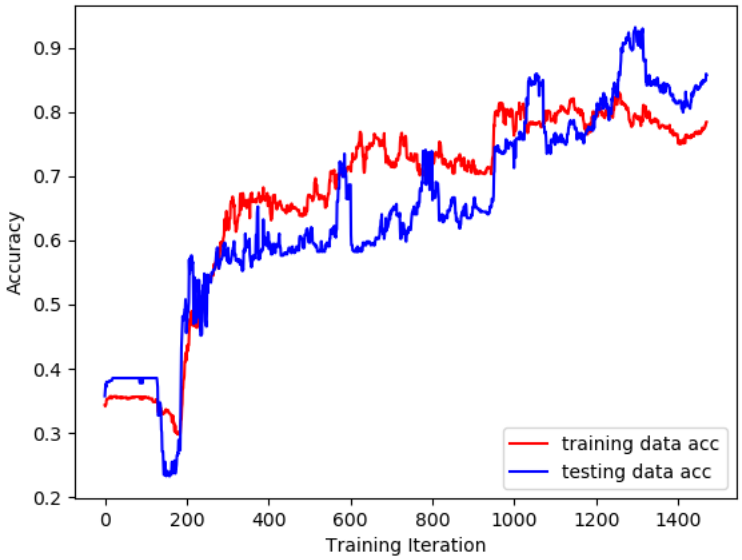
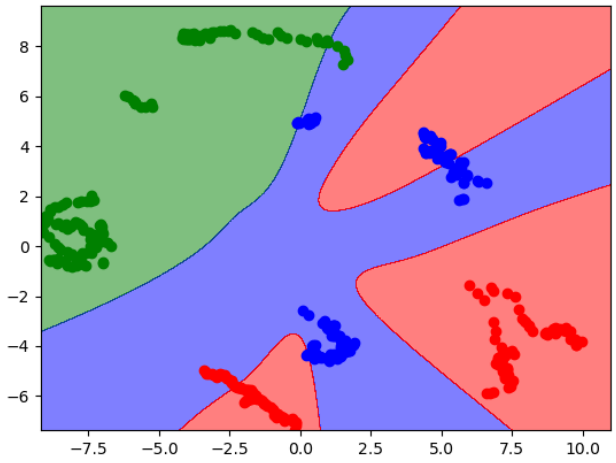
- How to run the code
  有兩個 py 檔，分別是 two_layer_NN.py 與 three_layer_NN.py，執行方式為在終端機輸入 python3 two_layer_NN.py 與 python3 three_layer_NN.py 即可。
- two_layer_NN.py 程式碼說明
  - 此為 two-layer neural network，分別有一層 input layer，一層 hidden layer(activation function 為 sigmoid)與一層 output layer(activation function 為 softmax)，input layer 為 2 個維度(用 PCA 降維至 2 維)與一個 bias，hidden layer 為 35 個 neuron 與一個 bias，output layer 為 3 個 neuron。Learning rate 設為 0.01。因為題目要求 SGD，每次更新參數的方式是從 1470 個 training data 中隨機不重複挑一個，先做 forward passing 得到 output(3 種水果的機率，總共三個數字，總和為 1)，再用此 output 與 ground true 計算多類別 cross entropy，並以 cross entropy 當作 loss function 做偏微分與 chain rule 得到各個參數的梯度(Backpropagation)並更新參數。參數總共更新 1470 次(讓 model 看過每一個 training data)。weight 於更新前的初始值為 mean=0 & variance=1 的 gaussian 隨機數，而 bia 於更新前的初始值為 0。此 model 在 testing data 上的 accuracy=85.74%。
  - 程式執行時會產出兩張圖，第一張圖如下

