下，其中 `start_fn_type = "block"`、`randomize_random_p = false`，Autoclass 對此 Data file 重複 5 次搜尋，可產生相同的最佳分類結果（附錄四）。

故相同的 Data file 只有當 `start_fn_type = "block"`、`randomize_random_p = false` 的參數設定下，才會產生相同的分類結果。否則，不僅最佳的分類結果會改變，連分成 N 群的概率值也可能有所不同。進一步測試，在參數設定相同下，其中 `start_fn_type = "block"`、`randomize_random_p = false`，若將資料輸入 Data file 的順序改變，則會產生不同的分類結果（附錄五）。

由以上測試結果，我們得知 Autoclass 無法對同一組資料產生相同的搜尋結果，即使我們可以控制 `start_fn_type` 與 `randomize_random_p` 參數，使相同輸入的 Data file 產生相同的分類結果，但此做法的缺點在於資料輸入的情況較好時，會產生不錯的分群結果；資料的輸入較不理想時，即會產生不好的分群結果。

4.2 測試最佳分類的準確性

由於我們無法要求 Autoclass 對同一組資料產生相同的搜尋結果，所以本論文就針對 `start_fn_type = "random"`、`randomize_random_p = true` 的設定下，討論最佳分類的準確性。

我們以機率值大小作為標準，若某一筆資料屬於某一群的機率值最大，則此筆資料就分到該群，因此，將 4.1 節所提的前五次最佳分類畫成圖形，發現 Autoclass 確實將 50 筆 (0, 0) 與 50 筆 (1, 1) 分開。五張圖形皆相同，見圖三。

進一步測試，我們將原 100 筆的資料加上 0 到 1 之間的雜訊（見圖四），讓 Autoclass 對此 Data file 重複多次搜尋，其分類結果參見附錄六。同樣的，我們以機率值大小作為標準，若某一筆資料屬於某一群的機率值最大，則此筆資料就分到該群。因此，將附錄六的分類結果畫成圖形，其圖形為圖五、圖六及圖七。

總結上述結果，我們將加雜訊後的 100 筆資料進行三十次搜尋，只有第二十六次及第二十九次產生三群的分類結果（見圖六、圖七），其餘的結果皆與圖五相同。我們觀看圖六與圖七的分類圖形，可以了解 Autoclass 將資料分成三類的結果並非全無道理，但是這只是少數情況，其餘二十八次的最佳分類結果都是相同的，若我們以投票的方式來將資料分類，統計分類結果證實 Autoclass 確實能將此資料正確分開。

第 5 章 Autoclass 實驗

根據第四章的測試結果，Autoclass 能處理較簡易的分類問題，因此本章中包含五個較困難的實驗，每個實驗中的動作電位皆為人工動作電位，雜訊都是從同一個真實訊號中取出，其最大標準差約為 $12 \mu V$ 。本論文採用的訊號並非真實訊號，而是由人工動作電位加上真實雜訊所產生。採用人工訊號的原因有二：一是人工訊號較容易控制；二是能得知正確的分類結果。

5.1 實驗一

圖八為兩個人工動作電位，我們將每個動作電位都加上一千筆真實雜訊，所以總共得到兩千個動作電位（見圖九），我們將以 Autoclass 處理圖九的兩千筆人工訊號。圖十為圖九的兩千筆動作電位經 PCA 轉換後，以第一及第二主成分所繪製的資料分布圖，同顏色的點屬於同一群。

5.1.1 測試一

實驗中的每筆訊號皆由 32 維的向量表示，所以本論文將每一維視為一種屬性輸入 Data file，而 Header file 的 error 設定為兩兩數據間最小差距的一半，每一屬性皆選擇 Single Normal CN Model，且 try 的次數設定為 100。在此條件設定下以 Autoclass 進行多次搜尋，每次最佳分類結果大致分成 30 群左右（附錄七），由於正確的群數只有兩群，所以我們認為 Autoclass 直接搜尋沒有經過處理的資料，其結果並不理想，因此我們試想透過 PCA 的概念，是否可以得到較佳的分類結果。

5.1.2 測試二

資料經由 PCA 轉換，每一筆訊號都得到 32 個主成分，在相同的設定之下，我們將 32 個主成分都輸入 Data file，進行多次搜尋之後，每次最佳分類結果都是四群（附錄八）。由此得知資料先經過 PCA 轉換再以 Autoclass 搜尋，分類結果會

有所改善。進一步，我們選取某次分類的最佳結果作為觀察（圖十一），探究此分類結果的可信度。由圖中可以發現四群中有兩群比較密集（藍色與綠色），檢驗後發現藍色點在原始訊號中確實屬於同一群，綠色點亦是如此。不過這樣的結果我們並不滿意，其原因在於藍綠兩群只包含了 809 個點，換句話說，假設將另外兩群視為未分類群，則未分類的點接近 60%。另一方面，分類結果的群數應至多為三群，不應把未分類群再分成兩群。我們認為造成此分類結果的原因有二，一是 Autoclass 認為每個屬性重要性是相同的，也就是 Autoclass 不知道前幾個主成分具有較大的影響力，若以較不具影響力的主成分分類，則會降低正確分類的點數（附錄九）；二是我們提供資料的維度愈高，分類的群數也隨之提高。

5.1.3 測試三

在此小節中，我們將測試主成分個數對於分類正確率的影響，並試著處理未分類點的問題。本論文提到的正確率是指分類正確的點在能分類的點中所佔的比率，而未分類率是指未分類的點在全部點中所佔的比率。首先我們只取第一個主成分來分類，在五次的測試中有一群、兩群或三群的分類結果，而平均正確率為 62%、未分類率平均為 6%；以前兩個主成分來分類，分類結果皆為三群，平均正確率為 91%、未分類率平均為 31%；以前三個主成分來分類，分類結果也皆為三群，平均正確率為 96%、未分類率平均為 43%；以前四個主成分來分類，分類結果也皆為三群，平均正確率為 99%、未分類率平均為 53%；以前五個主成分來分類，分類結果皆為三群，平均正確率為 99.9%、未分類率平均為 56%，其平均正確率之高使得我們不考慮繼續增加主成分的個數進行測試（見圖十二）。

此實驗結果顯示，前五個主成分的分類正確率幾乎與 32 個主成分的分類正確率一樣高，更佳的是除了未分類率稍微降低之外，分群數只有三群，所以我們認為以前五個主成分處理分類問題會比用 32 個主成分好。因此，我們假設：處理兩群的分類問題時，將資料經過 PCA 轉換後，Autoclass 只要以前五個主成分分類，即能得到具有可信度相當高的分類結果。更進一步驗證詳見於後續的實驗。

在此我們建議一種處理未分類點的方法，此方法適用於分類正確率 90% 以上且未分類率 60% 以下的結果。討論分類結果為三群的情況，我們判定有兩群為已分類群，同時將兩群對應的訊號取平均，即可得到兩個平均波形（**template**）。接著利用這兩個平均波形將其他未分類的訊號分類。我們實際以前五個主成分所得的分類結果測試，發現完全分類後的正確率可達 93% 之高，並且圖十三的資料分布圖顯示，此做法能準確且合理地處理兩群交界的點，比起我們直接以眼睛觀察而取一條分隔線來的好。因此，我們認為上述之方法除了可以解決未分類點的問題，同時仍保持相當高的正確率。

5.2 實驗二

此小節將以更難分類的資料驗證實驗一的假設。圖十四為兩個人工動作電位，我們將每個動作電位都加上一千筆真實雜訊，共得兩千個動作電位（見圖十五），圖十六為圖十五的兩千筆動作電位經 PCA 轉換後，以第一及第二主成分所繪製的資料分布圖，同顏色的點屬於同一群。

此實驗的參數設定與實驗一相同。我們分別以前兩個、前三個、前四個及前五個主成分來分類，最佳的分類結果皆為兩群（附錄十），正確率平均約為 51%，且資料分布類似於圖十七。

由此得知在此實驗中，**Autoclass** 的最佳分類結果並不盡理想，但是觀察每次的 **Report** 檔案（附錄十），報告內容都包含三群的分類結果，我們將最佳的三群分類結果繪製成圖，見圖十八。

從圖十八中我們發現前兩個主成分與前四個主成分的分類結果圖是相似的；前三個主成分與前五個主成分的分類結果圖亦是相似的。進一步計算，前三個主成分與前五個主成分的分類正確率平均可達 88% 之高，而前兩個主成分與前四個主成分的分類正確率卻只有 51% 左右（圖十九）。

藉由這個較難以分類的實驗，我們得知 **Autoclass** 仍然只需考慮前五個主成分，即可得到高正確率的分類結果，同時此實驗顯示 **Autoclass** 進行一次搜尋所

產生的分類群數，皆有可能是實際的分類群數，所以我們不應直接以最佳分類的群數作為真實的分類群數，而是需對所有的分類群數做進一步的探討。

5.3 實驗三

由 5.2 小節的實驗結果得知，以前三個主成分和前五個主成分處理兩類的分類問題，會產生較佳的分類結果。在此小節中，我們將實驗二中 spike4 的最低點位置慢慢靠近 spike3 的最低點位置，其餘位置上的點不動，藉此增加分類的困難度來觀察前三個主成分和前五個主成分分類正確率的變化情形。

