

蒙地卡羅模擬實驗：多種分類方法的評比

郭翊萱 711133115

2022.11

在前面的章節中，分別對簡單線性迴歸 (Simple Linear Regression)、加廣迴歸模型 (Augmented Regression)、羅吉斯迴歸 (Logistic Regression) 三種決定性 (Deterministic) 機器學習方法與線性判別分析 (LDA)、二次判別分析 (QDA)、K-近鄰演算法三種機率性 (Probabilistic) 機器學習方法跟類神經網路模型的分類問題進行了探討，也大致了解了不同方法的理論以及分類上可能遇到的問題以及特性，但是卻未同時對決定性機器學習方法、機率性機器學習方法與類神經網路的分類準確度進行比較，因此在此章節中將模擬生成多個兩群與三群的雙變量常態資料來同時比較這七種方法的分類準確度。另外，本章節亦會嘗試生成非常態的資料來進行分類問題之比較。

1 七種分類方法介紹

1.1 簡單線性迴歸簡述

線性迴歸 (Linear Regression) 是一種常見的監督式學習方法，是一種常見的預測定量響應變量 (Response) 的工具，它假定了 X 與 Y 之間存在線性關係，在數學上將這種關係記為：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_i \quad (1)$$

在上式 (1) 中假設存在兩個預測變量 X_1 與 X_2 ，假設有 N 筆已知的輸入與輸出資料 ($[x_1(i) \ x_2(i)], y(i)$)，利用最小平方法期望使觀察值與估計值的誤差越小越好。其計算方法如下矩陣：

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1(1) & X_2(1) \\ 1 & X_1(2) & X_2(2) \\ \vdots & & \\ 1 & X_1(n) & X_2(n) \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}$$

透過上述矩陣即可求得最佳解如式 (2)，式 (2) 假設反矩陣 $(X^T X)^{-1}$ 存在，且每個輸出值 $y(i)$ 根據其類別分類至 0 或 1。

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (2)$$

1.2 加廣迴歸簡述

加廣迴歸模型是一種非線性的迴歸模型，它透過擴充簡單線性迴歸模型來試圖得到較好的分類準確度，其重新假設迴歸模型如式 (3)：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \beta_4 X_1^2 + \beta_5 X_2^2 \quad (3)$$

1.3 羅吉斯迴歸簡述

假設存在兩個類別來解釋羅吉斯迴歸的簡單原理。在線性迴歸中利用 $p(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ 來表示機率，而在羅吉斯迴歸中，則使用羅吉斯函數式 (4) 結合最大概似估計 (MLE) 方法來擬合模型。

$$p(X) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2}} \quad (4)$$

接著將式 (4) 改寫成式 (5)，並將其稱之為勝算比 (odd)。

$$\frac{p(X)}{1 - p(X)} = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2} \quad (5)$$

若將式 (5) 取 log，則將其稱為對數勝算比 (log-odd)，寫作式 (6)

$$\log\left(\frac{p(X)}{1 - p(X)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \quad (6)$$

1.4 線性判別分析簡述

在線性判別分析 (Linear Discriminant Analysis) 中，試圖找出預測變數 X 在不同的反應變數 G (不同類別) 的分配，並利用貝氏定理去估計 $P(G = k | X = x)$ 。假設先驗分配 (prior probability) ($P(G = k)$) 代表隨機從類別 k 抽出觀察值的機率，而後驗分配 $f_k(x) = P(X = x | G = k)$ 代表觀察值來自類別 k 時 X 的機率密度函數，則透過貝氏定理可知：

$$P(G = k|X = x) = \frac{P(G = k)P(X = x|G = k)}{\sum_{l=1}^k P(G = l)P(x = X|g = l)} \quad (7)$$

