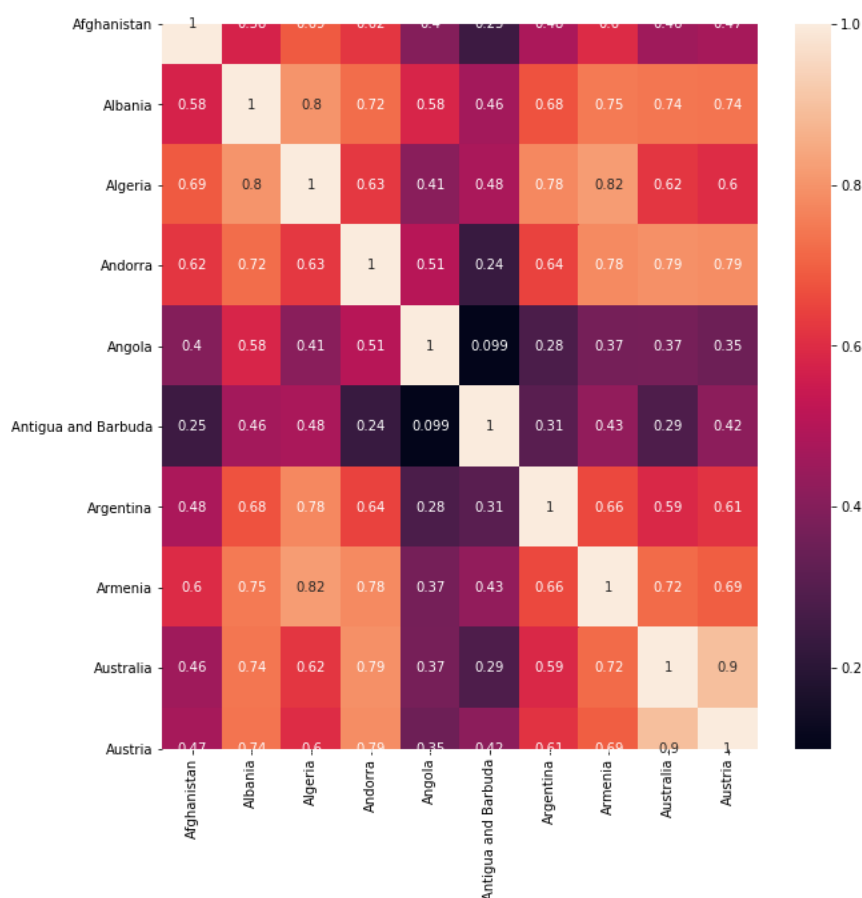


Deep Learning HW2_1 Report

0853411 劉書維

1. You should plot these coefficients in all the pairs of two countries.

我是用每天數間增加的案例 (differences) 當作一個 sequence，並計算出各國間的 correlation，因版面有限，故只印出 10 個國家的圖。

