

圖形識別與機器學習 –手指第一指節

系級：電機碩一 學號：11278041 姓名：陳大荃

學校系所：中原大學 電機工程學系

1. 摘要

量取各 20 名男女受測者小拇指指節的直徑並進行數據分布的量測。

2. 引言

在這次實驗中記錄男女各 20 名共 40 名受測者小拇指第一指節的直徑分布。在得到數據後將資料分性別個別探討其資料的分布範圍及是否符合常態分佈。

3. 方法

實驗器材與對象：



圖一：游標卡尺



圖二：受測者手掌

實驗一：

預設環境：無

測試方法：在受測者小拇指第一指節最凸出處以游標卡尺量測直徑。(如圖三所示) 收集的資料集為表一。以切成 5 等分來分析(mm)。

表一：男女小拇指第一指節直徑資料集

資料編號	男(mm)	女(mm)
1	10.18	10.81
2	10.18	9.23
3	10.51	9.33

4	10.37	10.74
5	10.46	9.72
6	10.53	10.38
7	9.58	10.92
8	11.11	10.48
9	9.29	9.79
10	11.44	9.75
11	10.56	10.51
12	10.83	9.44
13	10.36	9.9
14	10.43	10.41
15	9.85	10.72
16	10.73	11.46
17	10.1	9.65
18	11.33	9.97
19	10.32	9.35
20	9.68	10.34



圖三：直徑量測示意圖

以下列出各種數據的計算方式 ([GitHub](#))：

1. Min (\min_x): $\min \{x_1, \dots, x_n\}$
2. Max (\max_x): $\max \{x_1, \dots, x_n\}$
3. Mean (μ): $\frac{1}{n} \sum_{i=1}^n x_i$
4. Variance (σ^2): $\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$
5. Standard Deviation (σ): $\sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}}$

6. Split Step (stp_x): $\max_x - \min_x$
7. Split Points: $\min_x + \frac{i}{5} \times stp_x$
8. Split Count: The number of data fall into each regions.

以下列出決策邊界的計算方式 ([GitHub](#)) :

1. Intersection of smaller dataset (x_1, y_1):
Check the probability of the smaller dataset from the left until it is greater than the larger dataset.
2. Intersection of larger dataset (x_2, y_2):
Check the probability of the larger dataset from the right until it is smaller than the smaller dataset.
3. Average x (\bar{x}): $\frac{x_1 + x_2}{2}$
4. Minimum error ($\varepsilon_1 + \varepsilon_2$):

$$\int_{R_{2min}}^{\bar{x}} P[x|w_1]dx + \int_{\bar{x}}^{R_{1max}} P[x|w_2]dx$$

結果 1: (男性)

表 二：男性資料集分析結果(詳情請見 [GitHub](#))

Min	9.29				
Max	11.44				
Mean (μ)	10.392				
Variance (σ^2)	0.286166				
Standard Deviation (σ)	0.534945				
Split Step	0.43				
Split Points	9.29 / 9.72 / 10.15 / 10.58 / 11.01 / 11.44				
Split Count	3	2	10	2	3

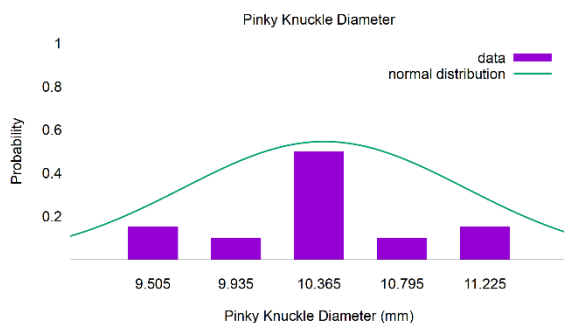


圖 四：男性資料五等份分布([GitHub](#))

結果 2: (女性)

表 三：女性資料集分析結果(詳情請見 [GitHub](#))

Min	9.23				
Max	11.46				
Mean (μ)	10.145				
Variance (σ^2)	0.364725				
Standard Deviation (σ)	0.603925				
Split Step	0.446				
Split Points	9.23 / 9.676 / 10.122 / 10.568 / 11.014 / 11.46				
Split Count	5	5	5	4	1

