

# 跨領域學習

## 機率在 AI 領域運用之基礎-簡易高斯辨識器(Gaussian Classifier)實作

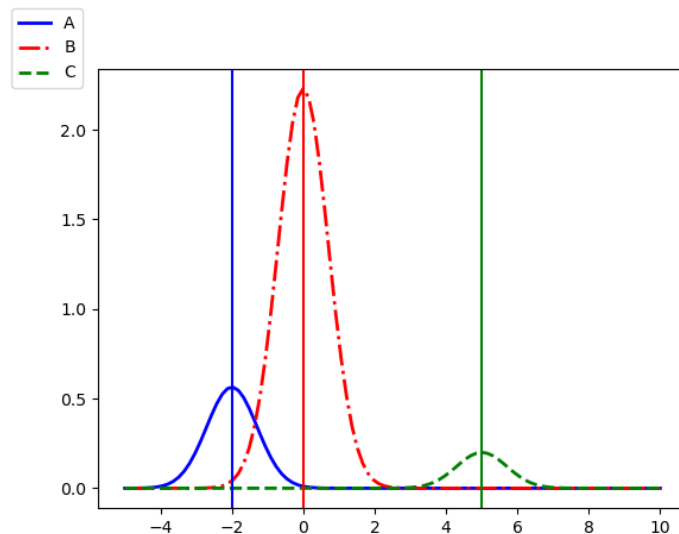
### 【1. 數學原理說明】

在 AI 大數據時代，利用電腦來進行協助大量資料分類已經是基本運用，而其中協助分類的分類器或是辨識器(classifier)設計就是核心之一。高斯辨識器(Gaussian classifier)利用常見的常態分布(normal distribution)的機率值大小來做為分類的依據。常態分布，又稱為「參數型分布」，也就說當我們假設資料是符合常態分布，那麼我們只要有平均數 $\mu$ 和標準差 $\sigma$ 就可以知道任何資料 $x$ 對應的機率值。當然，目前學到的是單變量的常態分布，如公式(1)；未來有機會就會看到多變量的機率分布，如公式(2)。

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

或

$$f_{\mathbf{x}}(x_1, \dots, x_k) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \mathbf{x}, \boldsymbol{\mu} \in \mathbb{R}^k, \Sigma \in \mathbb{R}^{k \times k} \quad (2)$$



由上圖中可以發現，因此若每一個分布代表一類，都有屬於自己的 $\mu$ 和 $\sigma$ ，黑色直線所指的位置的樣本(sample)  $x$ ，屬於綠色分布的機率最高，就會將該樣本判定給綠色的類別。因此，只要能計算該類別參數資料，就可以進行機器學習(machine learning)，利用機器直接來判定。