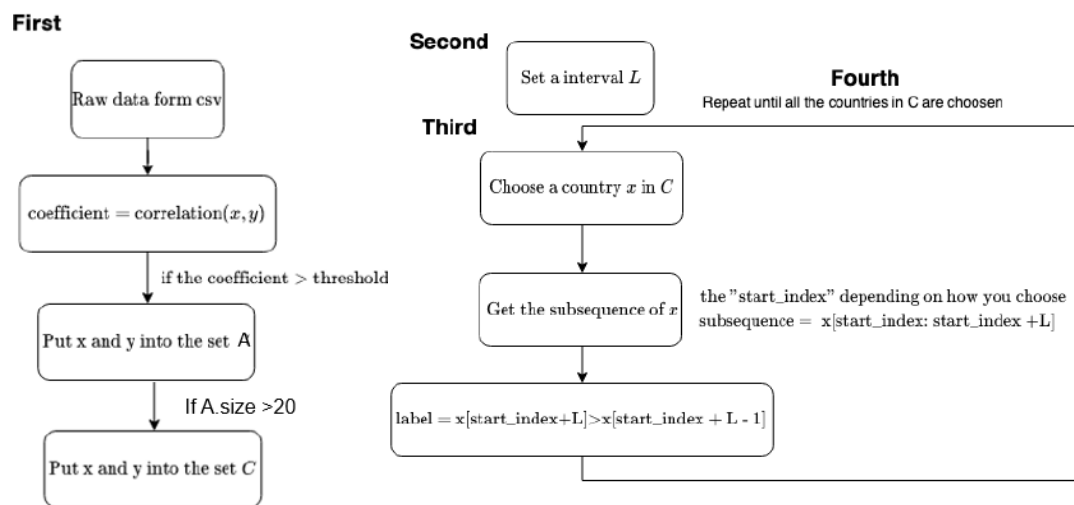


2. Process on you data.

我將 threshold 設為 0.8，就是將 correlation 大於等於 0.8 的國家配對找出，之後找出長度大於 20 的國家輸出 (表示這些國家彼此有相關性)。如下圖所示，size 大於 20 就挑選其國家。

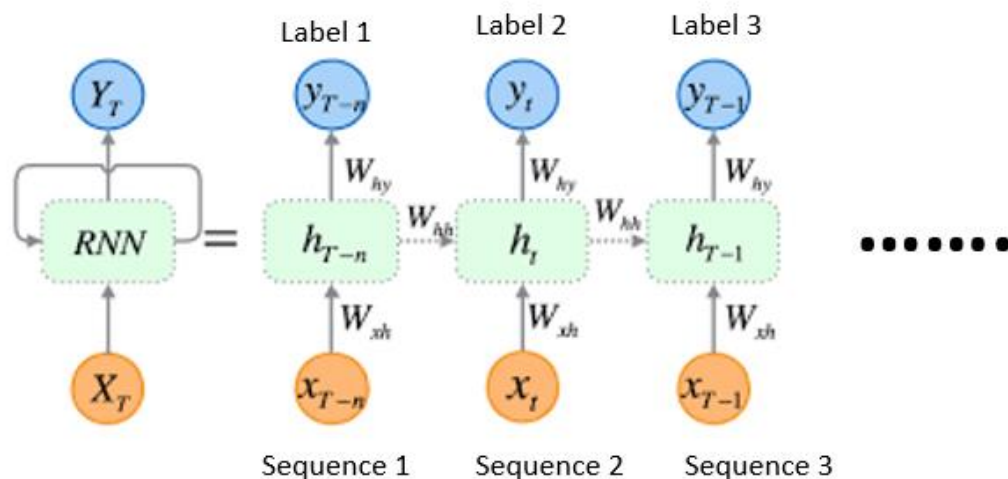
0	list	7	['Afghanistan', 'Azerbaijan', 'Canada', 'Indonesia', 'Poland', 'Saudi ...
1	list	17	['Albania', 'Algeria', 'Czechia', 'Denmark', 'Estonia', 'Germany', 'Ir ...
2	list	37	['Albania', 'Algeria', 'Armenia', 'Belgium', 'Bosnia and Herzegovina', ...
3	list	8	['Andorra', 'Costa Rica', 'Czechia', 'Germany', 'Iceland', 'Netherlands ...
4	list	1	['Angola']
5	list	1	['Antigua and Barbuda']
6	list	2	['Argentina', 'Panama']
7	list	12	['Algeria', 'Armenia', 'Belgium', 'Bulgaria', 'Costa Rica', 'Czechia', ...
8	list	16	['Australia', 'Austria', 'Costa Rica', 'Croatia', 'Germany', 'Iceland' ...

再將它們做好 label，即案例增加上升者標註為 1，反之為 0。增加案例的數目以 10 天一個單位當輸入，輸出第六天上升 (1)或下降 (0)。



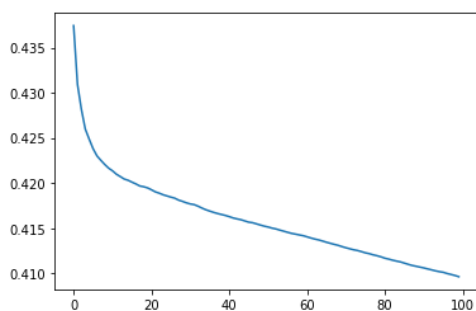
3. Build a recurrent neural network to predict the label based on the given sequence.