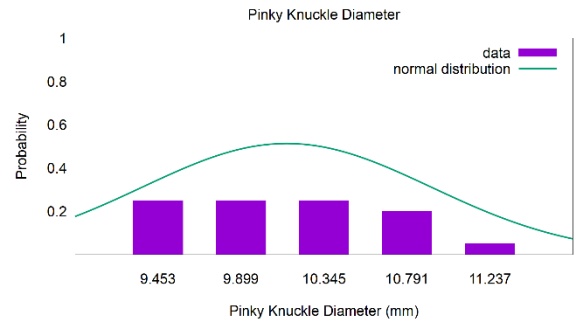
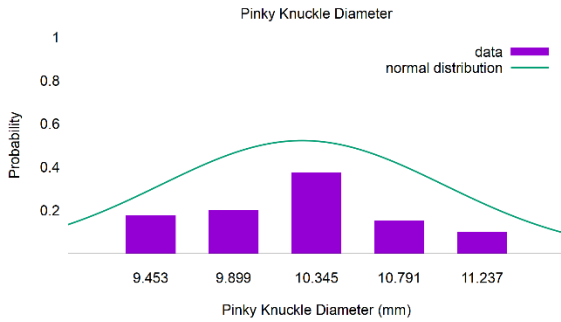


圖 五：女性資料五等份分布([GitHub](#))

結果 3: (男女混合)

表 四：男女混合資料集分析結果(詳情請見 [GitHub](#))

Min	9.23				
Max	11.46				
Mean (μ)	10.2685				
Variance (σ^2)	0.340698				
Standard Deviation (σ)	0.583693				
Split Step	0.446				
Split Points	9.23 / 9.676 / 10.122 / 10.568 / 11.014 / 11.46				
Split Count	7	8	15	6	4

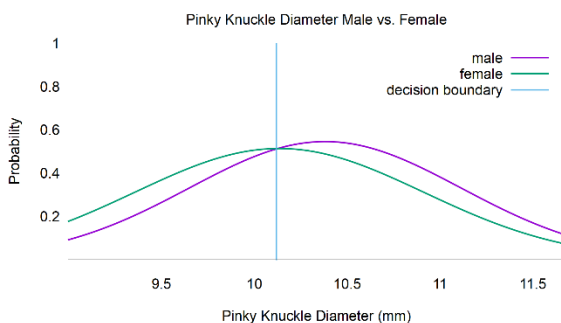


圖六：男女混合資料五等份分布([GitHub](#))

結果 4: (Decision Boundary)

表五：男女混合資料集決策邊界(詳情請見 [GitHub](#))

Intersection of smaller dataset (x_1, y_1)	10.128107	0.660324
Intersection of larger dataset (x_2, y_2)	10.128106	0.660324
Average x (\bar{x})	10.128107	
Minimum error ($\epsilon_1 + \epsilon_2$)	0.844838	



圖七：Decision Boundary ([GitHub](#))

以程式碼進行的分析已完成，請至

https://github.com/belongtothenight/PRML_Code/tree/main/src/hw2 依照 readme.md 執行即可得知其餘兩個結果的分析。

4. 結果

可以大略的從每個區域的數據量呈現如常態

分佈般的鐘形分佈。

5. 討論

若是有更多筆資料或許就能夠呈現如完美的常態分佈一般的分布圖。

在上週打程式的過程中發現若資料的 Standard Deviation 若是小於 0.5，即有很大的可能性機率分布會超過 1，這點有經過多方查證(以公式計算、Matlab[己]、GNUPlot[戊]、圖形產生器[戊]、Python[庚])。這個問題後來發現是公式變數的解讀錯誤，需先開根號再進行計算，已經完成程式修正。(請見 [GitHub commit 6c4c581](#))

6. 結論

可以從自然界中的尺寸分布觀察到常態分佈。而計算 Decision Boundary 若要以最好的效果(minimize minimum error)區分兩個群體，則資料的分布(mean)離得越開越好。

7. 參考文獻

- 甲、<https://www.calculatorsoup.com/calculators/statistics/mean-median-mode.php>
- 乙、<https://www.calculatorsoup.com/calculators/statistics/statistics.php>
- 丙、<https://www.calculatorsoup.com/calculators/statistics/descriptivestatistics.php>
- 丁、https://homepage.divms.uiowa.edu/~mbo_gnar/applets/normal.html
- 戊、<https://gist.github.com/JavierJia/e10f7c7d8ce3c42a6e469316a5738f0f>
- 己、<https://www.mathworks.com/help/stats/normal-distribution.html>
- 庚、<https://stackoverflow.com/questions/20011122/fitting-a-normal-distribution-to-1d-data>