其中，若使後驗分配 $P(X = x|G = k)$ 最大，則此分類的分類誤差 (error rate) 最少。

假設 $X \sim N(\mu, \Sigma)$ ，且其機率密度函數如式(8)：

$$f(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)) \quad (8)$$

而線性判別分析 (LDA) 會將觀察值 $X = x$ 分類至使式(9)的值最大的第 k 類。

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \ln P(G = k) \quad (9)$$

透過以上式(9)即可得到線性判別分析的分群結果。

1.5 二次判別分析簡述

在線性判別分析中假設母體服從常態分配且每個類別的共變異數相同 (共變異矩陣一樣)，但在實務上往往很難滿足此假設，而二次判別分析中完善了此一缺點。二次判別分析假設了每個類別都有各自的共變異數，因此式(9)可被改寫成如式(10)：

$$\begin{aligned} \delta_k(x) &= -\frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k) - \frac{1}{2} \ln |\Sigma_k| + \ln P(G = k) \\ &= -\frac{1}{2} x^T \Sigma_k^{-1} x + x^T \Sigma_k^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma_k^{-1} \mu_k - \frac{1}{2} \ln |\Sigma_k| + \ln P(G = k) \end{aligned} \quad (10)$$

其中，二次判別分析 (QDA) 需估計 Σ_k, μ_k 以及 $P(G = k)$ 並將觀察值 $X = x$ 分類至使式(10)的值最大的類別 k 。

另外，二次判別分析的分界線為變數的二次方，被寫作如式(11)

$$\{x | \delta_k(x) = \delta_l(x)\} \quad (11)$$

1.6 K-近鄰演算法簡述

雖然在機器學習中有幾種常見的方法是利用貝氏定理來求得條件機率 $P(Y|X = x)$ ，其中 x 與 y 分別是反應變數與預測變數，但在真實世界中，很難知道此條件機率的值，

因此利用貝氏定理來分類有時是不可行的，而有一些方法便嘗試去估計 $P(Y|X = x)$ ，其中包含 K-近鄰演算法 (KNN)。K-近鄰演算法透過給定 K 值以及任一測試點 x_0 ，觀察離該點 x_0 最近的 K 個點 (即 $N_k(x)$)，並將這些鄰近的 K 筆資料所對應的 y 值取平均，得到式 (12)：

$$\hat{y} = Ave(y_i | x_i \in N_k(x)) = \frac{1}{K} \sum_{x_i \in N_k(x)} y_i \quad (12)$$

且其給定 x 時 y 的條件機率如式 (13)：

$$P(Y = j | X = x_0) = \frac{1}{K} \sum_{i \in N_0} I(y_i = j) \quad (13)$$

1.7 類神經網路模型簡述

類神經網路是一種模仿大腦運作的機器學習方式，透過大量的人工神經元連結進行計算，並能自行透過學習更多資訊進而對模型進行改進，其結構一般包含輸入層 (Input)、隱藏層 (Hidden Layer) 以及輸出層 (Output)，其中隱藏層的數量是可以進行改變的，一般會遵循以下步驟：

- 建立包含未知參數的函數

假設函數為 $y = b + wx_1$ ，其中 x_1 : views on 2/25、 y : views on 2/26， w (weight) 與 b (bias) 為模型之未知參數，其從已有資料中進行學習。

- 定義損失函數 (Loss Function)

定義函數的輸入 (Input) 為未知參數 b 、 w ，而損失函數 (Loss Function) 則用來表示在將 input 丟入函數後，此函數的預測效果有多好，其函數可寫作如式 (14)。

$$Loss = \frac{1}{N} \sum_n e_n \quad (14)$$

其中 $e = (y - \hat{y})^2$ 。

- 利用不同的未知參數計算不同的損失函數並畫圖
- 最佳化模型 (Optimization)

最佳化模型有很多方法，其中有一個常見的方法是梯度下降 (Gradient Descent)，嘗試求得合適的未知參數 w 、 b 使得損失函數的值最小，在此文不進行過多討論。

在此步驟中，所有的模型調整全部建立在線性模型 (Linear Model) 之上進行調整，但線性模型往往太過簡單，就算選出了具有最小損失函數的模型，其依舊被初始設置的線性模型所限制 (model bias)。而如何設置更複雜的模型？Sigmoid Function 協助達成了此目的，其被表示如式 (15)。

$$\text{sigmoid} = \frac{1}{1 + e^{-x}} \quad (15)$$

故將新的預測模型寫作式 (16)。

$$y = c \frac{1}{1 + e^{-(b+wx_1)}} = c \times \text{sigmoid}(b + wx_1) \quad (16)$$

式 (16) 中的 w 控制函數坡度、 b 控制左右移動， c 則可以改變函數的高度，若是具有兩個隱藏層，則利用此原理可以將類神經網路模型寫作式 (17)。

$$\hat{y} = b_k^2 + \sum_{i=1}^q w_{ki}^2 g\left(\sum_{j=1}^p w_{ij}^1 x_j + b_i^1\right) + b_k^2, 1 \leq k \leq r \quad (17)$$