我使用標準的 RNN 架構。各參數如下：input_size = 10、hidden_size = 128、num_layers = 1、EPOCH = 100、BATCH_SIZE = 1、TIME_STEP = 1、INPUT_SIZE = 10、LR = 0.001。

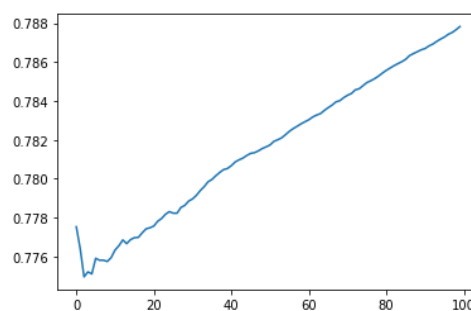


最後 train loss 和 test accuracy 如下圖所示：

Train loss：



Test acc：



可以看到 loss 穩定下降，且 test 的準確率在訓練 100 次後可以達到

78.8%。

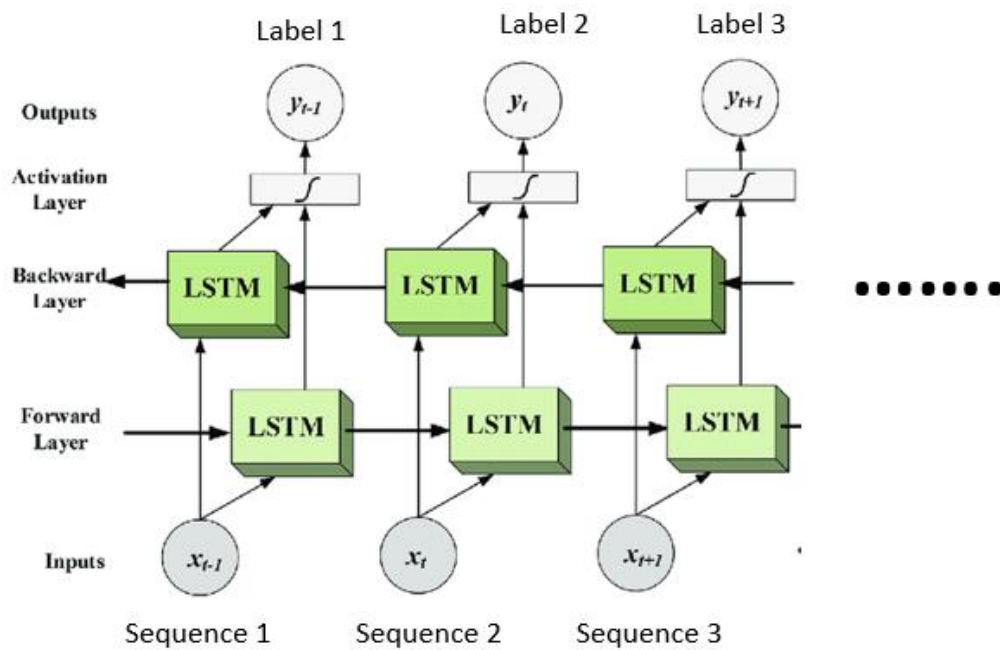
4. Please implement different recurrent neural network, LSTM.

