111652015曹晉嘉 – 機器學習hw4 (程式報告 + Written assignment)

1. **模型 :**

**甲、資料處理**

* + 1. 讀取 XML 資料，將資料展平成一維陣列，再依照經度、緯度重塑為二維網格
    2. 建立兩種資料集：
       1. 分類資料集：判斷該格點是否有效（0為無效，1為有效），共 8040 筆資料
       2. 回歸資料集：僅保留有效格點，輸出為溫度數值，共 3495 筆資料
    3. 將經緯度座標進行標準化（縮放到 [-1,1]），以利神經網路訓練
  1. 模型設計
     1. 分類模型
        1. 結構 : 稍加修改並使用第二次作業的神經網路 ( 兩層隱藏層的神經元都是10個 )
        2. 損失函數 : 使用Weighted BCE，並計算pos\_weight以避免有類別不平衡
        3. 特殊處理 : 將輸出的值放入sigmoid函數，使輸出在0到1之間
     2. 回歸模型
        1. 結構 : 稍加修改並使用第二次作業的神經網路 ( 兩層隱藏層的神經元都是10個 )
        2. 損失函數 : 使用均方誤差 (MSE)

1. 訓練過程
   1. 使用參數
      1. 分類模型 : batch size = 16、訓練300輪、學習率0.01
      2. 回歸模型 : batch size = 4 、訓練800輪、學習率0.005
   2. 資料切分比例
      1. 訓練 : 驗證 : 測試 = 6 : 2 : 2
2. 結果
   1. 分類模型
      1. 測試集Accuracy ≈ 97.8%
      2. 混淆矩陣顯示模型能預測大部分資料

一張含有 文字, 字型, 螢幕擷取畫面, 設計 的圖片

AI 產生的內容可能不正確。 一張含有 文字, 螢幕擷取畫面, 繪圖, 行 的圖片

AI 產生的內容可能不正確。

一張含有 文字, 螢幕擷取畫面, 繪圖, 行 的圖片

AI 產生的內容可能不正確。

* 1. 回歸模型
     1. 測試集RMSE ≈ 2.87
     2. 測試集R² ≈ 0.786
     3. True vs Pred圖片顯示模型能捕捉主要趨勢但仍有部分誤差
     4. 訓練曲線顯示 : 損失函數收斂但驗證集存在一定的波動

一張含有 文字, 字型, 螢幕擷取畫面, 數字 的圖片

AI 產生的內容可能不正確。

一張含有 文字, 繪圖, 螢幕擷取畫面, 行 的圖片

AI 產生的內容可能不正確。

一張含有 行, 文字, 繪圖, 圖表 的圖片

AI 產生的內容可能不正確。

* 1. 資料分布
     1. 畫出溫度分布，顯示資料在經緯度的樣子
     2. 圖上白色區域代表無效格點

一張含有 文字, 螢幕擷取畫面, 鮮豔, 圖表 的圖片

AI 產生的內容可能不正確。

1. 結論 :
   1. 分類模型非常成功
   2. 回歸模型改進方向
      1. 嘗試用不同的激活函數
      2. 增加隱藏層數量或神經元個數
      3. 增加資料量
      4. 增加迭代次數
2. Written assignment
   1. 如何選擇適合的隱藏層和神經元數量 ?