其中函式 $g(\bullet)$ 為映射函數 (Sigmoid Function)，其數學式如式 (18)。

$$g(z) = c_1 \frac{1 - e^{-c_2 z}}{1 + e^{-c_2 z}} \quad (18)$$

在式 (17) 中， q 為映射函數 (Sigmoid Function) 的數量、 p 為輸入的特徵 (feature) 的數量、 w_{ij}^1 與 b_i^1 代表第一個隱藏層的第 i 個神經元與第 j 個 input 的權重 (Weight) 跟偏誤 (Biases)，而 w_{ki}^2 與 b_k^2 則是第二層隱藏層的第 k 個神經元與前一隱藏層的第 i 個輸出的權重與偏誤，為了使模型最佳化，期望能得到使此模型的損失函數 (Loss Function) 最小的未知參數。

以上即是類神經網路的基本原理，但據此亦可知雖然其原理簡單有效，如果資料為高維資料時，如何求得使損失函數最小的未知參數也變得困難，如何定義需要多少隱藏層也成為挑戰。

2 七種機器學習方法對兩群組資料的分類問題

在此章節中將分別建立兩群組的雙變量常態資料跟三群組的雙變量常態資料，利用上面所述的七種機器學習方法來建立模型進行分類並比較其訓練資料準確度與測試資料準確度。與上面幾個章節相同，本節將模擬各種不同樣本數、不同平均向量與不同共

變異數的資料集來進行比較，首先介紹模擬兩群組雙變量常態資料的情況，並嘗試改變其樣本數大小來觀察其訓練資料準確度與測試資料準確度。

2.1 兩群組更改樣本數大小的分類問題

此節設置雙變量常態資料的參數如下，其散布圖亦呈現於圖 1。

$$\mu_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \mu_2 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

$$\Sigma_1 = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}$$

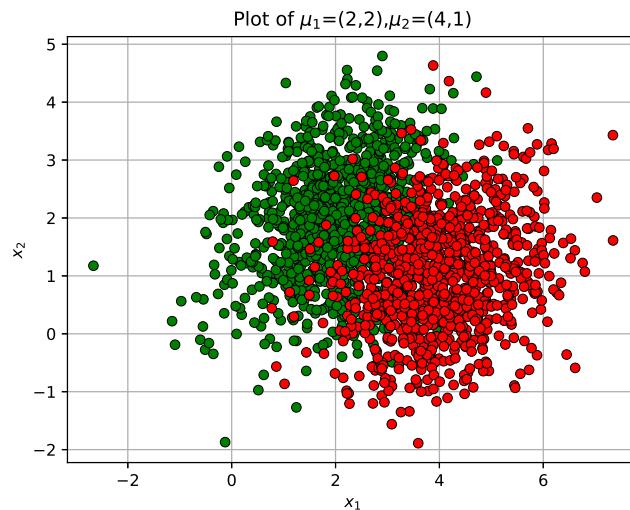


圖 1: 比較七種方法改變樣本數實驗的兩群組雙變量常態資料散布圖

根據圖 1 可以觀察到此兩群組雙變量常態資料的關係，接著嘗試利用七種機器學習方法建立分類模型，其中 KNN 模型的 K 值在此章節中都設定為 10，並利用 Bootstrapping 方法對資料抽樣 100 次並分別求其誤差，再將 100 次的誤差取平均後即可得到如表 1 與表 2 的結果。

表 1: 七種機器學習方法改變樣本數的兩群組訓練資料分類錯誤率

	Regression	Augmented	Logistic	LDA	QDA	KNN	ANN
$n_1 = n_2 = 200$	0.146	0.142	0.147	0.146	0.147	0.116	0.145
$n_1 = n_2 = 500$	0.120	0.116	0.120	0.120	0.115	0.096	0.118
$n_1 = n_2 = 1000$	0.112	0.111	0.112	0.112	0.116	0.099	0.114
$n_1 = 1000, n_2 = 500$	0.088	0.098	0.091	0.087	0.09	0.075	0.087

