Assignment 4

31652021羅瑋翔

Written Assignment

Why Logistic Regression does not use MSE, what would happen if we use it?

Program Assignment

Classification Model

資料處理

(1) 使用套件 'xml.etree.ElementTree' 解析xml檔案，提取<content>中的溫度。

(2) 生成經緯度網格(67*120)，再依據規定生成資料。

(3) 如果溫度是有效值→將label設置成1；無效值→將label設置成0

(4) 最後合併網格和label→得到(經度, 緯度, label)格式的資料, 並將資料以CSV檔儲存

切割Dataset

Training: 70%

Validation: 15%

Testing: 15%

模型

Input layer: 2 neurons

2 hidden layers with 16 and 32 neurons, respectively

Output layer: 1 neuron

Activation function: Use ReLU in the hidden layers and Sigmoid in the output layer

(因為分類的邊界通常不是直線，而是曲線或是複雜的區域→**非線性決策邊界**

引進ReLU產生非線性的結果，最後使用Sigmoid求0-1的機率。)

Loss Function: BCE Loss

Optimizer: Adam

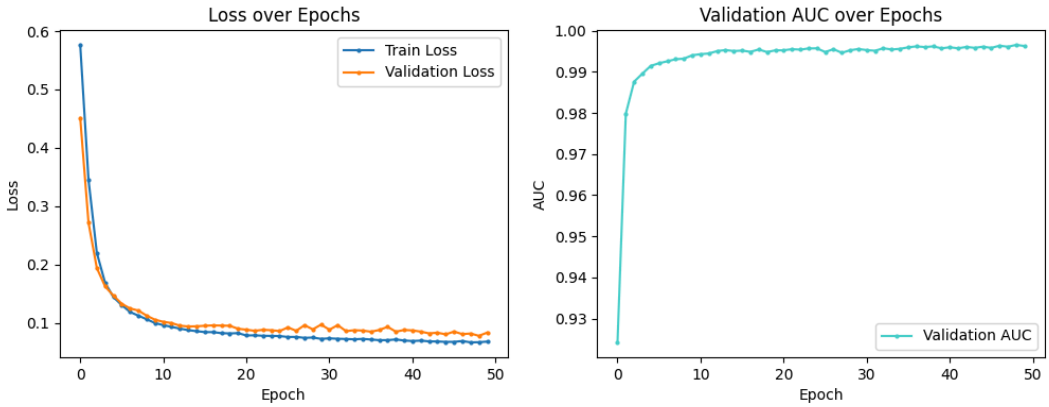

Figure 1. Left-hand side is the graph of Loss over epochs and the right-hand is the graph of validation AUC over epochs

Epoch 26/50 - Train Loss: 0.0741	Val Loss: 0.0852	Val AUC: 0.9956
Epoch 27/50 - Train Loss: 0.0736	Val Loss: 0.0801	Val AUC: 0.9962
Epoch 28/50 - Train Loss: 0.0730	Val Loss: 0.0801	Val AUC: 0.9963
Epoch 29/50 - Train Loss: 0.0729	Val Loss: 0.0891	Val AUC: 0.9951
Epoch 30/50 - Train Loss: 0.0721	Val Loss: 0.0840	Val AUC: 0.9963
Epoch 31/50 - Train Loss: 0.0718	Val Loss: 0.0792	Val AUC: 0.9963
Epoch 32/50 - Train Loss: 0.0703	Val Loss: 0.0802	Val AUC: 0.9962
Epoch 33/50 - Train Loss: 0.0705	Val Loss: 0.0775	Val AUC: 0.9965
Epoch 34/50 - Train Loss: 0.0697	Val Loss: 0.0838	Val AUC: 0.9961
Epoch 35/50 - Train Loss: 0.0692	Val Loss: 0.0764	Val AUC: 0.9966
Epoch 36/50 - Train Loss: 0.0684	Val Loss: 0.0752	Val AUC: 0.9967
Epoch 37/50 - Train Loss: 0.0684	Val Loss: 0.0794	Val AUC: 0.9962
Epoch 38/50 - Train Loss: 0.0684	Val Loss: 0.0814	Val AUC: 0.9960
Epoch 39/50 - Train Loss: 0.0674	Val Loss: 0.0767	Val AUC: 0.9965
Epoch 40/50 - Train Loss: 0.0672	Val Loss: 0.0823	Val AUC: 0.9959
Epoch 41/50 - Train Loss: 0.0665	Val Loss: 0.0821	Val AUC: 0.9968
Epoch 42/50 - Train Loss: 0.0671	Val Loss: 0.0806	Val AUC: 0.9965
Epoch 43/50 - Train Loss: 0.0655	Val Loss: 0.0798	Val AUC: 0.9965
Epoch 44/50 - Train Loss: 0.0665	Val Loss: 0.0793	Val AUC: 0.9968
Epoch 45/50 - Train Loss: 0.0648	Val Loss: 0.0763	Val AUC: 0.9967
Epoch 46/50 - Train Loss: 0.0659	Val Loss: 0.0768	Val AUC: 0.9968
Epoch 47/50 - Train Loss: 0.0643	Val Loss: 0.0714	Val AUC: 0.9970
Epoch 48/50 - Train Loss: 0.0656	Val Loss: 0.0719	Val AUC: 0.9968
Epoch 49/50 - Train Loss: 0.0633	Val Loss: 0.0784	Val AUC: 0.9965
Epoch 50/50 - Train Loss: 0.0643	Val Loss: 0.0744	Val AUC: 0.9967
---- Test Set AUC: 0.9978 ----		

