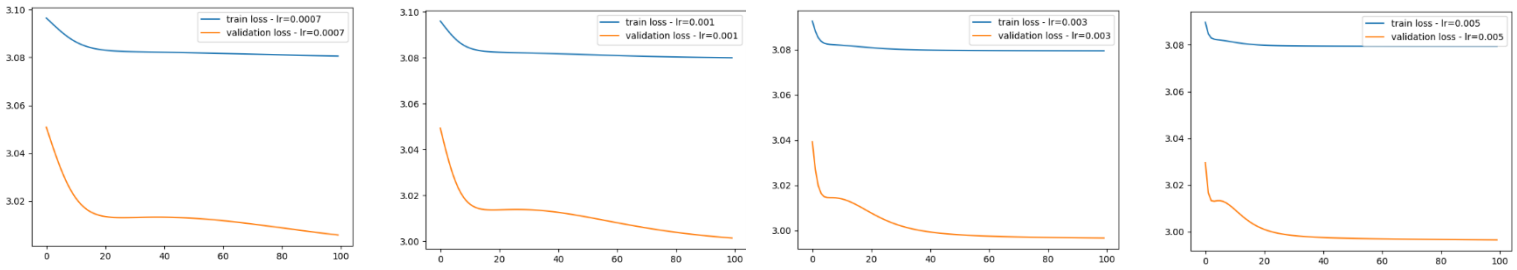
Introduction to Neural Networks Homework1

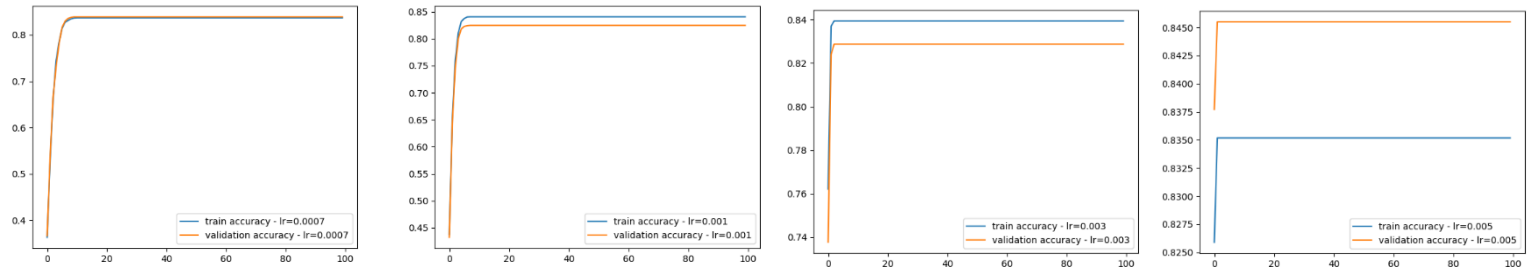
N26100618 李姵萱

I. Data preprocessing
   A. 類別型資料 ： 做 one-hot encoding。
   B. 數值型資料 ： 做標準化，使平均值=0，標準差=1。
   C. Y : 將 yes、no 轉換為 1 與 0 後，做 one-hot encoder。
   D. Imbalance data : 答案為 0 的資料筆數過多，隨機刪減至與答案為 1 的資料筆數相同，各 4640 筆。
   E. Split Data : 將資料分為 train 72%、validation 18%、test 1，並對 training data 做 shuffle。

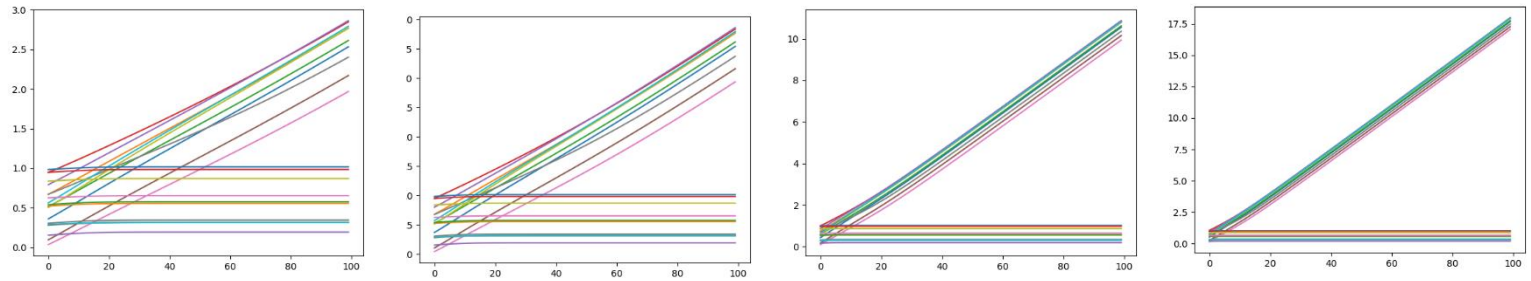
II. Neural Network (1 hidden layer)
   A. 設定 hidden layer 有 10 個 neuron，output size 為 2 (二分類，因此 output layer 的 activation function 固定為 softmax)，loss function 為 cross entropy。
   B. 變化 learn rate、hidden layer 的 activation function，訓練 100epochs，觀察 loss、accuracy、weight change。
      1. Activation function = sigmoid，learning rate = 0.0007、0.001、0.003、0.005。
         a. Loss 變化