表 2: 七種機器學習方法改變樣本數的兩群組測試資料分類錯誤率

	Regression	Augmented	Logistic	LDA	QDA	KNN	ANN
$n_1 = n_2 = 200$	0.135	0.134	0.135	0.135	0.15	0.15	0.137
$n_1 = n_2 = 500$	0.1195	0.1175	0.1188	0.1195	0.115	0.14	0.118
$n_1 = n_2 = 1000$	0.112	0.112	0.112	0.112	0.085	0.099	0.115
$n_1 = 1000, n_2 = 500$	0.086	0.097	0.089	0.085	0.073	0.087	0.086

根據表 1 與表 2 可知，雖然 KNN 模型無論在樣本數為多少時訓練誤差都是最小的，似乎應該是表現最好的分類模型，但是測試誤差卻是加廣迴歸模型 (Augmented Regression) 與 QDA 模型在不同樣本數下的表現最佳。

2.2 兩群組更改群平均的分類問題

接著觀察改變群平均時七種機器學習方法的訓練誤差與測試誤差，其雙變量常態資料的參數亦呈現如下，散佈圖則如圖 2，此兩組雙變量常態資料會與圖 1 的散佈圖資料一起比較，最後會得到表 3 與 4 的結果。

$$\mu_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \mu_2 = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

$$\mu'_1 = \begin{bmatrix} 2 \\ 6 \end{bmatrix}, \mu'_2 = \begin{bmatrix} 5 \\ 9 \end{bmatrix}$$

$$\Sigma_1 = \Sigma'_1 = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}, \Sigma_2 = \Sigma'_2 = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}$$

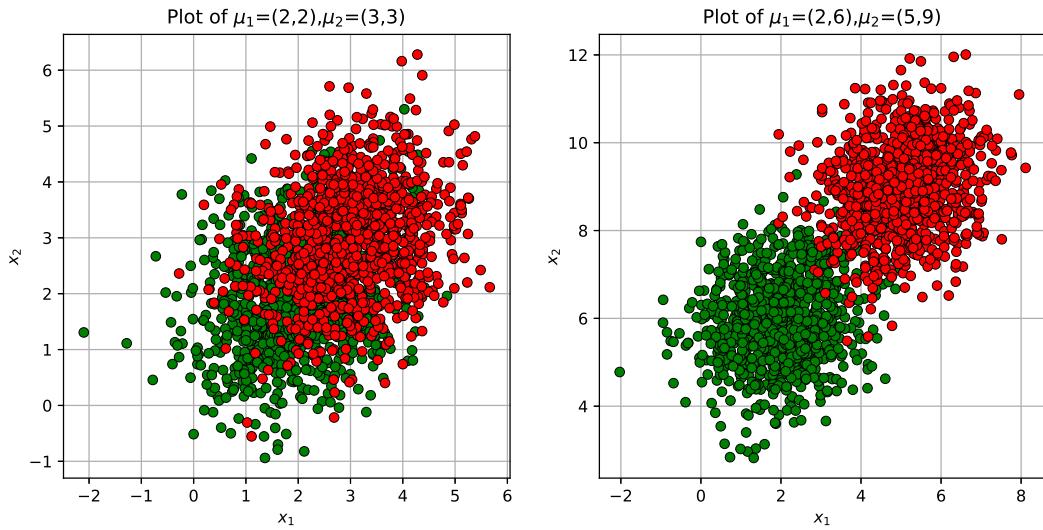


圖 2: 比較七種方法改變群平均實驗的兩群組雙變量常態資料散布圖

表 3: 七種機器學習方法改變群平均的兩群組訓練資料分類錯誤率

	Regression	Augmented	Logistic	LDA	QDA	KNN	ANN
$\mu_1 = (2, 2), \mu_2 = (4, 1)$	0.112	0.111	0.112	0.112	0.116	0.099	0.114
$\mu_1 = (2, 2), \mu_2 = (3, 3)$	0.274	0.273	0.274	0.274	0.276	0.211	0.274
$\mu_1 = (2, 6), \mu_2 = (5, 9)$	0.032	0.033	0.031	0.032	0.031	0.027	0.032

根據表 3 與表 4 可知，平均數向量的改變並不影響 KNN 模型是七種方法中訓練資料準確度最高的模型，但是與樣本數的實驗相同，KNN 模型並非測試資料中準確度最高的模型，反而 QDA 模型似乎與樣本數實驗相同，都是七種方法中表現最好的模型。

