

# Programming Assignment 6

---

## 1. 兩個 cell 分別的主要內容

**Cell 1** — 沿用 Assignment 4 的資料解析

**Cell 2** — 分類模型 : Gaussian Discriminant Analysis (GDA)

- 前處理 : (lon, lat) 標準化 (StandardScaler)
- 切分 : 60% / 20% / 20%

**1 理論 :**

假設資料各類樣本皆服從高斯分布且共用共變異矩陣  $\Sigma$  :

$$P(x|y = k) = N(x; \mu_k, \Sigma), \quad k \in \{0, 1\}$$

參數由最大概似估計 (MLE) 求得 :

$$\phi = \frac{1}{m} \sum_i y^{(i)}, \quad \mu_k = \frac{\sum_i 1_{y^{(i)=k}} x^{(i)}}{\sum_i 1_{y^{(i)=k}}}, \quad \Sigma = \frac{1}{m} \sum_i (x^{(i)} - \mu_{y^{(i)}})(x^{(i)} - \mu_{y^{(i)}})^T$$

- 條件機率 :

$$P(y = 1|x) = \frac{\phi P(x|y = 1)}{\phi P(x|y = 1) + (1 - \phi) P(x|y = 0)}$$

決策邊界為  $P(y = 1|x) = 0.5$

- 實作 :

```
phi, mu0, mu1, Sigma = fit_gda(X_train, y_train)
proba_test = predict_proba(X_test, phi, mu0, mu1, Sigma)
y_pred = (proba_test >= 0.5).astype(int)

acc = accuracy_score(y_test, y_pred)
cm = confusion_matrix(y_test, y_pred)
print(f"Accuracy: {acc:.4f}")
print("Confusion Matrix:\n", cm)
```

## 2 Piecewise Function

- 定義 :

$$h(x) = \begin{cases} R(x), & C(x) = 1 \\ -999, & C(x) = 0 \end{cases}$$

- 其中：

- $C(x)$ ：GDA 分類模型
- $R(x)$ ：(HW4) 神經網路迴歸模型

- 實作：

```
# === Piecewise Function 結合邏輯 ===
if C(x) == 1:
    output = R(x)          # 有效格點 → 輸出迴歸預測值
else:
    output = -999          # 無效格點 → 輸出固定 -999
```

### 3 輸出：

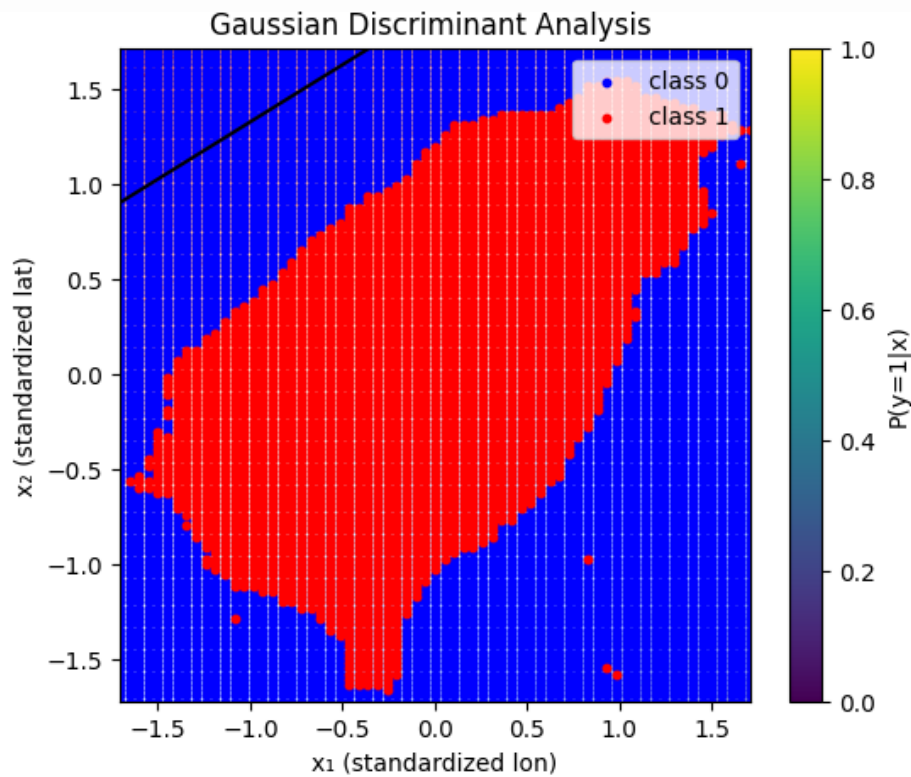
- Accuracy、Confusion Matrix
- Gaussian Discriminant Analysis圖檔

- 數據結果：

```
Accuracy: 0.5158
Confusion Matrix:
[[1244  119]
 [1049   0]]
```

GDA 幾乎將所有樣本預測為 class 0，造成 Recall = 0；  
然而，紅色區域形狀與台灣主島地理輪廓高度吻合，  
顯示模型成功捕捉資料的地理分群特徵

- 視覺化決策邊界圖說：



- 藍色：class 0（無效格點）
- 紅色：class 1（有效格點）
- 背景顏色：模型估計的 (  $P(y=1|x)$  ) 機率分布
- 黑線：(  $P(y=1|x)=0.5$  ) 決策邊界（理論上應為線性）

### 3. 結果與討論

項目	HW4 NN 分類	HW6 GDA 分類	分析
模型類型	判別式 (Discriminative)	生成式 (Generative)	不同方式建模
分布假設	無	高斯且共變異矩陣相同	GDA 假設較強
準確率	約 0.87	約 0.52	NN 明顯優於 GDA
邊界型態	非線性	線性	符合理論 (LDA)
可視化結果	機率分布平滑	幾乎全藍	GDA 未能擬合非線性區域
計算速度	需多輪訓練	解析式即時	GDA 無迭代，速度快

## 4. 分析重點

- **優點**：GDA 理論簡潔、計算快速、封閉解可解析
- **缺點**：假設過強，分類結果偏向多數類別
- **優勢**：雖然準確率僅約 **0.52**，但輸出圖清楚勾勒出台灣主島形狀
  - GDA：理論可解析、速度快；
  - NN：擬合能力強、精度高。