以下所提的 pick_n_std 定義為兩動作電位在最低點位置上的雜訊標準差，pick_dn_rat 定義為最低點位置的差距除以 pick_n_std。實驗二採用的真實雜訊 pick_n_std 是 $11.95 \mu V$ 。在較好的分類結果下，當 spike3 與 spike4 的 pick_dn_rat 約為 1.67 時，前三個主成分所得的分類正確率為 84.34%，前五個主成分所得的分類正確率為 87.78%；pick_dn_rat 約為 1.5 時，前三個主成分所得的分類正確率為 84.16%，前五個主成分所得的分類正確率為 86.22%；pick_dn_rat 約為 1.35 時，前三個主成分所得的分類正確率為 83.69%，前五個主成分所得的分類正確率為 87.24%；pick_dn_rat 約為 1.25 時，前三個主成分所得的分類正確率為 83.56%，前五個主成分所得的分類正確率為 86.45%；pick_dn_rat 約為 1.15 時，前三個主成分所得的分類正確率為 82.52%，前五個主成分所得的分類正確率為 86.46%；pick_dn_rat 約為 1 時，前三個主成分所得的分類正確率為 81.24%，前五個主成分所得的分類正確率為 84.63%；pick_dn_rat 約為 0.5 時，前三個主成分所得的分類正確率為 79%，前五個主成分所得的分類正確率為 82.54%；當最低點重疊時，前三個主成分所得的分類正確率為 78.94%，前五個主成分所得的分類正確率為 79.35%。我們將上述結果繪製成圖二十。

我們先說明圖二十中實線與虛線的差別。一組資料經由 Autoclass 分類之後，其分類結果的群數大多會重複，進一步觀察 influence 檔案可以發現，雖然群數相同，但影響分類結果的主成分重要性順序可能不完全相

同。圖二十中的實線是分類結果受到第一主成分影響最大，第二主成分次之，依序第三、第四、第五主成分所得的分類正確率，虛線則是主成分重要性任意排列所得的分類正確率。觀察圖二十，我們可以得知三件事：

1. 比較同色的實線與虛線，可以發現實線的正确率都比虛線高。而我們又知道 PCA 轉換後第一主成分至第五主成分的重要性是由大到小，就理論或實際而言，我們可以提出合理的結論：以前 N 個主成分個數分類時，選擇受到第一主成分影響最大，第二主成分次之，依序至第 N 個主成分所得的分類結果，其正確率會較同群數的分類結果來的高。
2. 觀看 5pca_ord 與 3pca_ord，可以得知前五個主成分所得的分類正確率較前三個主成分高，換句話說，若前五個主成分與前三個主成分皆能處理的分類問題，只需考慮前五個主成分的分類結果。
3. 圖二十中的實線皆呈現平穩的型態，顯示出只縮小兩動作電位最低點的距離，對分類結果的影響並不大，我們應考慮兩動作電位最大差距與雜訊最大標準差的關係。

5.4 實驗四

在此小節中，我們以圖二十一所造出的兩千筆訊號（圖二十二）進行實驗，圖二十三為圖二十二的兩千筆動作電位經 PCA 轉換後，以第一及第二主成分所繪製的資料分布圖，同顏色的點屬於同一群。觀察圖二十三的點圖，若不考慮顏色，我們似乎可以認為它們只有一群。事實上，圖二十一的兩筆原始訊號，由外觀判斷幾乎可以視為同一個神經訊號，底下我們將測試 Autoclass 是否仍能以前五個主成分將此種訊號確實分開。

此實驗的參數設定與前三個實驗相同。我們以前五個主成分來分類，產生的分類結果只有兩群和三群（附錄十一），其資料分布圖大致如圖二十四及圖二十五，平均正確率皆約為 51%。

明顯的，Autoclass 分類的結果並不是我們所期望的，這也說明了 5.13 小節中所提的假設只適用於某些條件限制下的分類問題。因此在後續實驗中，我們試著找出 Autoclass 利用前五個主成分處理分類問題時，其條件限制為何。

5.5 實驗五

此小節中我們探討兩動作電位最大差距與雜訊最大標準差的關係，以下所提的 m_dn_rat 定義為兩動作電位最大的差距除以雜訊最大的標準差。本實驗真實雜訊最大標準差為 $11.95 \mu V$ 。圖二十六中 spike7 與 spike3 的 m_dn_rat 為 0.85，實際上若只看齒狀的部份，spike7 是 spike3 的向上平移。圖二十七中 spike8 與 spike3 的 m_dn_rat 為 1.5。我們考慮平移過程中 m_dn_rat 分別為 0.85、1、1.05、1.1、1.15、1.2、1.25、1.3、1.35、1.4、1.45、1.5 時，前五個主成分所得的分類正確率變化情形（圖二十八）。

從圖二十八中我們可以發現 m_dn_rat 高於 1.3 時，其分類正確率皆大於 80%， m_dn_rat 低於 1.25 時，其分類正確率小於 60%，因此我們得到一個結論：Autoclass 以前五個主成分處理 m_dn_rat 低於 1.3 的分類問題時，其分類結果是不可靠的。進一步推廣此結論，由於我們知道在雜訊為正規分布的情況下，原始動作電位加上雜訊之後，其分布範圍幾乎都在原始動作電位的正負三倍雜訊最大標準差內。故上述結論提供我們一個處理分類問題的先驗標準：假設雜訊為正規分布，每一筆訊號皆由 N 個維度構成，我們計算每一維度訊號間的兩兩最大差距，若得到的 N 個距離之最大值小於等於 7.3 倍的雜訊最大標準差，則我們不建議以 Autoclass 處理此分類問題。同時此結論也可推得一個後驗方法：分類後的每一群可個別求得一個平均波形，若兩波形的 m_dn_rat 低於 1.3，則此分類結果的可信度可能不高。

第 6 章 結論與建議

資料分類一直是普遍存在的問題，在本論文中我們將 Autoclass 分類軟體應用於生物訊號的分類上。本章節將歸納出「Autoclass 的優缺點」與「Autoclass 處理分類問題步驟之建議」。

6.1 Autoclass 的優缺點

Autoclass 是一種 unsupervised 的分類法，我們認為使用上其優缺點如下：

優點：

1. 可以處理連續型及離散型的資料，即可以處理大多數的分類問題。
2. 當輸入的資料不完整時，仍可將此筆資料分類。
3. 可以處理多群的分類問題。

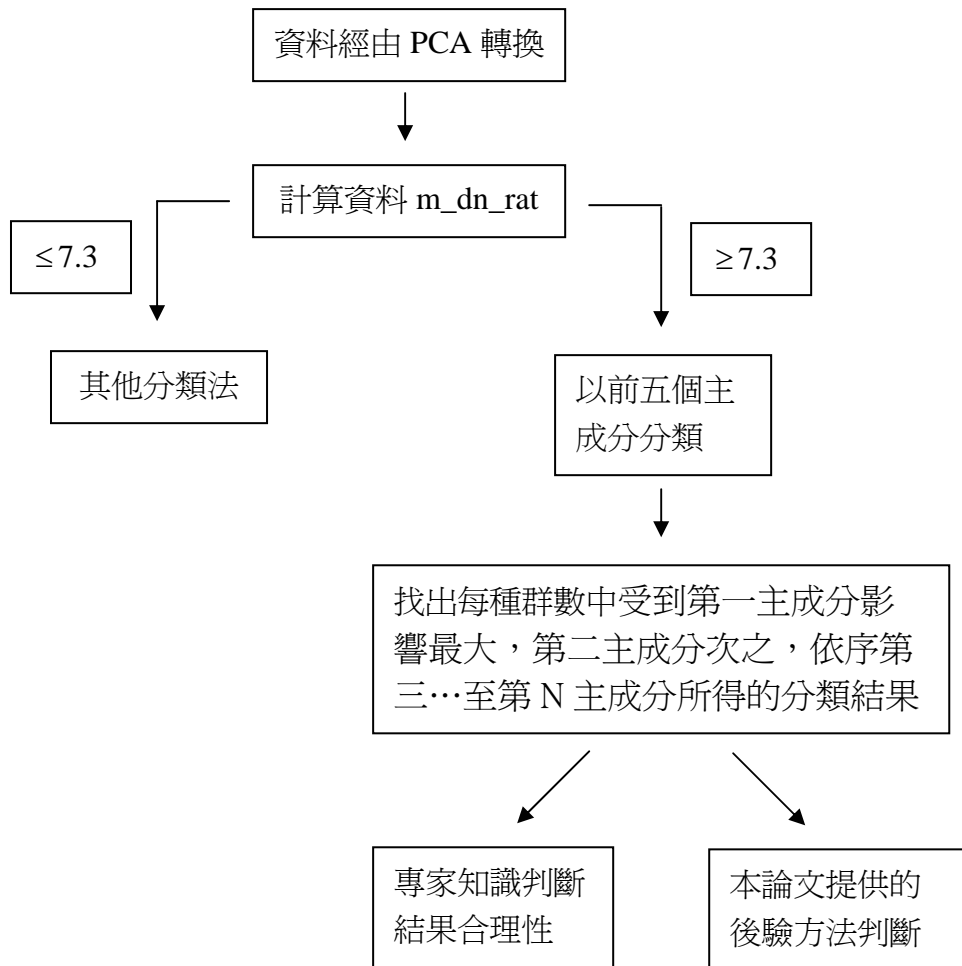
缺點：

1. 不是一個完全自動化的分類系統，使用者仍需主觀地決定資料適當的群數範圍，而此問題卻是分類上的一大難題。
2. 在使用上需反覆處理資料，因而花費大量的時間。
3. 沒有提供一個先驗標準來預測一組資料是否應該被分類。
4. 沒有提供一個後驗方法來評估分類的結果是否可以信賴。

Bayesian classification (autoclass) [2] 中描述：使用 Autoclass 處理分類問題，必須不斷的重複假設與測試，並結合專家知識與程式，才能得到良好的結果。由於缺點 1 是牽涉專家知識的層面，缺點 2 是屬於不可避免的時間花費，因此本論文對此不加以探討，主要著重於缺點 3 與缺點 4 的改進。本論文在 5.5 小節中提出的先驗與後驗判別法，即為改進缺點 3 和缺點 4 之方法，希望能提供未來使用 Autoclass 的使用者參考。

6.2 Autoclass 處理分類問題步驟之建議

本論文提供使用者一個處理分類問題的步驟如下：



參考書目

- [1] 李易霖 (2004) 。《神經細胞訊號的一種分類法》。國立中央大學數學系碩士論文，6-9。
- [2] 1996 Bayesian classification (autoclass) . theory and results *Advances in Knowledge Discovery and Data Mining* ed U M Fayyad *et al* (Menlo Park, CA: AAAI Press) pp 153-80.
- [3] Michael S Lewicki (1998) . *A review of methods for spike sorting: the detection and classification of neural action potentials*. pp R56.