表 4: 七種機器學習方法改變群平均的兩群組測試資料分類錯誤率

	Regression	Augmented	Logistic	LDA	QDA	KNN	ANN
$\mu_1 = (2, 2), \mu_2 = (4, 1)$	0.112	0.112	0.112	0.112	0.085	0.099	0.115
$\mu_1 = (2, 2), \mu_2 = (3, 3)$	0.276	0.276	0.276	0.276	0.265	0.27	0.276
$\mu_1 = (2, 6), \mu_2 = (5, 9)$	0.031	0.032	0.031	0.031	0.037	0.04	0.031

2.3 兩群組更改群共變異數的分類問題

接著來看改變共變異數時七種機器學習方法的分類效果，其參數設定如下所示，此小節中除了比較圖 1 中的雙變量常態資料外，也比較另外四種雙變量常態資料，其散布圖如圖 3。

$$\mu_1 = \mu'_1 = \mu''_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \mu_2 = \mu'_2 = \mu''_2 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

$$\Sigma_1 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

$$\Sigma'_1 = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}, \Sigma'_2 = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$

$$\Sigma''_1 = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}, \Sigma''_2 = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix}$$

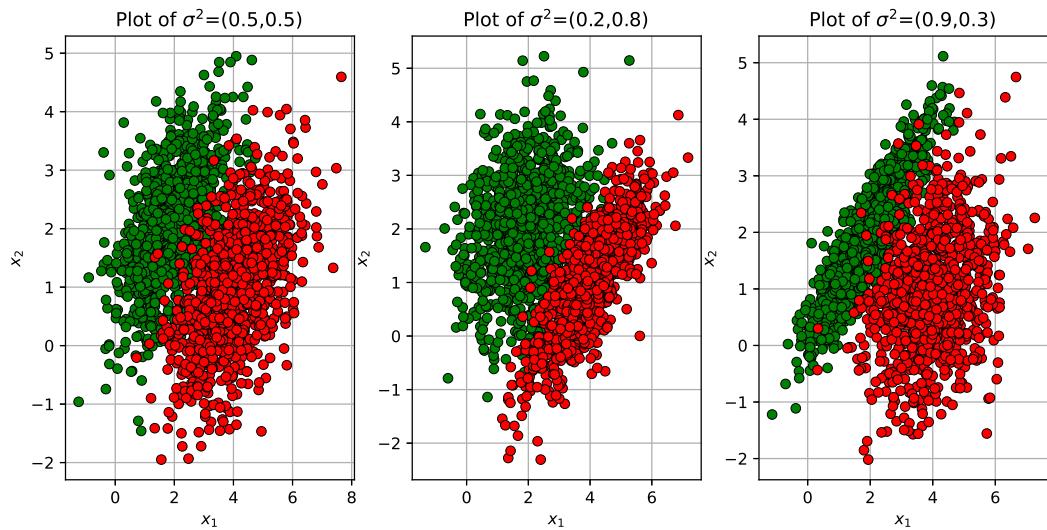


圖 3: 比較七種方法改變群共變異數實驗的兩群組雙變量常態資料散佈圖

表 5: 七種機器學習方法改變群共變異數的兩群組訓練資料分類錯誤率

	Regression	Augmented	Logistic	LDA	QDA	KNN	ANN
$\sigma^2 = (0.2, 0.2)$	0.112	0.111	0.112	0.112	0.116	0.099	0.114
$\sigma^2 = (0.5, 0.5)$	0.032	0.033	0.031	0.032	0.031	0.027	0.032
$\sigma^2 = (0.2, 0.8)$	0.053	0.052	0.053	0.053	0.057	0.046	0.052
$\sigma^2 = (0.9, 0.3)$	0.052	0.032	0.032	0.052	0.027	0.027	0.031

表 6: 七種機器學習方法改變群共變異數的兩群組測試資料分類錯誤率

	Regression	Augmented	Logistic	LDA	QDA	KNN	ANN
$\sigma^2 = (0.2, 0.2)$	0.112	0.112	0.112	0.112	0.085	0.099	0.115
$\sigma^2 = (0.5, 0.5)$	0.031	0.032	0.031	0.031	0.037	0.04	0.031
$\sigma^2 = (0.2, 0.8)$	0.054	0.050	0.052	0.054	0.027	0.04	0.05
$\sigma^2 = (0.9, 0.3)$	0.052	0.032	0.052	0.027	0.026	0.0958	0.12

