

Report of Programming assignment

模型與說明

本專案針對氣象資料觀測平台中「溫度分布-小時溫度觀測分析格點資料」進行機器學習建模。資料每個格點含有溫度觀測值($^{\circ}\text{C}$)，經緯度解析度均為 0.03 度，格點布局為 67×120 的網格。資料中無效值標為 -999。

構建兩種模型：

- 分類模型：利用經緯度特徵，預測該格點是否為有效資料(標記 0 或 1)。模型採用兩層含 ReLU 激活函數及 dropout 的神經網路結構，優化目標為交叉熵(Cross Entropy)損失，使用 Adam 優化器。
- 回歸模型：同樣利用經緯度特徵，預測該格點的溫度值。模型同樣為類似結構的神經網路，最後輸出一個實值，目標為均方誤差(MSE)損失。

分類模型採用 Cross Entropy Loss 的理由：

交叉熵損失透過計算真實標籤概率分佈與預測概率分佈間的信息差異，促使模型將預測概率朝向真實類別的方向靠近。當預測概率越接近真實標籤，交叉熵損失值就越小，模型表現越好。

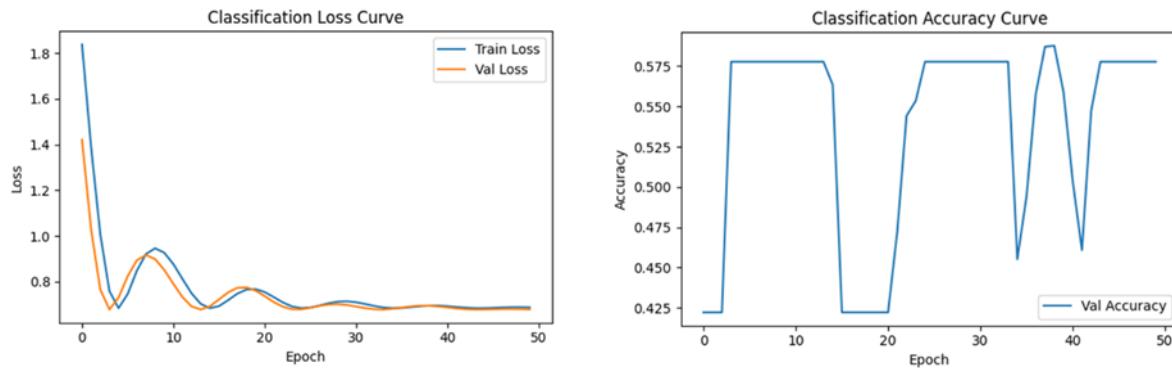
因此，對於分類模型中判斷格點資料是否有效的任務(即二分類問題)，使用交叉熵損失能有效反映預測與真實標籤的不一致性，並引導模型透過反向傳播優化參數，提升分類準確度。