附錄

附錄一

第一次執行分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-625.474) N_CLASSES	2 FOUND ON TRY	1 *SAVED*	
PROBABILITY	exp(-647.587) N_CLASSES	3 FOUND ON TRY	2 DUPS	1 *SAVED*
PROBABILITY	exp(-647.587) N_CLASSES	3 FOUND ON TRY	4 DUPS	12 *SAVED*
PROBABILITY	exp(-668.103) N_CLASSES	4 FOUND ON TRY	5 DUPS	9 *SAVED*
PROBABILITY	exp(-668.103) N_CLASSES	4 FOUND ON TRY	10 *SAVED*	
PROBABILITY	exp(-687.527) N_CLASSES	5 FOUND ON TRY	3 DUPS	1 *SAVED*

第二次執行分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-625.474) N_CLASSES	2 FOUND ON TRY	1 DUPS	1 *SAVED*
PROBABILITY	exp(-647.587) N_CLASSES	3 FOUND ON TRY	3 DUPS	1 *SAVED*
PROBABILITY	exp(-668.103) N_CLASSES	4 FOUND ON TRY	4 DUPS	1 *SAVED*
PROBABILITY	exp(-687.527) N_CLASSES	5 FOUND ON TRY	11 DUPS	9 *SAVED*
PROBABILITY	exp(-708.202) N_CLASSES	6 FOUND ON TRY	5 *SAVED*	
PROBABILITY	exp(-746.077) N_CLASSES	8 FOUND ON TRY	6 DUPS	12 *SAVED*

第三次執行分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-625.474) N_CLASSES	2 FOUND ON TRY	1 DUPS	1 *SAVED*
PROBABILITY	exp(-647.583) N_CLASSES	3 FOUND ON TRY	8 DUPS	1 *SAVED*
PROBABILITY	exp(-647.583) N_CLASSES	3 FOUND ON TRY	2 DUPS	2 *SAVED*
PROBABILITY	exp(-668.102) N_CLASSES	4 FOUND ON TRY	4 *SAVED*	
PROBABILITY	exp(-668.102) N_CLASSES	4 FOUND ON TRY	15 *SAVED*	
PROBABILITY	exp(-687.526) N_CLASSES	5 FOUND ON TRY	9 DUPS	2 *SAVED*

第四次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-625.474) N_CLASSES	2 FOUND ON TRY	1 DUPS	4 *SAVED*
PROBABILITY	exp(-647.587) N_CLASSES	3 FOUND ON TRY	2 DUPS	3 *SAVED*
PROBABILITY	exp(-669.097) N_CLASSES	4 FOUND ON TRY	5 DUPS	9 *SAVED*
PROBABILITY	exp(-708.850) N_CLASSES	6 FOUND ON TRY	6 DUPS	1 *SAVED*
PROBABILITY	exp(-727.594) N_CLASSES	7 FOUND ON TRY	7 DUPS	8 *SAVED*

第五次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-625.474)	N_CLASSES	2 FOUND ON TRY	1 DUPS 1 *SAVED*
PROBABILITY	exp(-647.583)	N_CLASSES	3 FOUND ON TRY	3 DUPS 1 *SAVED*
PROBABILITY	exp(-647.587)	N_CLASSES	3 FOUND ON TRY	10 DUPS 5 *SAVED*
PROBABILITY	exp(-669.097)	N_CLASSES	4 FOUND ON TRY	6 DUPS 1 *SAVED*
PROBABILITY	exp(-687.526)	N_CLASSES	5 FOUND ON TRY	4 *SAVED*
PROBABILITY	exp(-706.097)	N_CLASSES	6 FOUND ON TRY	8 *SAVED*

第二十二次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-647.587)	N_CLASSES	3 FOUND ON TRY	2 DUPS 2 *SAVED*
PROBABILITY	exp(-668.103)	N_CLASSES	4 FOUND ON TRY	6 *SAVED*
PROBABILITY	exp(-669.093)	N_CLASSES	4 FOUND ON TRY	4 DUPS 23 *SAVED*
PROBABILITY	exp(-726.453)	N_CLASSES	7 FOUND ON TRY	7 *SAVED*
PROBABILITY	exp(-2010.079)	N_CLASSES	2 FOUND ON TRY	1 *SAVED*

附錄二

第一次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-625.474)	N_CLASSES	2 FOUND ON TRY	1 DUPS 1 *SAVED*
PROBABILITY	exp(-647.583)	N_CLASSES	3 FOUND ON TRY	2 DUPS 1 *SAVED*
PROBABILITY	exp(-647.587)	N_CLASSES	3 FOUND ON TRY	22 *SAVED*
PROBABILITY	exp(-669.089)	N_CLASSES	4 FOUND ON TRY	12 *SAVED*
PROBABILITY	exp(-689.149)	N_CLASSES	5 FOUND ON TRY	3 *SAVED*
PROBABILITY	exp(-689.154)	N_CLASSES	5 FOUND ON TRY	7 DUPS 1 *SAVED*
PROBABILITY	exp(-708.202)	N_CLASSES	6 FOUND ON TRY	23 DUPS 7 *SAVED*

第二次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-625.474)	N_CLASSES	2 FOUND ON TRY	1 DUPS 6 *SAVED*
PROBABILITY	exp(-647.583)	N_CLASSES	3 FOUND ON TRY	2 *SAVED*
PROBABILITY	exp(-668.103)	N_CLASSES	4 FOUND ON TRY	5 DUPS 1 *SAVED*
PROBABILITY	exp(-687.527)	N_CLASSES	5 FOUND ON TRY	6 DUPS 14 *SAVED*
PROBABILITY	exp(-708.202)	N_CLASSES	6 FOUND ON TRY	7 *SAVED*
PROBABILITY	exp(-726.453)	N_CLASSES	7 FOUND ON TRY	14 DUPS 3 *SAVED*

第三次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-625.474)	N_CLASSES	2 FOUND ON TRY	1 DUPS 5 *SAVED*
PROBABILITY	exp(-647.583)	N_CLASSES	3 FOUND ON TRY	2 DUPS 7 *SAVED*
PROBABILITY	exp(-669.093)	N_CLASSES	4 FOUND ON TRY	12 *SAVED*
PROBABILITY	exp(-669.097)	N_CLASSES	4 FOUND ON TRY	7 DUPS 1 *SAVED*
PROBABILITY	exp(-687.527)	N_CLASSES	5 FOUND ON TRY	16 DUPS 4 *SAVED*
PROBABILITY	exp(-689.150)	N_CLASSES	5 FOUND ON TRY	3 *SAVED*
PROBABILITY	exp(-706.097)	N_CLASSES	6 FOUND ON TRY	5 DUPS 1 *SAVED*

第四次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-647.587)	N_CLASSES	3 FOUND ON TRY	2 DUPS 2 *SAVED*
PROBABILITY	exp(-668.103)	N_CLASSES	4 FOUND ON TRY	4 DUPS 2 *SAVED*
PROBABILITY	exp(-687.527)	N_CLASSES	5 FOUND ON TRY	5 *SAVED*
PROBABILITY	exp(-687.527)	N_CLASSES	5 FOUND ON TRY	7 DUPS 6 *SAVED*
PROBABILITY	exp(-689.150)	N_CLASSES	5 FOUND ON TRY	3 DUPS 2 *SAVED*
PROBABILITY	exp(-706.097)	N_CLASSES	6 FOUND ON TRY	9 *SAVED*
PROBABILITY	exp(-726.453)	N_CLASSES	7 FOUND ON TRY	8 DUPS 1 *SAVED*

第五次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-625.474)	N_CLASSES	2 FOUND ON TRY	1 DUPS 12 *SAVED*
PROBABILITY	exp(-647.587)	N_CLASSES	3 FOUND ON TRY	2 DUPS 3 *SAVED*
PROBABILITY	exp(-726.453)	N_CLASSES	7 FOUND ON TRY	7 *SAVED*
PROBABILITY	exp(-762.930)	N_CLASSES	9 FOUND ON TRY	8 *SAVED*
PROBABILITY	exp(-830.756)	N_CLASSES	13 FOUND ON TRY	10 DUPS 10 *SAVED*

附錄三

第一次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-625.474)	N_CLASSES	2 FOUND ON TRY	1 DUPS 2 *SAVED*
PROBABILITY	exp(-669.097)	N_CLASSES	4 FOUND ON TRY	3 DUPS 2 *SAVED*
PROBABILITY	exp(-708.850)	N_CLASSES	6 FOUND ON TRY	4 DUPS 2 *SAVED*
PROBABILITY	exp(-727.594)	N_CLASSES	7 FOUND ON TRY	16 *SAVED*
PROBABILITY	exp(-746.077)	N_CLASSES	8 FOUND ON TRY	27 *SAVED*
PROBABILITY	exp(-781.392)	N_CLASSES	10 FOUND ON TRY	5 DUPS 4 *SAVED*
PROBABILITY	exp(-798.353)	N_CLASSES	11 FOUND ON TRY	17 DUPS 2 *SAVED*
PROBABILITY	exp(-814.791)	N_CLASSES	12 FOUND ON TRY	28 *SAVED*