Figure 2. The table of Train Loss, validation Loss and Validation AUC over Epochs, and Test AUC is manifest at the bottom

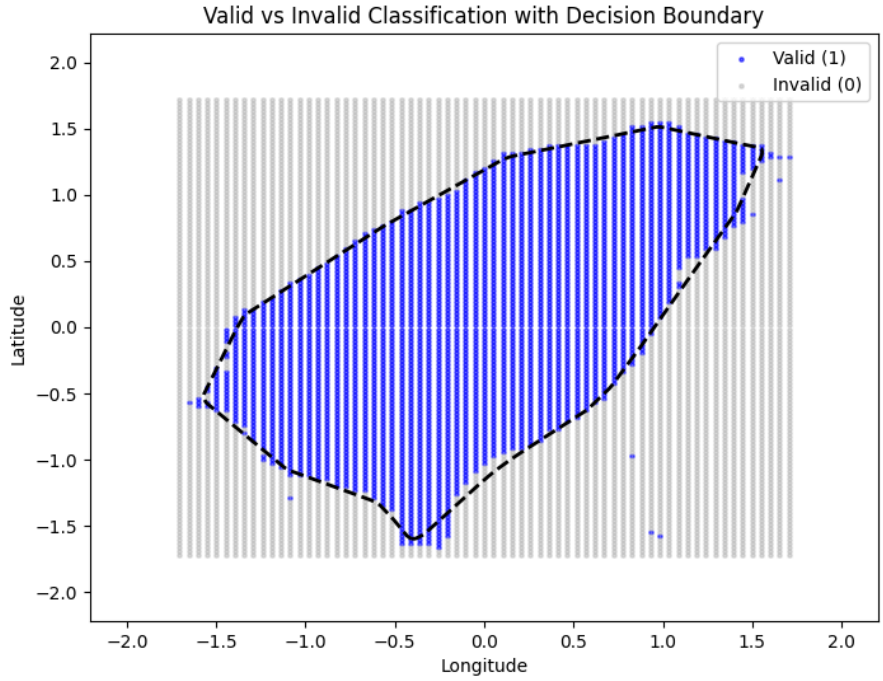

Figure 3. The predicted boundary of valid points vs. invalid points

Summary

By figure 1, we can see that the model coverges well.
And by figure 3, we can see the decition boundary looks like the shape of Taiwan, the prediction of boundary looks well, but on some place which has bigger gradient,
the model performs worse.

Regression Model

資料處理

(1) 使用套件 'xml.etree.ElementTree' 解析xml檔案，提取<content>中的溫度。
(2) 生成經緯度網格(67*120)，再依據規定生成資料，取除invalid資料後，將value設為溫度的key
(3) 最後合併網格和values→得到(經度, 緯度, value)格式的資料, 並將資料以CSV檔儲存

切割Dataset

Training: 70%
Validation: 15%
Testing: 15%

模型

Input layer: 2 neurons
2 hidden layers with 16 and 16 neurons, respectively
Output layer: 1 neuron
Activation function: Use ReLU in the hidden layers
(因為分類的邊界通常不是直線，而是曲線或是複雜的區域→**非線性決策邊界**
引進ReLU產生非線性的結果)
Loss Function: MSE Loss
Optimizer: Adam

