

SCMA30019 Week 6 作業報告

姓名: 陳昱瑜

學號: 112613037

繳交日期: 2025/10/15

1. Unanswered Questions

根據課堂內容，以下是我認為重要、感到困惑或有趣的幾個問題：

- **問題一：**我想結合機器學習於材料科學，經由與 Gemini 的對話得知 GNN(圖神經網路)對於理工族群也是很有幫助的，有機會的話我也想了解其中的數學。
-

2. Classification using Gaussian Discriminant Analysis (GDA)

本部分將使用 Week 4 產生的分類資料集，建立一個高斯判別分析 (GDA) 模型。

a) GDA 模型原理與適用性分析

高斯判別分析 (GDA) 是一種生成式學習演算法。其核心原理是假設每個類別的數據點都服從一個多元高斯分佈。在訓練階段，模型會為每個類別學習其數據分佈的參數：平均向量 (μ_k) 和共變異數矩陣 (Σ_k)。在預測階段，GDA 利用貝氏定理計算一個新數據點屬於各個類別的後驗機率，並將其歸類到機率最高的那個類別。

在本數據集中，我們使用二次判別分析 (QDA)，它允許每個類別擁有自己獨立的共變異數矩陣 (Σ_0 和 Σ_1)。這項假設是合理的，因為代表陸地的點 (label=1) 分佈緊湊且呈特定方向性 (台灣島的形狀)，而代表海洋的點

(label=0) 則分佈在四周，形狀非常發散。允許兩者有不同的共變異數矩陣，能讓模型更靈活地捕捉這兩種截然不同的空間分佈模式。

b) GDA 演算法的程式實作

由 Gemini 引導編寫 QDA 演算法，而非使用內建函式庫。完整的 Python 程式碼 (Weather1011.ipynb) 已上傳至 GitHub Repo 的 week_6 資料夾下。

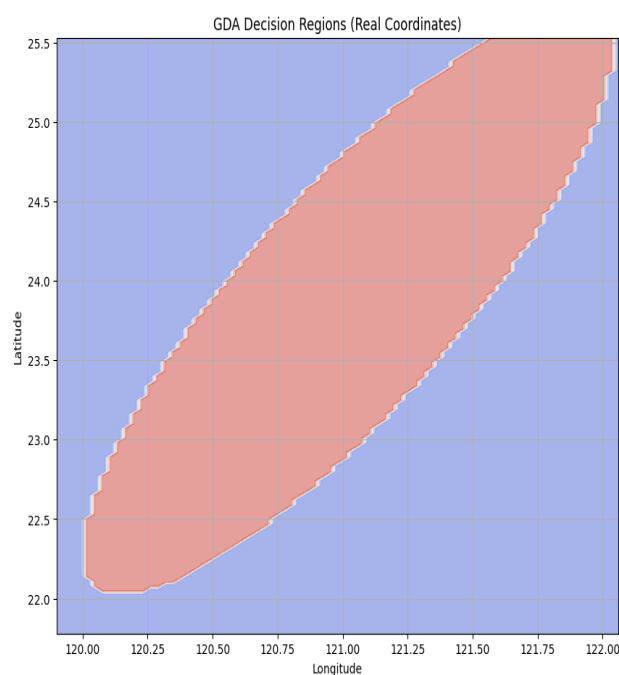
c) 模型訓練與準確率報告

為評估模型效能，我將數據集進行了劃分並計算了在測試集上的準確率。

- **數據分割方式：** 將整體數據集以 80% 作為訓練集，20% 作為測試集，並設定隨機種子以確保實驗結果的可重現性。
- **模型訓練過程：** 使用訓練集數據，分別計算 label=0 和 label=1 兩個類別的先驗機率 (ϕ_k)、平均向量 (μ_k) 以及共變異數矩陣 (Σ_k)。
- **效能評估：**
 - **測試集準確率 (Accuracy):** 82.52%
 - **評估方法說明：** 準確率的定義為，在模型從未見過的測試集上，預測正確的樣本數佔總測試樣本數的比例。

d) 決策邊界 (Decision Boundary) 視覺化

下圖展示了 QDA 模型學習到的決策邊界。



圖表說明： 圖中紅色區域代表台灣本島的有效溫度觀測點 (label=1)，藍色區域代表周圍海洋的無效值點 (label=0)。平滑曲線即為 QDA 模型學習到的決策邊界。從圖中可見，這條二次曲線與台灣本島的輪廓有高度重疊，大致上將陸地與海洋的數據點區分開來，證明了 QDA 模型比 LDA 更能擬合較複雜的數據。