圖 3 是實驗中的雙變量常態資料散布圖，而表 5 與表 6 則是七種機器學習方法的分類準確度，根據兩個表格可知，KNN 模型依舊是訓練模型中表現最好的模型，但是 KNN 模型的測試資料準確度並非最高，QDA 模型則是整體表現最好的分類模型。至此本節完成了對兩群組雙變量常態的分類效果之觀察，下一節中則將繼續觀察七種機器學習方法在三群組雙變量常態資料的分類準確度。

3 七種機器學習方法對三群組資料的分類問題

在此章節中將分別針對七種機器學習方法建立分類模型，並比較這七種方法的訓練資料誤差與測試資料誤差，判斷出哪個模型在三群組分類中的表現是最好的，此節與前述章節相同，資料都是來自於雙變量常態母體資料，並透過三種實驗，改變樣本數大小、改變平均數與改變共變異數來比較這七種機器學習方法。

3.1 三群組更改樣本數大小的分類問題

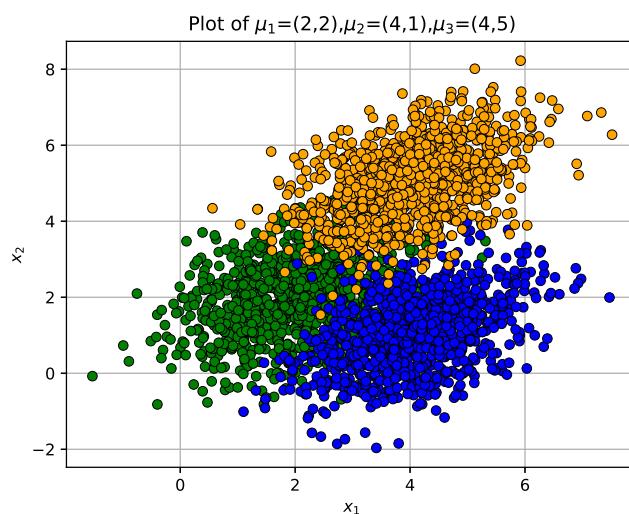


圖 4: 比較七種方法改變群樣本數實驗的三群組雙變量常態資料散布圖

$$\mu_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \mu_2 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \mu_3 = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

$$\Sigma_1 = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}$$

首先觀察改變樣本數大小造成的分類準確度變化，圖 4 是此實驗中用到的雙變量常態資料的散布圖，其資料的參數亦呈現於上方。

表 7: 七種機器學習方法改變樣本數的三群組訓練資料分類錯誤率

	Regression	Augmented	Logistic	LDA	QDA	KNN	ANN
$n_1 = n_2 = n_3 = 200$	0.467	0.480	0.366	0.365	0.356	0.262	0.368
$n_1 = n_2 = n_3 = 500$	0.477	0.489	0.390	0.385	0.372	0.286	0.384
$n_1 = n_2 = n_3 = 1000$	0.481	0.493	0.408	0.405	0.401	0.288	0.404
$n_1 = 1000, n_2 = 500, n_3 = 200$	0.247	0.226	0.294	0.202	0.199	0.165	0.203

表 7 與表 8 即為改變樣本數時七種機器學習的訓練誤差與測試誤差的比較，從表中可以知道，KNN 模型是訓練資料中誤差最小表現最好的模型，但是與兩群組實驗相同的是 QDA 模型才是七個模型中綜合表現最好的模型。

表 8: 七種機器學習方法改變樣本數的三群組測試資料分類錯誤率

	Regression	Augmented	Logistic	LDA	QDA	KNN	ANN
$n_1 = n_2 = n_3 = 200$	0.463	0.475	0.383	0.401	0.433	0.358	0.397
$n_1 = n_2 = n_3 = 500$	0.474	0.486	0.396	0.399	0.363	0.445	0.397
$n_1 = n_2 = n_3 = 1000$	0.477	0.489	0.413	0.415	0.388	0.437	0.415
$n_1 = 1000, n_2 = 500, n_3 = 200$	0.248	0.226	0.304	0.203	0.218	0.247	0.204

