

# DSP homework 1 report

B04902028    資工三    洪浩翔

## Environment:

使用系上工作站測試，可以執行 c code 且記憶體不要太小應該都可以跑。

## How to execute:

Type "make" to compile it first. Then, execute train\_hmm with input like `"/train_hmm iteration model_init.txt seq_model_01.txt model_01.txt "` to execute training program with 5 training input respectively. Training program will produce 5 different models. Finishing training, execute test\_hmm with input like `"/test_hmm modellist.txt testing_data1.txt result1.txt"`. Then, it will produce result file, and if testing data is "testing\_data1.txt", the program will produce acc.txt at the end, which is the accuracy rate of comparing our result with "testing\_answer.txt" file. After finishing all works, type "make clean" to remove all executing file.

## Summary of results:

結果的部分，若 iteration 為 1 的情況下，正確率大概 0.75 左右，而當 iteration 增加到 5 甚至是 50，正確率會下降到 0.6x 甚至快掉到 0.5，但當次數增加，正確率又會再次回升，當 iteration 為 1000 時，正確率大概 0.85，當 iteration 增加到 1500 時，正確率會上升到 0.87。最後，若是將 iteration 設成非常大，則正確率應會收斂至一個定值。資料夾中所包含的 5 個 model 則是在 iteration 是 1500 時 train 出來的。

關於程式的部分，考慮到沒定義輸入是否為同一長度情況下，所有陣列配置都是用 malloc 配置，這使得程式運行會有點花時間，另外，考慮到可讀性的關係，有些可以合併寫成的迴圈部分還是保持各自處理，這會增加運行時間，但是要重新讀 code 會比較好懂。目前 iteration 是 1500 的狀況下 train 完全部 5 個 model 至少要花 15-20 以上。