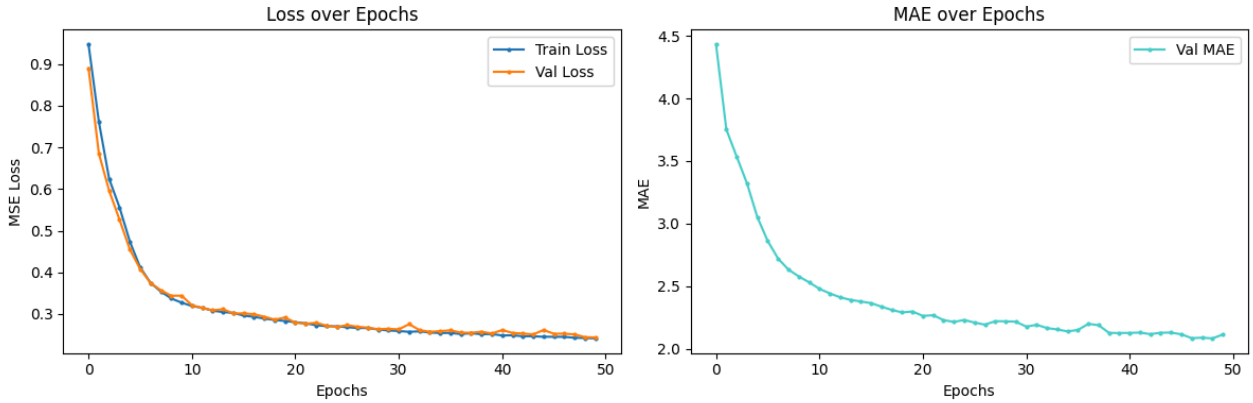

Figure 4. Left-hand side is loss over epochs graph and right-hand side is mean absolute error over epochs

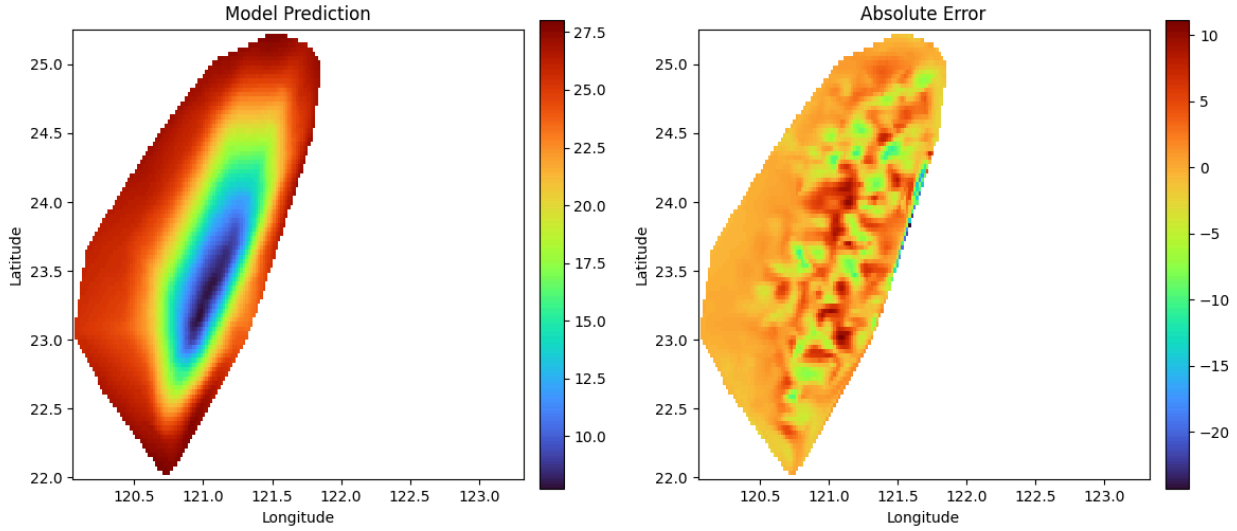

Figure 5. The above graph shows the prediction of the temprature and the absolute error over each data point

Summary

Though, from figure 4, we can see that prediction converge, but the MAE is higher than 2, it seems not like a good prediction.

I think the main reason why this phenomenon occurs is, we do not have enough big size of dataset. Hence, this model cannot learn the relationship between longitude/latitude and temperature well.