3.2 三群組更改群平均的分類問題

$$\mu_1 = \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \mu_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix}, \mu_3 = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$$

$$\mu'_1 = \begin{bmatrix} 1 \\ 6 \end{bmatrix}, \mu'_2 = \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \mu'_3 = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$$

$$\Sigma_1 = \Sigma'_1 = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix}, \Sigma_2 = \Sigma'_2 = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix}, \Sigma_3 = \Sigma'_3 = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix}$$

接著來觀察三群組雙變量常態資料改變群平均實驗的七種機器學習方法的分類效果，其使用的資料如圖 4 與 5，其資料參數亦呈現如上。

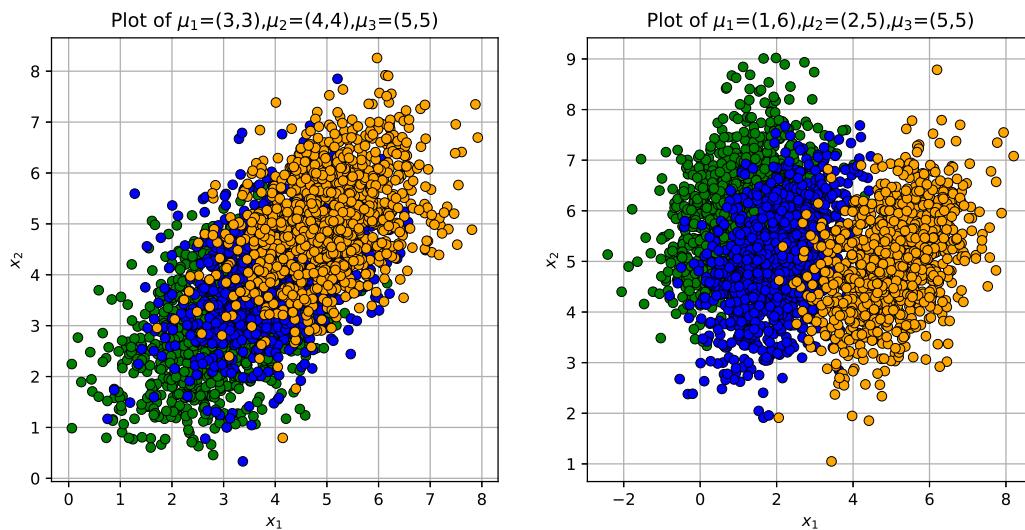


圖 5: 比較七種方法改變群平均實驗的三群組雙變量常態資料散布圖

表 9: 七種機器學習方法改變群平均的三群組訓練資料分類錯誤率

	Regression	Augmented	Logistic	LDA	QDA	KNN	ANN
$\mu_1 = (2, 2), \mu_2 = (4, 1), \mu_3 = (4, 5)$	0.481	0.493	0.408	0.405	0.401	0.288	0.404
$\mu_1 = (3, 3), \mu_2 = (4, 4), \mu_3 = (5, 5)$	0.608	0.598	0.512	0.515	0.505	0.384	0.516
$\mu_1 = (1, 6), \mu_2 = (2, 4), \mu_3 = (5, 5)$	0.477	0.486	0.397	0.398	0.395	0.283	0.397

根據表 9 與表 10 可知，KNN 模型是訓練資料中表現最好的模型，羅吉斯迴歸 (Logistic Regression) 則是測試誤差綜合最小的模型。

表 10: 七種機器學習方法改變群平均的三群組測試資料分類錯誤率

	Regression	Augmented	Logistic	LDA	QDA	KNN	ANN
$\mu_1 = (2, 2), \mu_2 = (4, 1), \mu_3 = (4, 5)$	0.477	0.489	0.413	0.415	0.388	0.437	0.415
$\mu_1 = (3, 3), \mu_2 = (4, 4), \mu_3 = (5, 5)$	0.609	0.599	0.514	0.521	0.543	0.565	0.522
$\mu_1 = (1, 6), \mu_2 = (2, 4), \mu_3 = (5, 5)$	0.478	0.488	0.404	0.412	0.422	0.438	0.410