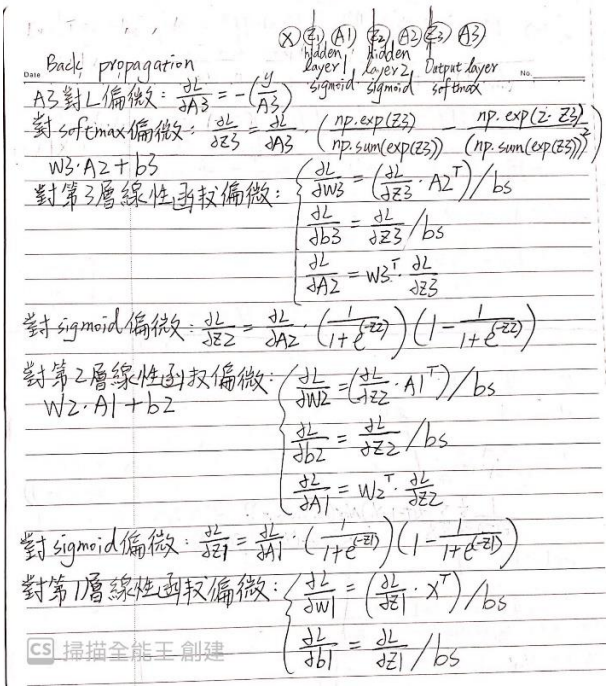


  我記錄下參數每次更新時對 training dataset 與 testing dataset 的 accuracy 變化，可發現隨著 model 更新次數越多，看過的 training data 也越多，training dataset 與 testing dataset 的 accuracy 都有上升的趨勢。
  - 第二張圖如下(decision regions)

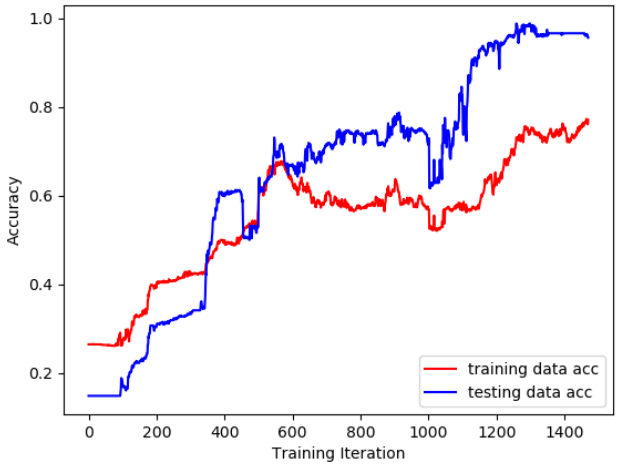


圖中的橫軸與縱軸分別是 testing dataset 做完 PCA 後得到了兩個維度，色塊區域是各種 2 維 feature 組合下餵給 model 的 predict 結果(3 種顏色代表三種水果)，而綠藍紅點是 testing dataset ground true，有些位置像是綠點點在藍色色塊就代表 model 預測錯誤，另外像是綠點點在綠色色塊代表 model 預測正確。

- three_layer_NN.py
  - 此為 three-layer neural network，與 two-layer neural network 的設置上幾乎相同，差別在多一層 hidden layer，而該層 hidden layer 的 activation function 是 sigmoid，兩層 hidden layer 的 neuron 數皆為 35，並且 learning rate 設為 0.003。此 model 在 testing data 上的 accuracy=95.58%。
  - 下圖為三層 Backpropagation 的公式推導

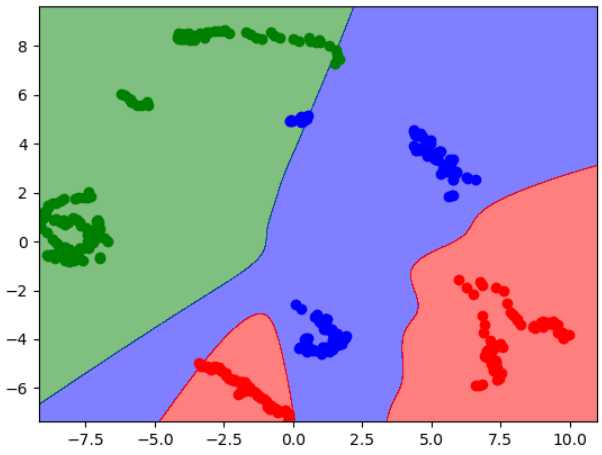


- 程式執行時會產出兩張圖，第一張圖如下



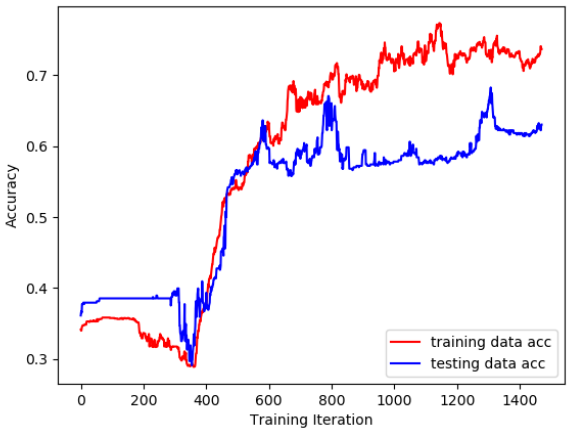
Training data acc 比 testing data acc 還低，這種情況是不常見的(2 layer model 也有此狀況)，訓練模型通常是相反的狀況，並且兩者 acc 落差懸殊而有 overfitting 的情形發生。個人推測 testing data 之 acc 較高可能的原因是 testing data 的辨識難度較 training data 簡單很多，並且兩組 dataset 有相似的 distribution，故 model 除了能在 training data 上學到對 testing data 有用的概念，且因為 testing data 更簡單，所以能在 testing data 上有更好的表現。