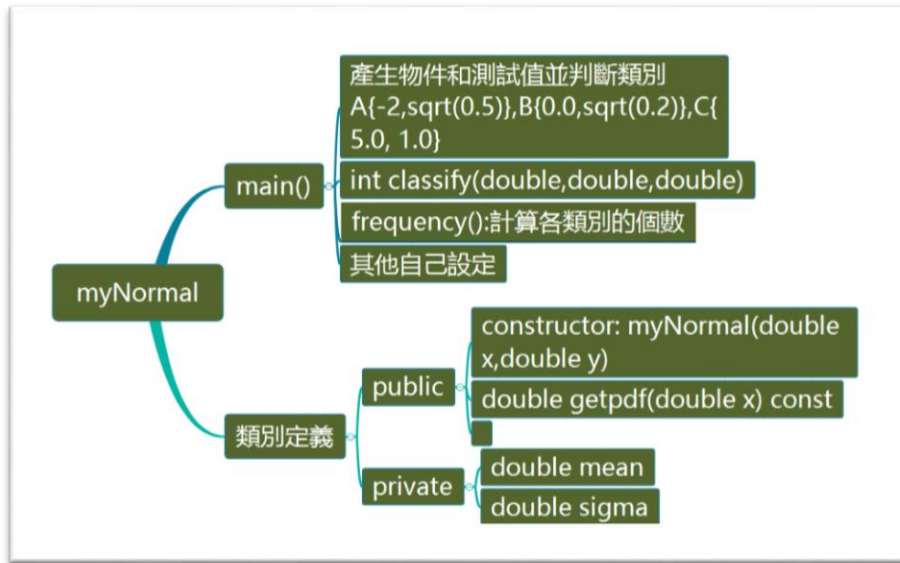
評分：依照寫的格式給予分數

(1)結果正確 60%。

(2)架構完備 40%。

### 【3. C++參考架構說明】

1. 設計 myNormal 類別(class)，資料結構如下圖。



2. 本程式設定有三類{1,2,3}的資料，**資料名稱：normaldata3.txt**，測試時請讀入該檔案有 100 個資料。(給分的測試檔案，我可能從新產生)
3. Main()主程式中有 **classify()** 的作用是**分類**，將某個樣本 $x$ 在每個類別的機率算出後，有該程式辨識所屬類別，傳回所屬**類別標籤(label)**，型別是 int。
4. **類別建構子**有兩個參數，是來設定物件的平均數 $\mu$ 和標準差 $\sigma$ ，**內建值**是 0 和 1。
5. 設定物件：**myNormal A{-2,sqrt(0.5)},B{0.0,sqrt(0.2)},C{5.0, 1.0}**；其中 A 是第 1 類，B 是第 2 類，C 是第 3 類
6. 由資料檔 testdata.txt 讀入測試樣本，並存到 array 或 vector，然後計算機率後，進行分類，並**統計各類樣本數**是多少。

```
G:\我的雲端硬碟\00授課\數教系必修\物件導向\normalClassifier\myNormalReadFileData\bin\Debug\myNormalReadFileData.exe
Test samples:.....
2.050 6.720 3.077 7.956 4.946 -0.271 -0.527 -2.409 4.939 5.478 -1.680 1.654
4.380 1.604 3.322 -0.706 4.563 -1.620 7.986 4.077 -2.180 4.340 -2.532 0.227
2.929 3.567 4.667 4.181 -1.682 2.583 -0.031 6.344 3.935 0.668 7.867 4.480
6.404 4.640 0.505 2.807 0.087 1.852 2.812 -1.830 -2.919 0.172 4.993 6.139
-1.717 5.221 1.669 -1.127 7.703 7.156 6.159 1.782 -1.123 7.023 5.734 7.226
5.616 -1.327 -1.013 5.160 7.146 4.826 -1.208 -1.099 4.666 4.074 -2.185 6.475
1.720 2.395 2.500 4.313 3.243 -1.779 -0.255 1.626 0.123 2.422 5.927 -0.699
7.348 3.065 6.101 5.762 -1.598 -0.786 6.620 -0.363 -2.438 1.431 3.522 7.574
-1.760 -1.131 3.424 -0.372

Classifying results:
3 3 3 2 2 1 3 3 1 3
3 2 3 1 3 3 1 3 1 2
3 3 3 1 3 2 3 2 3 3
3 3 2 3 3 3 1 2 3 3
1 3 3 1 3 3 3 1 3 3
3 1 1 3 3 3 1 3 3 1
3 3 3 3 3 1 2 3 3 3
3 3 3 3 1 2 3 2 3 3
1 1 3 2 2 3 2 1 2 3

Frequency .....
class 1      class 2      class 3
21          17          62
```

7. 函數的架構，包括參數、傳回值和實作請自行完成，也可以加上自己的函數。
8. 練習測試資料

2.05017  
6.72025  
3.07692  
7.95632  
4.94621  
-0.270658  
-0.526847  
-2.40911  
4.93878  
5.47762  
-1.67951  
1.65447  
4.37985  
1.60358  
3.32181  
-0.705515  
4.56317  
-1.61969  
7.98603  
4.07733  
-2.17972  
4.33998  
-2.53155  
0.226907  
2.92908  
3.56744  
4.66744  
4.18114  
-1.68174  
2.58285  
-0.0314115  
6.34355  
3.93527  
0.667972  
7.86726  
4.48046  
6.40443  
4.63964

0.50495  
2.80682  
0.0865199  
1.8516  
2.81223  
-1.83042  
-2.91879  
0.1719  
4.9933  
6.13859  
-1.71702  
5.22128  
1.66937  
-1.12701  
7.70309  
7.15644  
6.15878  
1.78233  
-1.1227  
7.02335  
5.73376  
7.22568  
5.61564  
-1.32746  
-1.01267  
5.15975  
7.14582  
4.8259  
-1.20791  
-1.09933  
4.66591  
4.07364  
-2.18529  
6.47461  
1.71988  
2.39514  
2.49998  
4.31319

3.24255  
-1.77901  
-0.255149  
1.62629  
0.122889  
2.42199  
5.92665  
-0.698869  
7.34818  
3.06546  
6.10116  
5.76151  
-1.59808  
-0.785982  
6.62035  
-0.363185  
-2.43802  
1.43139  
3.5221  
7.57416  
-1.76039  
-1.13111  
3.42425  
-0.372361