第二次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-625.474) N_CLASSES	2 FOUND ON TRY	1 DUPS	1 *SAVED*
PROBABILITY	exp(-669.097) N_CLASSES	4 FOUND ON TRY	3 DUPS	3 *SAVED*
PROBABILITY	exp(-689.154) N_CLASSES	5 FOUND ON TRY	11 DUPS	3 *SAVED*
PROBABILITY	exp(-708.850) N_CLASSES	6 FOUND ON TRY	4 DUPS	4 *SAVED*
PROBABILITY	exp(-727.594) N_CLASSES	7 FOUND ON TRY	8 DUPS	1 *SAVED*
PROBABILITY	exp(-746.077) N_CLASSES	8 FOUND ON TRY	9 DUPS	1 *SAVED*
PROBABILITY	exp(-781.392) N_CLASSES	10 FOUND ON TRY	5 DUPS	3 *SAVED*
PROBABILITY	exp(-798.353) N_CLASSES	11 FOUND ON TRY	17 *SAVED*	

第三次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-625.474) N_CLASSES	2 FOUND ON TRY	1 DUPS	1 *SAVED*
PROBABILITY	exp(-669.097) N_CLASSES	4 FOUND ON TRY	3 DUPS	1 *SAVED*
PROBABILITY	exp(-689.154) N_CLASSES	5 FOUND ON TRY	10 DUPS	3 *SAVED*
PROBABILITY	exp(-708.850) N_CLASSES	6 FOUND ON TRY	4 DUPS	1 *SAVED*
PROBABILITY	exp(-727.594) N_CLASSES	7 FOUND ON TRY	15 DUPS	2 *SAVED*
PROBABILITY	exp(-746.077) N_CLASSES	8 FOUND ON TRY	13 *SAVED*	
PROBABILITY	exp(-781.392) N_CLASSES	10 FOUND ON TRY	5 DUPS	3 *SAVED*
PROBABILITY	exp(-847.259) N_CLASSES	14 FOUND ON TRY	6 DUPS	2 *SAVED*

第四次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-625.474) N_CLASSES	2 FOUND ON TRY	1 DUPS	4 *SAVED*
PROBABILITY	exp(-669.097) N_CLASSES	4 FOUND ON TRY	3 DUPS	5 *SAVED*
PROBABILITY	exp(-689.154) N_CLASSES	5 FOUND ON TRY	9 DUPS	2 *SAVED*
PROBABILITY	exp(-708.850) N_CLASSES	6 FOUND ON TRY	4 DUPS	1 *SAVED*
PROBABILITY	exp(-727.594) N_CLASSES	7 FOUND ON TRY	13 DUPS	2 *SAVED*
PROBABILITY	exp(-746.077) N_CLASSES	8 FOUND ON TRY	15 *SAVED*	
PROBABILITY	exp(-781.392) N_CLASSES	10 FOUND ON TRY	5 DUPS	2 *SAVED*
PROBABILITY	exp(-798.353) N_CLASSES	11 FOUND ON TRY	18 *SAVED*	

第五次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-625.474) N_CLASSES	2 FOUND ON TRY	1 DUPS	1 *SAVED*
PROBABILITY	exp(-669.097) N_CLASSES	4 FOUND ON TRY	3 DUPS	1 *SAVED*
PROBABILITY	exp(-689.154) N_CLASSES	5 FOUND ON TRY	12 DUPS	4 *SAVED*
PROBABILITY	exp(-708.850) N_CLASSES	6 FOUND ON TRY	4 DUPS	1 *SAVED*

PROBABILITY	exp(-727.594)	N_CLASSES	7 FOUND ON TRY	16 *SAVED*
PROBABILITY	exp(-746.077)	N_CLASSES	8 FOUND ON TRY	28 *SAVED*
PROBABILITY	exp(-781.392)	N_CLASSES	10 FOUND ON TRY	5 DUPS 7 *SAVED*
PROBABILITY	exp(-814.791)	N_CLASSES	12 FOUND ON TRY	15 *SAVED*

附錄四

第一次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-625.474)	N_CLASSES	2 FOUND ON TRY	1 DUPS 2 *SAVED*
PROBABILITY	exp(-669.097)	N_CLASSES	4 FOUND ON TRY	3 DUPS 2 *SAVED*
PROBABILITY	exp(-708.850)	N_CLASSES	6 FOUND ON TRY	4 DUPS 2 *SAVED*
PROBABILITY	exp(-727.594)	N_CLASSES	7 FOUND ON TRY	26 DUPS 1 *SAVED*
PROBABILITY	exp(-746.077)	N_CLASSES	8 FOUND ON TRY	11 DUPS 3 *SAVED*
PROBABILITY	exp(-781.392)	N_CLASSES	10 FOUND ON TRY	5 DUPS 3 *SAVED*
PROBABILITY	exp(-798.353)	N_CLASSES	11 FOUND ON TRY	9 DUPS 2 *SAVED*
PROBABILITY	exp(-814.791)	N_CLASSES	12 FOUND ON TRY	22 *SAVED*

第二次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-625.474)	N_CLASSES	2 FOUND ON TRY	1 DUPS 2 *SAVED*
PROBABILITY	exp(-669.097)	N_CLASSES	4 FOUND ON TRY	3 DUPS 2 *SAVED*
PROBABILITY	exp(-708.850)	N_CLASSES	6 FOUND ON TRY	4 DUPS 2 *SAVED*
PROBABILITY	exp(-727.594)	N_CLASSES	7 FOUND ON TRY	26 DUPS 1 *SAVED*
PROBABILITY	exp(-746.077)	N_CLASSES	8 FOUND ON TRY	11 DUPS 3 *SAVED*
PROBABILITY	exp(-781.392)	N_CLASSES	10 FOUND ON TRY	5 DUPS 3 *SAVED*
PROBABILITY	exp(-798.353)	N_CLASSES	11 FOUND ON TRY	9 DUPS 2 *SAVED*
PROBABILITY	exp(-814.791)	N_CLASSES	12 FOUND ON TRY	22 *SAVED*

第三次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-625.474)	N_CLASSES	2 FOUND ON TRY	1 DUPS 2 *SAVED*
PROBABILITY	exp(-669.097)	N_CLASSES	4 FOUND ON TRY	3 DUPS 2 *SAVED*
PROBABILITY	exp(-708.850)	N_CLASSES	6 FOUND ON TRY	4 DUPS 2 *SAVED*
PROBABILITY	exp(-727.594)	N_CLASSES	7 FOUND ON TRY	26 DUPS 1 *SAVED*
PROBABILITY	exp(-746.077)	N_CLASSES	8 FOUND ON TRY	11 DUPS 3 *SAVED*
PROBABILITY	exp(-781.392)	N_CLASSES	10 FOUND ON TRY	5 DUPS 3 *SAVED*
PROBABILITY	exp(-798.353)	N_CLASSES	11 FOUND ON TRY	9 DUPS 2 *SAVED*
PROBABILITY	exp(-814.791)	N_CLASSES	12 FOUND ON TRY	22 *SAVED*

附錄五

原輸入結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-625.474)	N_CLASSES	2 FOUND ON TRY	1 DUPS 2 *SAVED*
PROBABILITY	exp(-669.097)	N_CLASSES	4 FOUND ON TRY	3 DUPS 2 *SAVED*
PROBABILITY	exp(-708.850)	N_CLASSES	6 FOUND ON TRY	4 DUPS 2 *SAVED*
PROBABILITY	exp(-727.594)	N_CLASSES	7 FOUND ON TRY	26 DUPS 1 *SAVED*
PROBABILITY	exp(-746.077)	N_CLASSES	8 FOUND ON TRY	11 DUPS 3 *SAVED*
PROBABILITY	exp(-781.392)	N_CLASSES	10 FOUND ON TRY	5 DUPS 3 *SAVED*
PROBABILITY	exp(-798.353)	N_CLASSES	11 FOUND ON TRY	9 DUPS 2 *SAVED*
PROBABILITY	exp(-814.791)	N_CLASSES	12 FOUND ON TRY	22 *SAVED*

更改輸入順序結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-625.474)	N_CLASSES	2 FOUND ON TRY	1 DUPS 2 *SAVED*
PROBABILITY	exp(-647.587)	N_CLASSES	3 FOUND ON TRY	11 DUPS 4 *SAVED*
PROBABILITY	exp(-647.587)	N_CLASSES	3 FOUND ON TRY	17 DUPS 2 *SAVED*
PROBABILITY	exp(-669.097)	N_CLASSES	4 FOUND ON TRY	4 DUPS 3 *SAVED*
PROBABILITY	exp(-687.527)	N_CLASSES	5 FOUND ON TRY	20 DUPS 2 *SAVED*
PROBABILITY	exp(-689.154)	N_CLASSES	5 FOUND ON TRY	9 DUPS 1 *SAVED*
PROBABILITY	exp(-708.202)	N_CLASSES	6 FOUND ON TRY	22 *SAVED*
PROBABILITY	exp(-763.836)	N_CLASSES	9 FOUND ON TRY	6 DUPS 1 *SAVED*

附錄六

第一次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-1964.812)	N_CLASSES	2 FOUND ON TRY	1 DUPS 11 *SAVED*
PROBABILITY	exp(-1970.521)	N_CLASSES	3 FOUND ON TRY	2 *SAVED*
PROBABILITY	exp(-1986.373)	N_CLASSES	5 FOUND ON TRY	3 DUPS 1 *SAVED*
PROBABILITY	exp(-1993.178)	N_CLASSES	6 FOUND ON TRY	4 *SAVED*
PROBABILITY	exp(-2004.794)	N_CLASSES	8 FOUND ON TRY	10 DUPS 10 *SAVED*