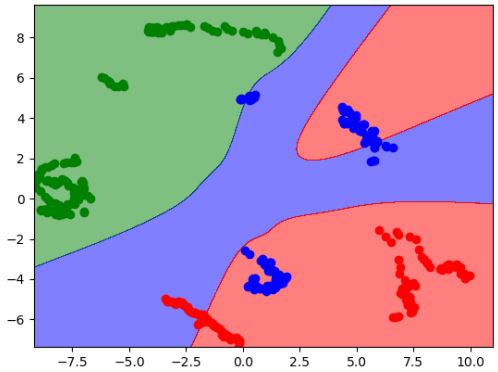
- 第二張圖如下(decision regions)



- two-layer neural network & three-layer neural network 比較
  - three-layer neural network 是在 two-layer neural network 的基礎上多加一層 hidden layer，故參數量較多，所以 model size 較大，可以處理較複雜的 task，並且在 testing data 上的表現有機會更好，而上文提到兩個 model 的 testing data acc，的確驗證了這件事。

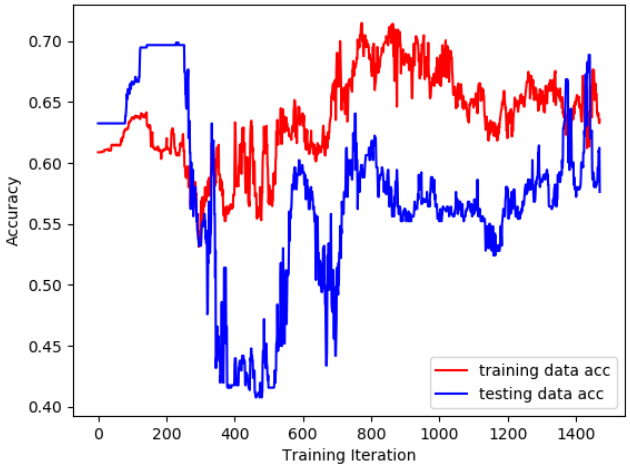
- Different settings on two-layer neural network
  我將原本 two_layer_NN.py 內的 learning_rate 從 0.01 設成 0.005，model 在 testing data 上的 acc 變成 63.05%，兩張結果圖如下:

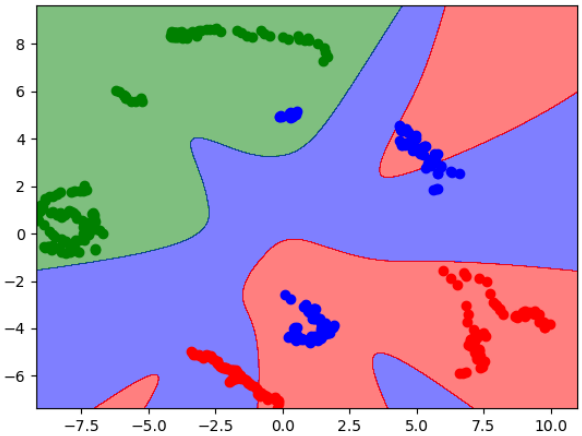




  折線圖的尾端可看到 acc 都還在攀升的階段，代表 model 還沒達到收斂，意味著 learning_rate 設成 0.005 太小了，應該調大 learning_rate 或是增加參數更新次數。

- Different settings on three-layer neural network
  我將原本 three_layer_NN.py 內的兩個 hidden layers 之 neuron 數量都從原本 35 調到 50，參數量變得更多。照理說 model 可以有更好的 performance，但是 model 在 testing data 上的 acc 變成 57.63%，結果圖如下:





Model size 變大可以處理更複雜的 task 或是有更好的表現，但是也不是絕對，要看 task 難易度與 training data 的量來決定，若在 task 過於簡單或是 training data 量不夠多的情況下提升 model size 會造成反效果。上面的結果即可證明此事。