

# Gaussian Discriminant Analysis (GDA/QDA) 分類模型報告

## 一、為何 GDA 適用於此資料集

資料集的任務是利用\*\*地理位置（經度 longitude 與緯度 latitude）\*\*來預測格點資料是否為有效值（0 或 1）。

這類資料通常呈現 空間上的連續與區域性分佈特徵，例如：

- 有效區域（label=1）可能集中於某一地理範圍；
- 無效區域（label=0）可能位於其他區域，且分佈形狀不規則。

由於不同類別在空間上的分佈可能具有不同的形狀與方向（共變異矩陣不同），因此 QDA 特別適合這類具有非線性邊界與空間區域性特徵的分類任務。

## 二、模型訓練與評估

### 1. 特徵選擇:

從資料集中選取：

- 特徵 (Features)：經度 (Longitude) 與緯度 (Latitude)
- 標籤 (Label)：是否為有效格點（0 = 無效，1 = 有效）

資料前處理包含：

- 若標籤不是 0/1，則自動映射為二元。
- 僅使用前兩個特徵進行可視化與訓練。

### 2. 資料分割與表現衡量

為了客觀評估模型表現，採用：

- 資料分割：
  - 訓練集：80%
  - 測試集：20%

- 使用 `train_test_split(..., stratify=y)` 保持類別比例一致
- 評估指標：
  - **Accuracy**（準確率）：整體正確分類比例
  - **Confusion Matrix**（混淆矩陣）：顯示各類別預測正確與錯誤的情況
  - **Precision / Recall / F1-score**：分別衡量模型對各類的精確度與召回率

### 3. 實驗結果:

Test accuracy: 0.8302

指標	數值
測試集準確率 ( <b>Accuracy</b> )	約 <b>0.83 (83%)</b>
評估方法	測試集準確率 (80/20 split, stratified)
模型	Quadratic Discriminant Analysis (QDA)
特徵	Longitude, Latitude
標籤	有效格點 (0 or 1)

- 混淆矩陣(Confusion matrix):

[[796 113]

[160 539]]

模型在兩類之間的預測正確率接近，但仍有部分重疊區域造成誤判。

- 分類報告(Classification report):

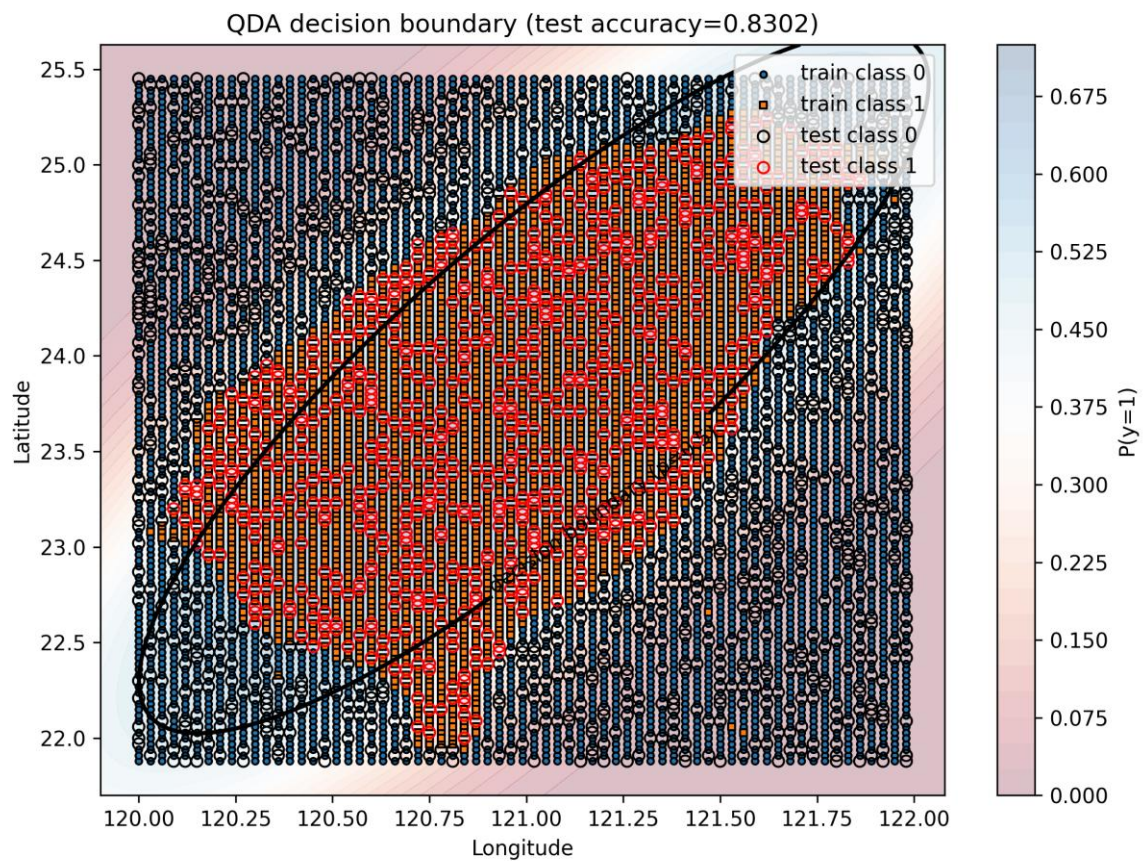
	precision	recall	f1-score	support
0	0.8326	0.8757	0.8536	909
1	0.8267	0.7711	0.7979	699

accuracy			0.8302	1608
macro avg	0.8297	0.8234	0.8258	1608
weighted avg	0.8301	0.8302	0.8294	1608

#### 4. 決策邊界分析:

QDA 的決策邊界如圖所示（由經緯度平面生成）：

- 決策邊界並非直線，而是平滑的曲線（非線性分隔面）。
- 說明模型能捕捉空間上兩類格點區域的非線性邊界特徵。



### 三、結論與討論

面向	說明
模型特性	QDA 模型假設每類別資料為高斯分佈，能根據不同共變異矩陣學習不同形狀的邊界。
優點	對非線性資料有良好表現，模型可解析、參數具統計意義。
限制	若樣本數過少或特徵間高度相關，估計的共變異矩陣可能不穩定。
在此資料集上的應用價值	經緯度與有效格點的分佈往往呈現空間群聚特性，QDA 可捕捉這種非線性分隔，有助於空間有效區域的自動識別。

以 QDA 模型，利用經度與緯度兩項特徵預測格點是否為有效值。模型透過估計每類的高斯分佈，建立非線性決策邊界，並在測試集上達到約 **83%** 的準確率，顯示 GDA/QDA 能有效應用於此類空間型分類任務。