第二次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-1964.812)	N_CLASSES	2 FOUND ON TRY	1 DUPS 1 *SAVED*
PROBABILITY	exp(-1970.695)	N_CLASSES	3 FOUND ON TRY	2 DUPS 21 *SAVED*
PROBABILITY	exp(-1979.843)	N_CLASSES	5 FOUND ON TRY	3 *SAVED*
PROBABILITY	exp(-1984.978)	N_CLASSES	7 FOUND ON TRY	5 *SAVED*
PROBABILITY	exp(-1985.076)	N_CLASSES	6 FOUND ON TRY	4 *SAVED*

第三次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-1964.812) N_CLASSES	2 FOUND ON TRY	1 *SAVED*	
PROBABILITY	exp(-1973.045) N_CLASSES	3 FOUND ON TRY	2 *SAVED*	
PROBABILITY	exp(-1979.838) N_CLASSES	5 FOUND ON TRY	3 *SAVED*	
PROBABILITY	exp(-1987.905) N_CLASSES	6 FOUND ON TRY	4 DUPS	6 *SAVED*
PROBABILITY	exp(-1988.895) N_CLASSES	7 FOUND ON TRY	14 DUPS	4 *SAVED*

第四次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-1964.812) N_CLASSES	2 FOUND ON TRY	1 *SAVED*	
PROBABILITY	exp(-1971.998) N_CLASSES	3 FOUND ON TRY	2 DUPS	23 *SAVED*
PROBABILITY	exp(-1980.334) N_CLASSES	4 FOUND ON TRY	3 *SAVED*	
PROBABILITY	exp(-1986.830) N_CLASSES	6 FOUND ON TRY	4 *SAVED*	
PROBABILITY	exp(-2002.102) N_CLASSES	8 FOUND ON TRY	5 *SAVED*	

第五次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-1964.812) N_CLASSES	2 FOUND ON TRY	1 *SAVED*	
PROBABILITY	exp(-1972.807) N_CLASSES	3 FOUND ON TRY	10 *SAVED*	
PROBABILITY	exp(-1974.251) N_CLASSES	3 FOUND ON TRY	2 *SAVED*	
PROBABILITY	exp(-1979.665) N_CLASSES	5 FOUND ON TRY	3 *SAVED*	
PROBABILITY	exp(-1986.795) N_CLASSES	7 FOUND ON TRY	4 *SAVED*	

第二十六次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-1972.413) N_CLASSES	3 FOUND ON TRY	14 DUPS	8 *SAVED*
PROBABILITY	exp(-1972.598) N_CLASSES	3 FOUND ON TRY	2 *SAVED*	
PROBABILITY	exp(-1973.811) N_CLASSES	4 FOUND ON TRY	3 DUPS	1 *SAVED*
PROBABILITY	exp(-1984.803) N_CLASSES	6 FOUND ON TRY	4 DUPS	1 *SAVED*
PROBABILITY	exp(-1991.747) N_CLASSES	7 FOUND ON TRY	5 *SAVED*	
PROBABILITY	exp(-2001.988) N_CLASSES	10 FOUND ON TRY	6 DUPS	1 *SAVED*

第二十九次分類結果

----- SUMMARY OF 10 BEST RESULTS -----				
PROBABILITY	exp(-1966.736) N_CLASSES	3 FOUND ON TRY	2 *SAVED*	
PROBABILITY	exp(-1970.554) N_CLASSES	3 FOUND ON TRY	30 *SAVED*	
PROBABILITY	exp(-1979.303) N_CLASSES	5 FOUND ON TRY	3 DUPS	6 *SAVED*
PROBABILITY	exp(-1983.530) N_CLASSES	6 FOUND ON TRY	4 *SAVED*	

附錄七

SUMMARY OF 15 BEST RESULTS					
PROBABILITY	exp(-560160.856)	N_CLASSES	29 FOUND ON TRY	49 *SAVED*	-1
PROBABILITY	exp(-560191.962)	N_CLASSES	31 FOUND ON TRY	94 *SAVED*	-2
PROBABILITY	exp(-560257.893)	N_CLASSES	31 FOUND ON TRY	98	

SUMMARY OF 15 BEST RESULTS					
PROBABILITY	exp(-560016.100)	N_CLASSES	34 FOUND ON TRY	34 *SAVED*	-1
PROBABILITY	exp(-560064.999)	N_CLASSES	39 FOUND ON TRY	95 *SAVED*	-2
PROBABILITY	exp(-560095.518)	N_CLASSES	39 FOUND ON TRY	58	

SUMMARY OF 15 BEST RESULTS					
PROBABILITY	exp(-560084.514)	N_CLASSES	34 FOUND ON TRY	69 *SAVED*	-1
PROBABILITY	exp(-560118.014)	N_CLASSES	31 FOUND ON TRY	63 *SAVED*	-2
PROBABILITY	exp(-560183.783)	N_CLASSES	29 FOUND ON TRY	76	

附錄八

SUMMARY OF 15 BEST RESULTS					
PROBABILITY	exp(-883120.978)	N_CLASSES	4 FOUND ON TRY 13 DUPS	9 *SAVED*	-1
PROBABILITY	exp(-883121.378)	N_CLASSES	4 FOUND ON TRY 14 DUPS	11 *SAVED*	-2
PROBABILITY	exp(-883122.232)	N_CLASSES	3 FOUND ON TRY	82	
PROBABILITY	exp(-883156.088)	N_CLASSES	2 FOUND ON TRY 1 DUPS	4	

SUMMARY OF 15 BEST RESULTS					
PROBABILITY	exp(-883120.860)	N_CLASSES	4 FOUND ON TRY 18 DUPS	8 *SAVED*	-1
PROBABILITY	exp(-883121.382)	N_CLASSES	4 FOUND ON TRY 9 DUPS	9 *SAVED*	-2
PROBABILITY	exp(-883121.986)	N_CLASSES	4 FOUND ON TRY	98	

SUMMARY OF 15 BEST RESULTS						##
PROBABILITY	exp(-883119.767)	N_CLASSES	4 FOUND ON TRY 95	*SAVED*	-1	
PROBABILITY	exp(-883120.979)	N_CLASSES	4 FOUND ON TRY 19 DUPS	10 *SAVED*	-2	
PROBABILITY	exp(-883121.370)	N_CLASSES	4 FOUND ON TRY 10 DUPS	16		

附錄九

REAL ATTRIBUTE (t = R)									
Num	t	mtt	description						
21	22	R SNcn	pc22	0.007	(5.95e-02	2.80e+00)	2.12e-02 (-3.36e-07	2.59e+00)
13	14	R SNcn	pc14	0.006	(-2.50e-01	3.49e+00)	7.18e-02 (1.05e-07	3.29e+00)
16	17	R SNcn	pc17	0.005	(9.29e-02	3.13e+00)	2.96e-02 (-1.12e-07	2.94e+00)

00 01 R SNcn	pc1	0.004 (1.96e+00	4.88e+01)	4.02e-02 (-2.01e-05	4.62e+01)
02 03 R SNcn	pc3	0.004 (-6.45e-01	2.85e+01)	2.26e-02 (-4.16e-06	2.68e+01)

附錄十

前兩個主成分

_____ SUMMARY OF 15 BEST RESULTS _____					
PROBABILITY	exp(-62985.366)	N_CLASSES	2 FOUND ON TRY 42	*SAVED*	-1
PROBABILITY	exp(-62985.479)	N_CLASSES	2 FOUND ON TRY 80	*SAVED*	-2
PROBABILITY	exp(-62985.647)	N_CLASSES	2 FOUND ON TRY 62 DUPS 2	*SAVED*	-3
PROBABILITY	exp(-62986.434)	N_CLASSES	2 FOUND ON TRY 34 DUPS 7	*SAVED*	-4
PROBABILITY	exp(-62987.089)	N_CLASSES	2 FOUND ON TRY 71	*SAVED*	-5
PROBABILITY	exp(-62987.465)	N_CLASSES	2 FOUND ON TRY 8 DUPS 6	*SAVED*	-6
PROBABILITY	exp(-63000.731)	N_CLASSES	3 FOUND ON TRY 2	*SAVED*	-7
PROBABILITY	exp(-63000.786)	N_CLASSES	3 FOUND ON TRY 68 DUPS 1	*SAVED*	-8
PROBABILITY	exp(-63001.400)	N_CLASSES	3 FOUND ON TRY 44	*SAVED*	-9

前三個主成分

_____ SUMMARY OF 15 BEST RESULTS _____					
PROBABILITY	exp(-95378.106)	N_CLASSES	2 FOUND ON TRY 1 DUPS 10	*SAVED*	-1
PROBABILITY	exp(-95379.032)	N_CLASSES	2 FOUND ON TRY 29 DUPS 25	*SAVED*	-2
PROBABILITY	exp(-95392.979)	N_CLASSES	3 FOUND ON TRY 18	*SAVED*	-3
PROBABILITY	exp(-95393.045)	N_CLASSES	3 FOUND ON TRY 42	*SAVED*	-4
PROBABILITY	exp(-95393.370)	N_CLASSES	3 FOUND ON TRY 24 DUPS 1	*SAVED*	5