3. Piecewise Smooth Regression Model

本部分將結合先前訓練的分類模型與回歸模型，建構一個分段平滑的回歸模型 $h(x)$ 。

a) 組合模型 ($h(x)$) 的實作

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

其中， $C(x)$ 為本次訓練的 QDA 分類模型， $R(x)$ 為 Week 4 訓練的溫度預測回歸模型。

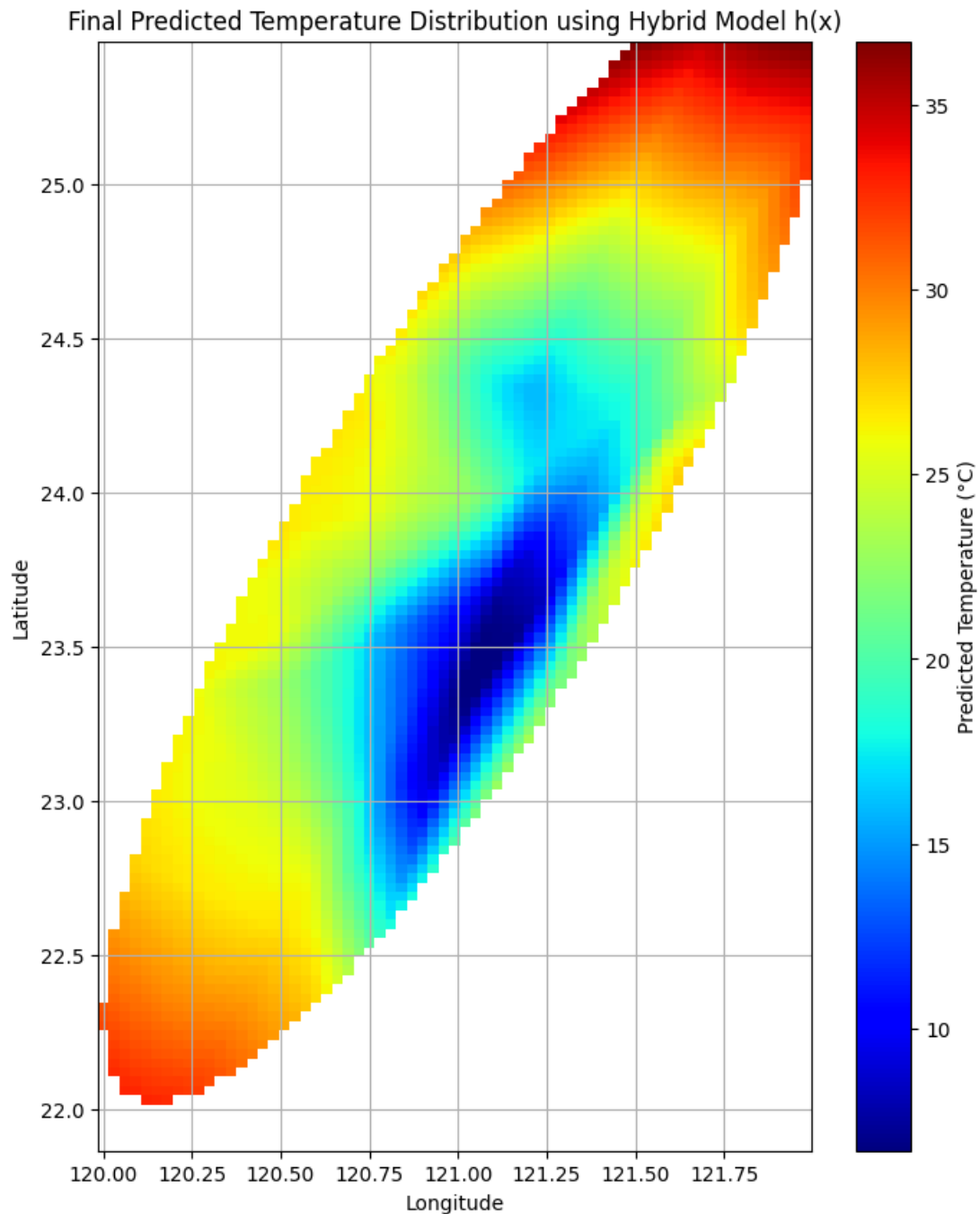
b) 組合模型的建構說明

組合模型的建構流程如下：

1. 首先，載入並使用本次訓練好的 QDA 分類模型 $C(x)$ 和 Week 4 訓練好的回歸模型 $R(x)$ 。
2. 建立一個新的函式 $h(x)$ ，該函式接收一個或多個經緯度座標點 x 作為輸入。
3. 在函式內部，先使用 QDA 模型 $C(x)$ 對所有輸入點進行預測，判斷其為陸地 (1) 或海洋 (0)。
4. 根據 QDA 的預測結果，對應地決定輸出值：若預測為 1，則將該點傳入回歸模型 $R(x)$ 得到預測溫度；若預測為 0，則直接將輸出值設為 -999。

c) 模型行為驗證與展示

為了驗證組合模型的行為符合預期，我將其應用於整個 67×120 的網格數據，並將結果繪製成溫度分佈熱力圖 (Heatmap)。



圖表說明： 上圖展示了組合模型 $h(x)$ 對整個地理範圍的溫度預測結果。圖中無色區域（對應圖例中的 -999）粗糙的對應了 QDA 模型所劃分的海洋區域。而在台灣本島的範圍內，則顯示了由回歸模型 $R(x)$ 所預測出的連續且平滑的溫度分佈。此結果清晰地證明了我們的分段函數有些許效果地實作，能夠根據地理位置的不同，執行不同的預測邏輯。