訓練過程與結果分析

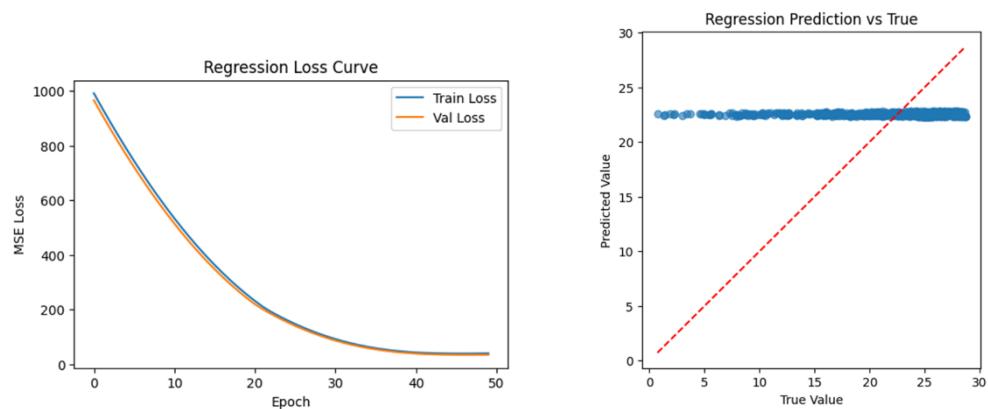
- 前期嘗試

首次嘗試直接使用原始未經標準化的資料，結果表現不佳：

- 分類模型損失下降不穩定，驗證準確率波動較大，決策邊界不清晰，無法有效區分有效與無效資料格點。



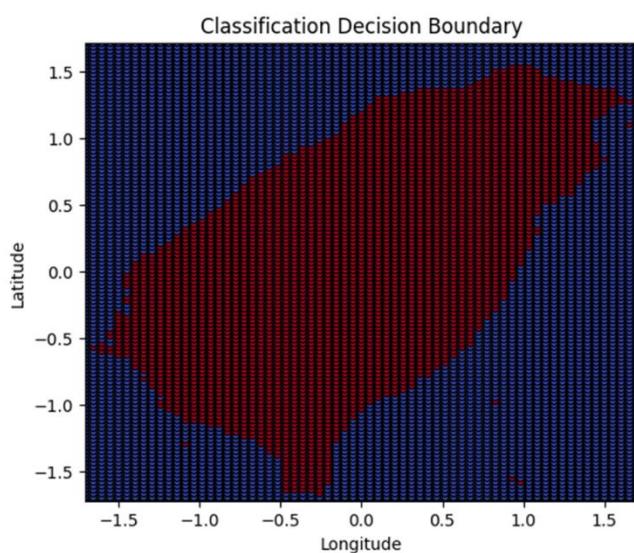
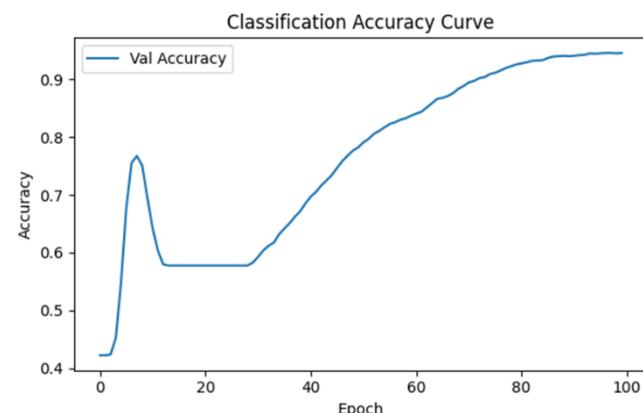
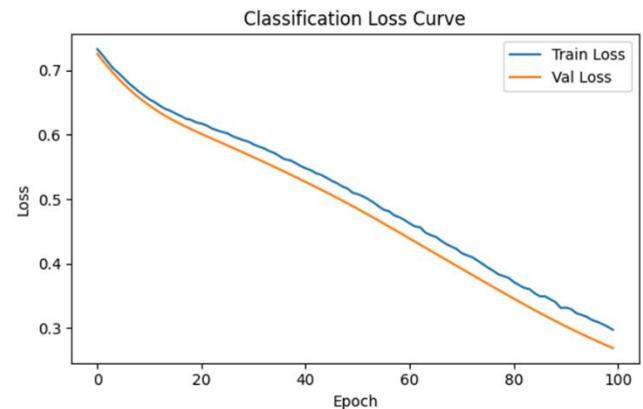
- 回歸模型預測結果集中於單一值，難以擬合真實溫度變化，預測誤差過大，模型無法捕捉溫度的空間分布特徵。



- 標準化後模型訓練

將經緯度資料標準化後，模型訓練效果顯著改善：

- 分類模型經過 100 次迭代，損失和驗證準確率曲線均穩定收斂，分辨力強，決策邊界與臺灣地區有效觀測範圍高度吻合。



- 回歸模型訓練 1000 次迭代後，損失曲線趨於平緩，預測值與真實值散點呈現明顯線性關係，能有效擬合空間溫度分布，預測誤差控制在合理範圍。