前四個主成分

_____ SUMMARY OF 15 BEST RESULTS _____					
PROBABILITY	exp(-127330.252)	N_CLASSES	2 FOUND ON TRY 9 DUPS 12	*SAVED*	-1
PROBABILITY	exp(-127330.835)	N_CLASSES	2 FOUND ON TRY 1 DUPS 3	*SAVED*	-2
PROBABILITY	exp(-127351.120)	N_CLASSES	3 FOUND ON TRY 2	*SAVED*	-3
PROBABILITY	exp(-127351.499)	N_CLASSES	3 FOUND ON TRY 8 DUPS 1	*SAVED*	-4
PROBABILITY	exp(-127351.572)	N_CLASSES	3 FOUND ON TRY 15 DUPS 3	*SAVED*	5

前五個主成分

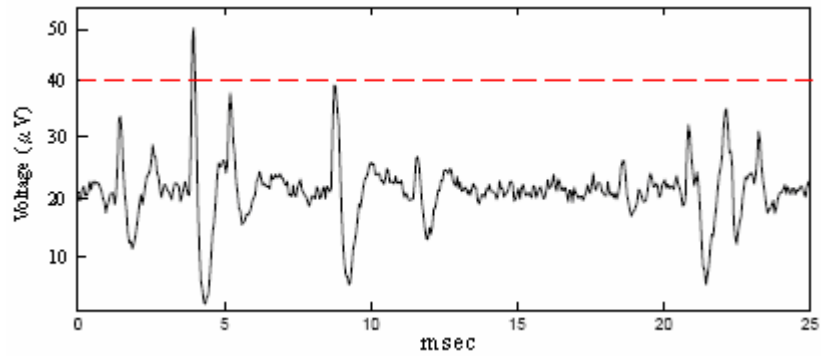
_____ SUMMARY OF 15 BEST RESULTS _____					
PROBABILITY	exp(-156982.778)	N_CLASSES	2 FOUND ON TRY 1	*SAVED*	-1
PROBABILITY	exp(-156983.300)	N_CLASSES	2 FOUND ON TRY 6 DUPS 28	*SAVED*	-2
PROBABILITY	exp(-156995.947)	N_CLASSES	3 FOUND ON TRY 58 DUPS 1	*SAVED*	-3
PROBABILITY	exp(-156996.070)	N_CLASSES	3 FOUND ON TRY 52	*SAVED*	-4

附錄十一
前五個主成分

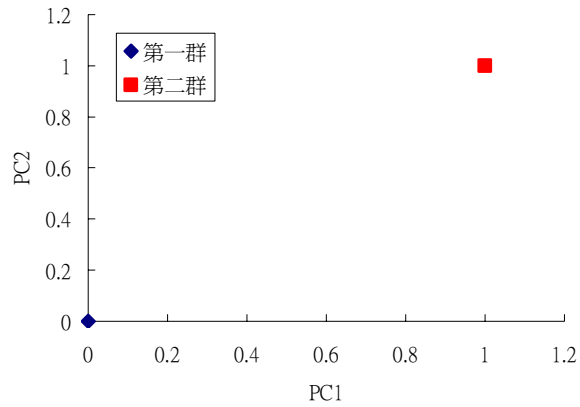
SUMMARY OF 15 BEST RESULTS			
PROBABILITY	$\exp(-146216.061)$	N_CLASSES	2 FOUND ON TRY 48 DUPS 2 *SAVED* -1
PROBABILITY	$\exp(-146216.510)$	N_CLASSES	2 FOUND ON TRY 1 DUPS 12 *SAVED* -2
PROBABILITY	$\exp(-146230.507)$	N_CLASSES	3 FOUND ON TRY 78 DUPS 4 *SAVED* -3
PROBABILITY	$\exp(-146232.263)$	N_CLASSES	3 FOUND ON TRY 55 *SAVED* -4

附圖

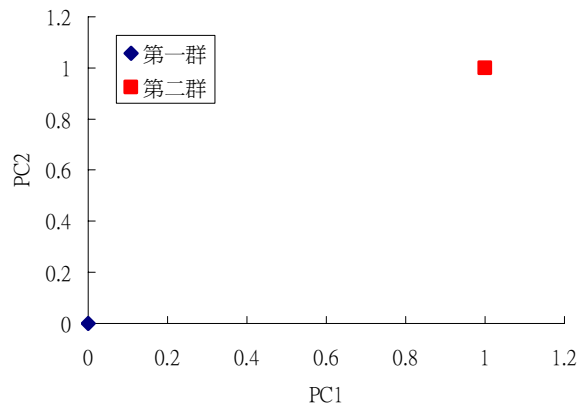
圖一 動作電位門檻圖【虛線為動作電位之門檻值】



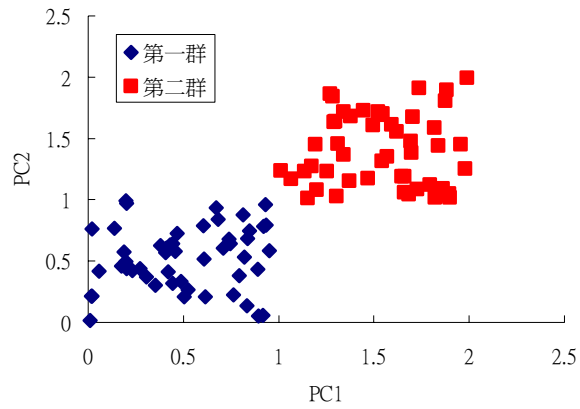
圖二 測試資料之分布圖【含 50 筆 (0, 0) 及 50 筆 (1, 1)】



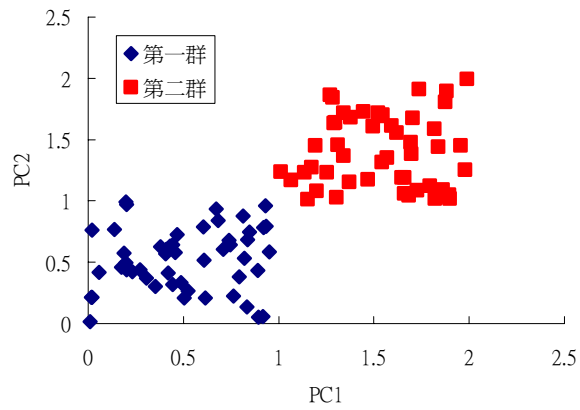
圖三 五次的最佳分類圖【分類結果確實把 50 筆 (0, 0) 及 50 筆 (1, 1) 分開】



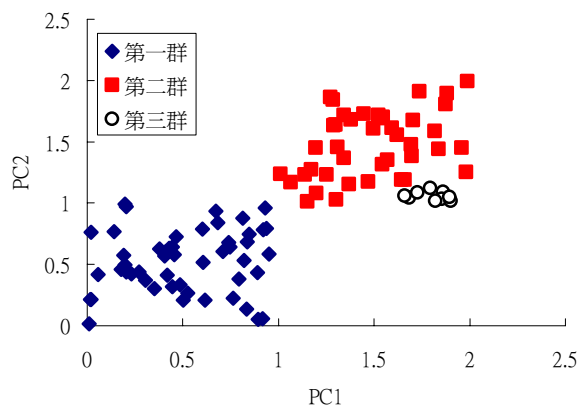
圖四 測試資料之分布圖【50筆(0,0)及50筆(1,1)分別加上0到1之間的雜訊】