我使用的 LSTM 架構跟 RNN 雷同。各參數如下：input_size = 10、

hidden_size = 128、num_layers=1、EPOCH = 100、BATCH_SIZE = 1、

TIME_STEP = 1、INPUT_SIZE = 10、LR = 0.001。雖然訓練較慢一些，但

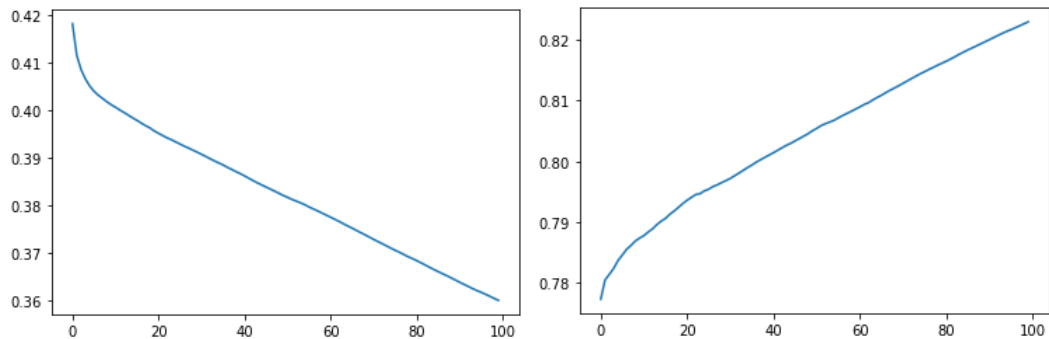
是準確率有比 RNN 更高。



最後 train loss 和 test accuracy 如下圖所示：

Train loss：

Test acc：



可以看到 loss 穩定下降，且 test 的準確率在訓練 100 次後可以達到

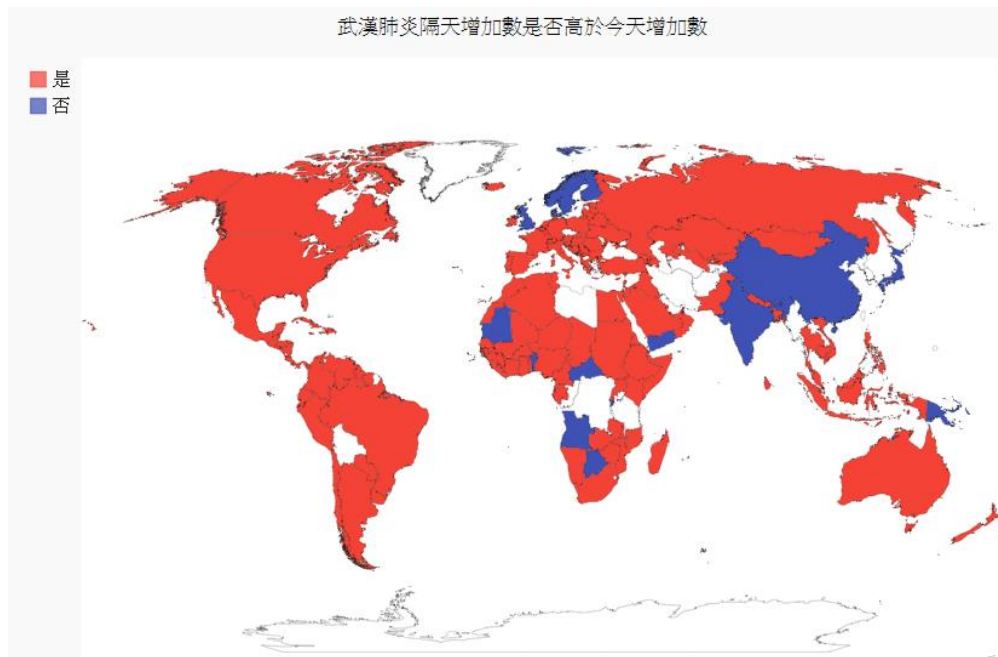
82.5%。

5. Compute the probability for each country and plot on a world map.

最後，將 LSTM 的 model 丟入各國最近 10 天的預測後一天確診案例是增

加還是減少，可以從 output.csv 中得知結果，而後將其可視化可以得到下

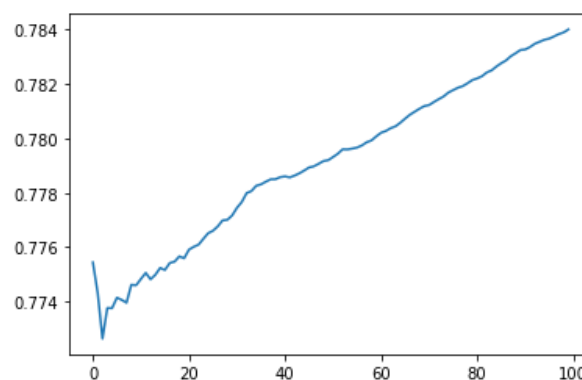
圖。



6. Do some discussion based on your result.

從過程和結果，可以得知和實際情形很類似，都是高峰過後的國家慢慢趨緩，而當時的西歐和美洲整處於水深火熱階段，所以也是慢慢上升，但是因為有些國家名稱和套件的國名沒有對上，所以沒有呈現在地圖上（包含 label 是 1 的台灣），也因此這個模型還是有參考價值的，可以套用在這次的疫情中做參考。

另外，我有將 LSTM 的 LR 放大十倍，也就是 0.01，發現 test 的 acc 反而無法快速上升，如下圖：



所以 learning rate 的調整也是很重要的。