3.3 三群組更改群共變異數的分類問題

最後，此章節透過改變共變異數來比較七種機器學習方法的訓練資料誤差與測試資料誤差，其使用的雙變量常態資料的參數如下，圖 6 則是這些資料的散布圖。

$$\mu_1 = \mu'_1 = \mu''_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \mu_2 = \mu'_2 = \mu''_2 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \mu_3 = \mu'_3 = \mu''_3 = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

$$\Sigma_1 = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix}$$

$$\Sigma'_1 = \begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix}, \Sigma'_2 = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix}, \Sigma'_3 = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$$

$$\Sigma''_1 = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}, \Sigma''_2 = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix}, \Sigma''_3 = \begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix}$$

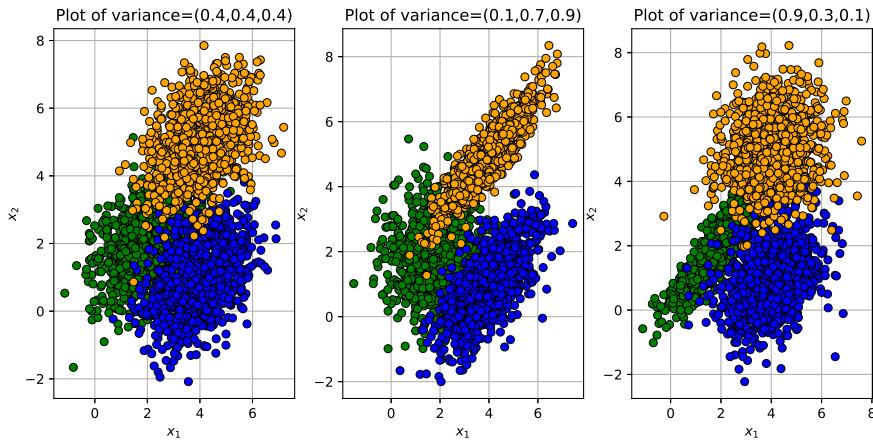


圖 6: 改變群共變異數實驗的三群組雙變量常態資料散布圖

表 11: 七種機器學習方法改變群共變異數的三群組訓練資料分類錯誤率

	Regression	Augmented	Logistic	LDA	QDA	KNN	ANN
$\sigma^2 = (0.2, 0.2, 0.2)$	0.481	0.493	0.408	0.405	0.401	0.288	0.404
$\sigma^2 = (0.4, 0.4, 0.4)$	0.032	0.033	0.031	0.032	0.031	0.027	0.032
$\sigma^2 = (0.1, 0.7, 0.9)$	0.447	0.442	0.369	0.371	0.368	0.242	0.367
$\sigma^2 = (0.9, 0.3, 0.1)$	0.450	0.469	0.371	0.372	0.370	0.255	0.371

根據表 11 與表 12 可觀察到七種方法的訓練誤差與測試誤差，其中無論共變異數是多少，KNN 模型都是訓練誤差最小的機器學習方法，而羅吉斯迴歸與 QDA 模型則是比較測試誤差時綜合表現較好的模型。

表 12: 七種機器學習方法改變群共變異數的三群組測試資料分類錯誤率

	Regression	Augmented	Logistic	LDA	QDA	KNN	ANN
$\sigma^2 = (0.2, 0.2, 0.2)$	0.477	0.489	0.413	0.415	0.388	0.437	0.415
$\sigma^2 = (0.4, 0.4, 0.4)$	0.031	0.032	0.031	0.031	0.037	0.04	0.031
$\sigma^2 = (0.1, 0.7, 0.9)$	0.448	0.443	0.376	0.383	0.371	0.378	0.38
$\sigma^2 = (0.9, 0.3, 0.1)$	0.453	0.473	0.377	0.380	0.363	0.393	0.380

4 結論

在此章節中依次重新簡述了前面章節所介紹到的七種機器學習方法，包含簡單線性迴歸 (Simple Linear Regression)、加廣迴歸 (Augmented Regression)、羅吉斯迴歸 (Logistic Regression)、線性判別分析 (LDA)、二次判別分析 (QDA)、K-近鄰演算法 (KNN) 與類神經網路 (ANN)，並就每種方法模擬生成兩群組與三群組雙變量常態資料建立分類模型並透過改變樣本數、平均數與共變異數來比較各自的訓練誤差與測試誤差，根據此章節的實驗可知，KNN 模型是所有方法中訓練誤差最小表現最好的模型，但是 QDA 模型才是其中測試誤差最低的模型。