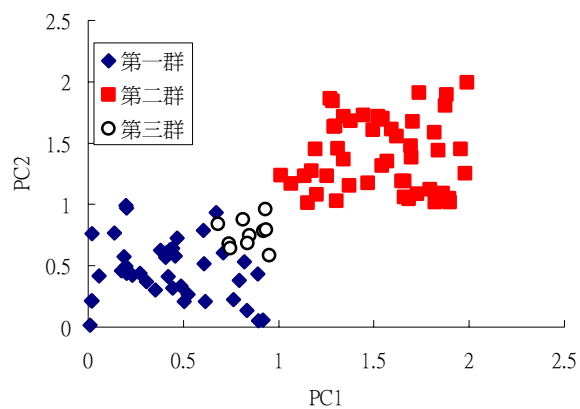
圖五 多次的最佳分類圖【加上雜訊後的資料仍被正確分開】



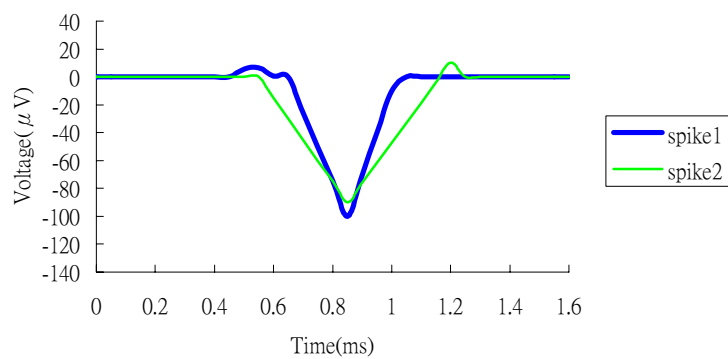
圖六 第二十六次的最佳分類圖



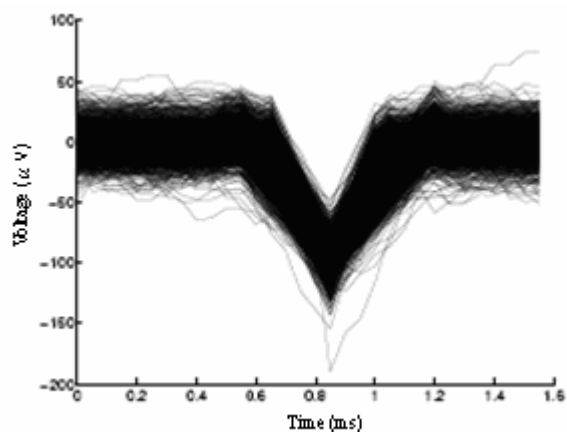
圖七 第二十九次的最佳分類圖



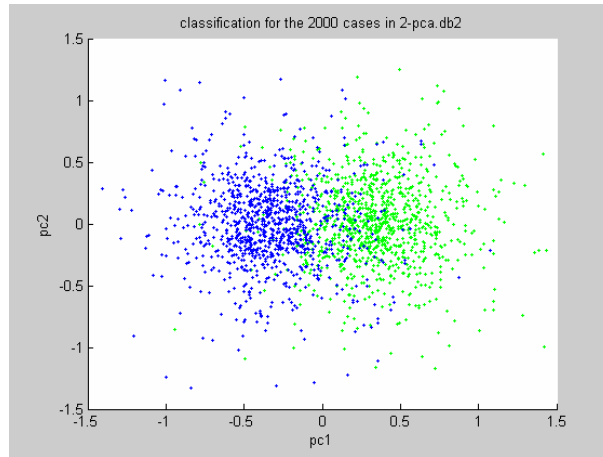
圖八 人工動作電位圖【spike1 與 spike2 之波形】



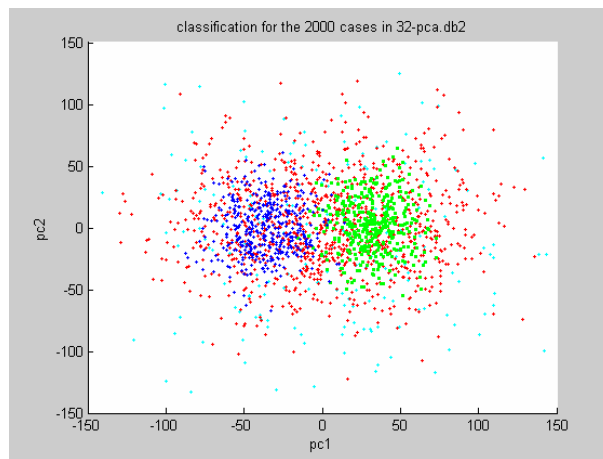
圖九 人工動作電位加上真實雜訊【spike1 與 spike2 各加上一千組真實雜訊所得的兩千筆動作電位】



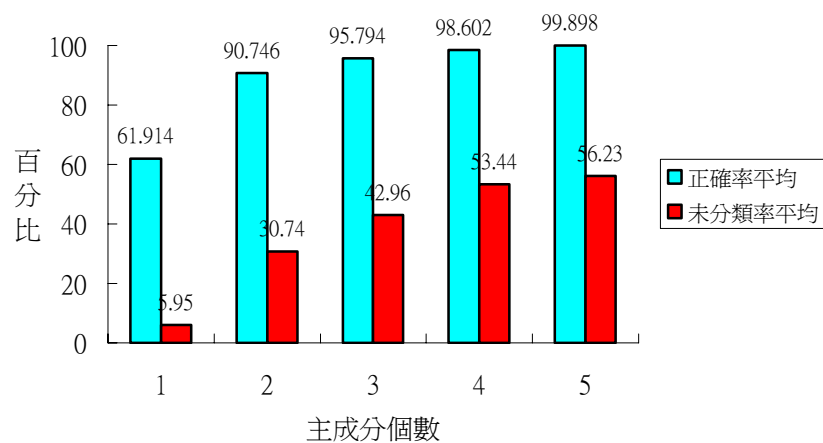
圖十 PCA 圖【圖九的兩千筆動作電位經 PCA 轉換後，以第一及第二主成分所繪製的資料分布圖】



圖十一 PCA 圖【圖九的兩千筆動作電位經 PCA 轉換後，Autoclass 對其 32 個主成分分類所得的結果圖】



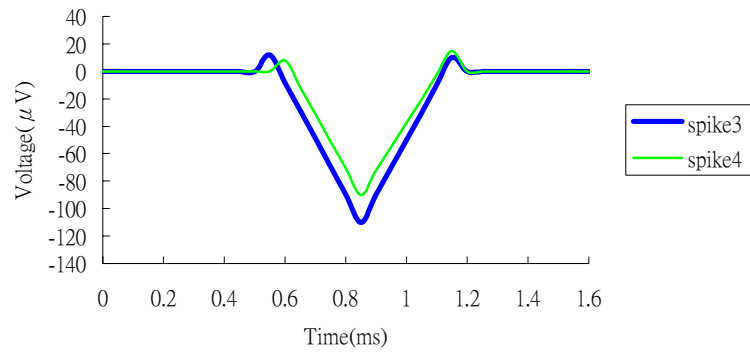
圖十二 實驗一之主成分個數與分類正確率關係圖



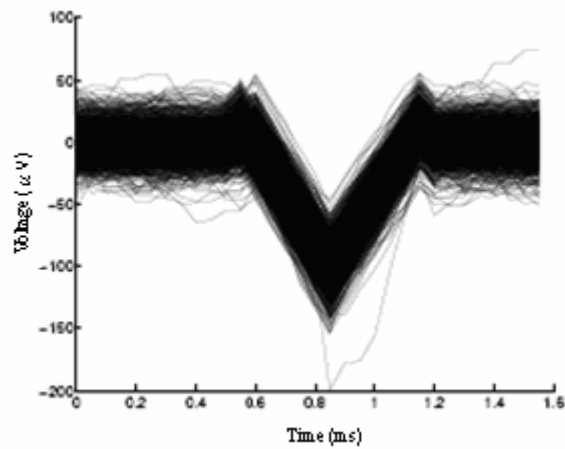
圖十三 PCA 圖【未分類點強制分類後以第一及第二主成分所繪製的資料分布圖】



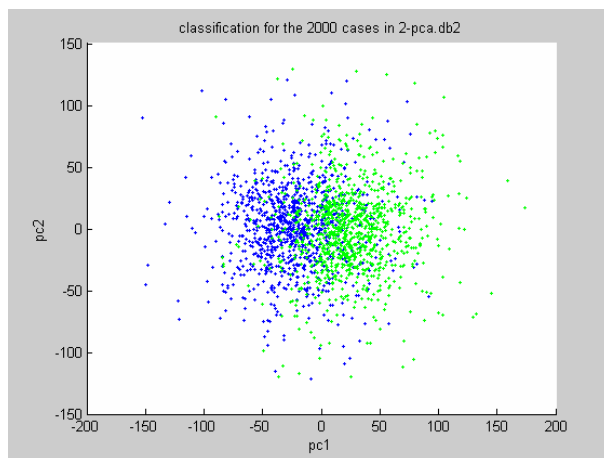
圖十四 人工動作電位圖【spike3 與 spike4 之波形】



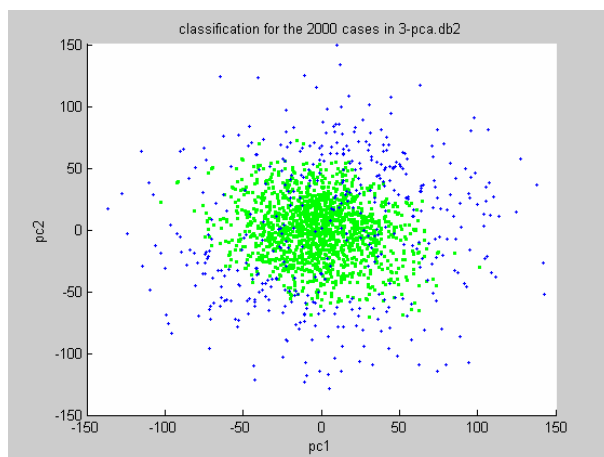
圖十五 人工動作電位加上真實雜訊【spike3 與 spike4 各加上一千組真實雜訊所得的兩千筆動作電位】



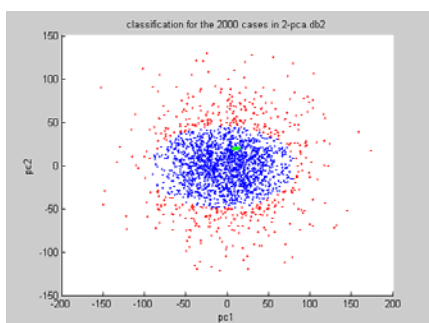
圖十六 PCA 圖【圖十五的兩千筆動作電位經 PCA 轉換後，以第一及第二主成分所繪製的資料分布圖】



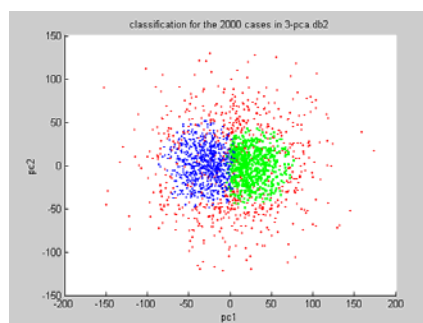
圖十七 PCA 圖【以前幾個主成分分類後，用第一及第二主成分所繪製的資料分布圖】



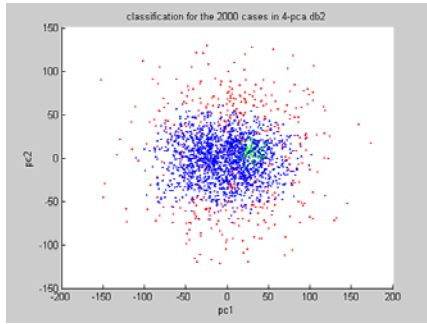
圖十八 PCA 圖【以前五個主成分分類後，用第一及第二主成分所繪製的資料分布圖】



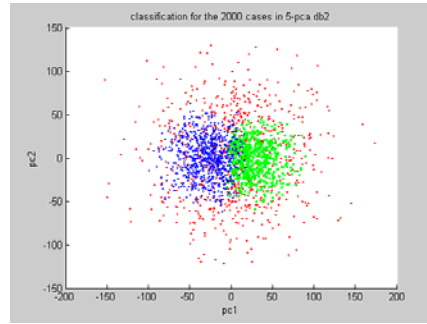
前兩個主成分



前三個主成分

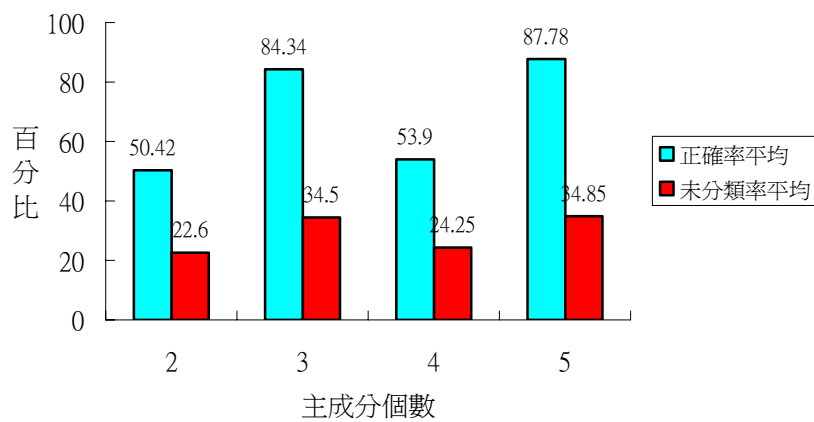


前四個主成分

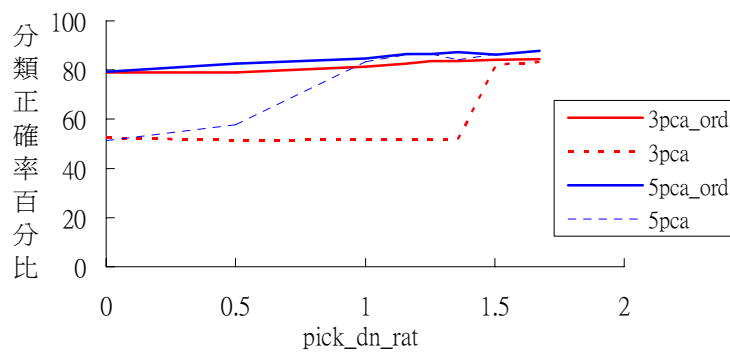


前五個主成分

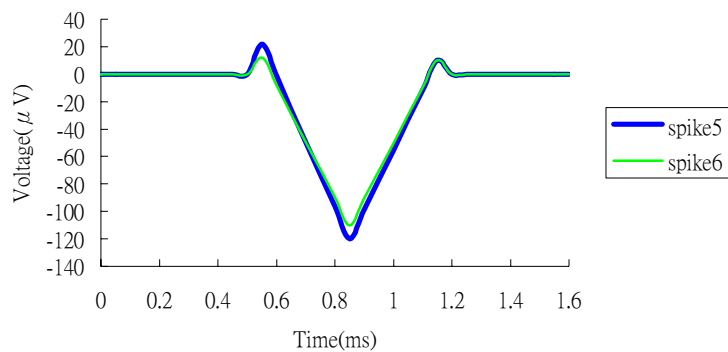
圖十九 實驗二之主成分個數與分類正確率關係圖



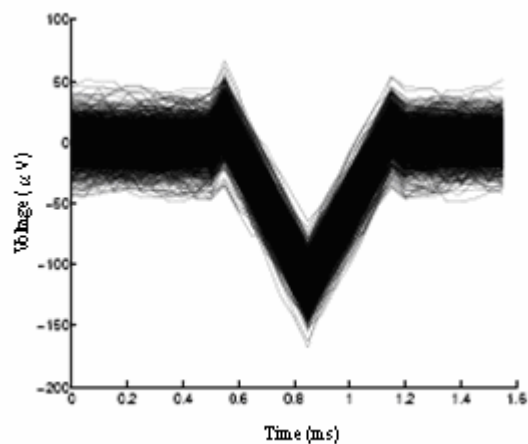
圖二十 主成分重要性與分類正確率關係圖【pick_dn_rat 為兩動作電位最低點位置的差距除以兩動作電位最低點位置上的雜訊標準差。實線是由分類結果受到第一主成分影響最大，第二主成分次之，依序排列的分類正確率所構成。虛線則是由主成分重要性任意排列所得的分類正確率構成】



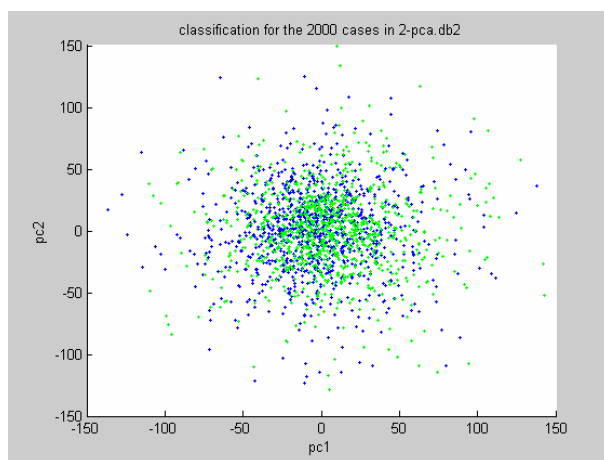
圖二十一 人工動作電位圖【spike5 與 spike6 之波形】



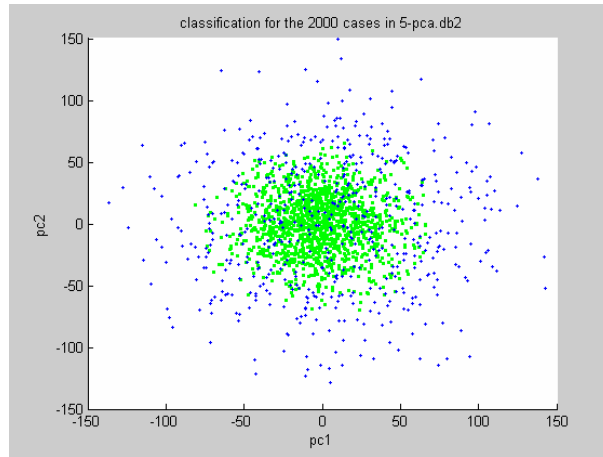
圖二十二 人工動作電位加上真實雜訊【spike5 與 spike6 各加上一千組真實雜訊所得的兩千筆動作電位】



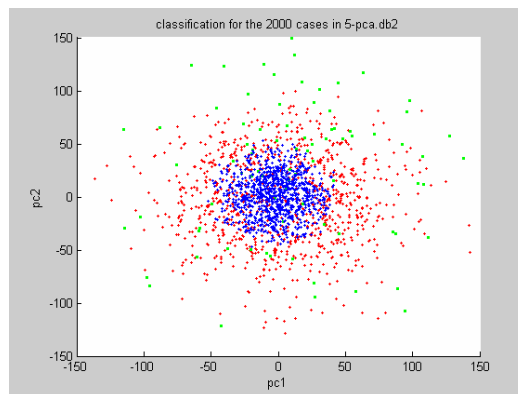
圖二十三 PCA 圖【圖二十二的兩千筆動作電位經 PCA 轉換後，以第一及第二主成分所繪製的資料分布圖】



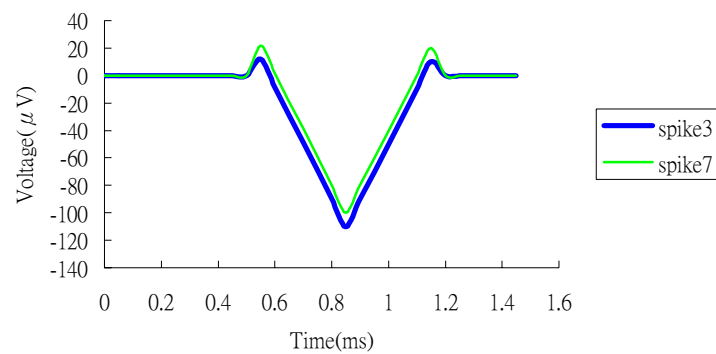
圖二十四 PCA 圖【以前五個主成分分類所得的兩群分類結果，用第一及第二主成分所繪製的資料分布圖】



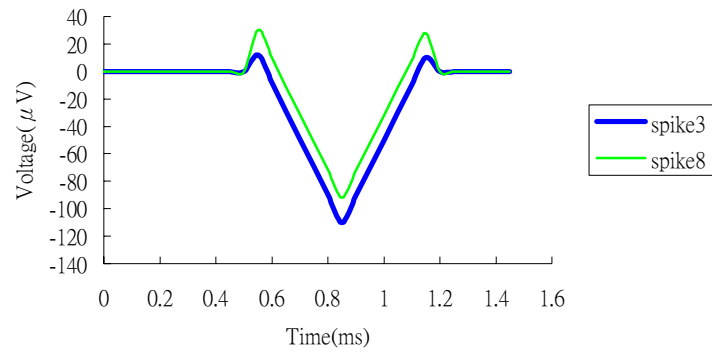
圖二十五 PCA 圖【以前五個主成分分類所得的三群分類結果，用第一及第二主成分所繪製的資料分布圖】



圖二十六 人工動作電位圖【spike3 與 spike7 之波形。其波形最大差距為 0.85 倍的雜訊最大標準差】



圖二十七 人工動作電位圖【spike3 與 spike8 之波形。其波形最大差距為 1.5 倍的雜訊最大標準差】



圖二十八 m_dn_rat 與分類正確率關係圖【m_dn_rat 為兩動作電位最大的差距除以雜訊最大的標準差】