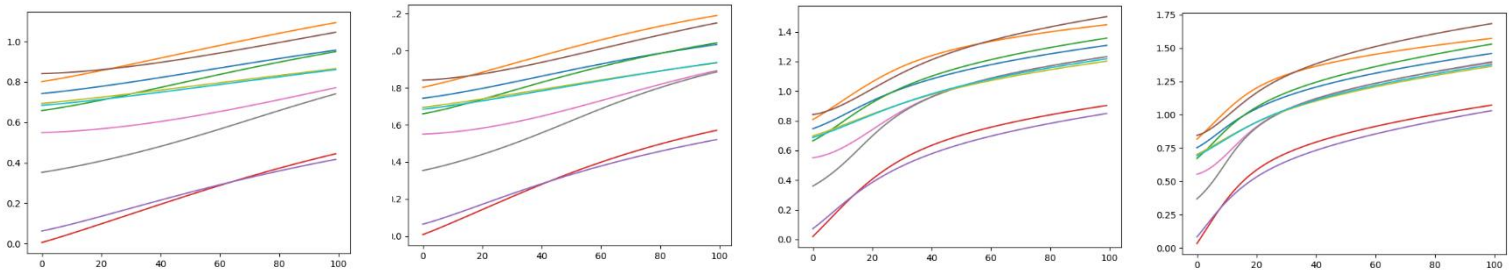
         b. Accuracy 變化



         c. Weight 變化
            ✧ Output layer weights

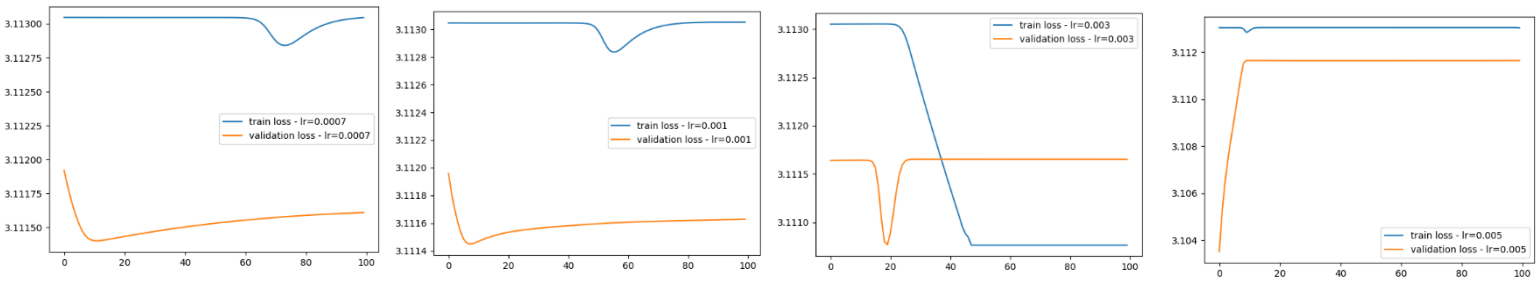


✧ Hidden layer bias

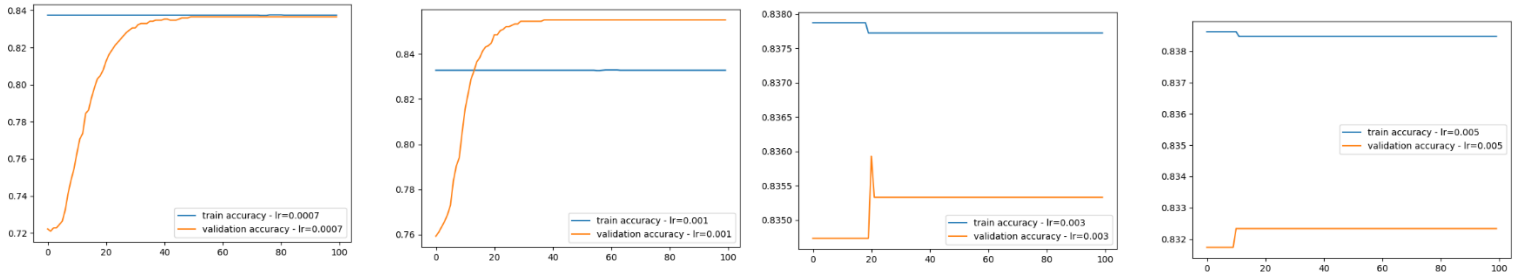


2. Activation function = relu，learning rate = 0.0007、0.001、0.003、0.005

a. Loss 變化

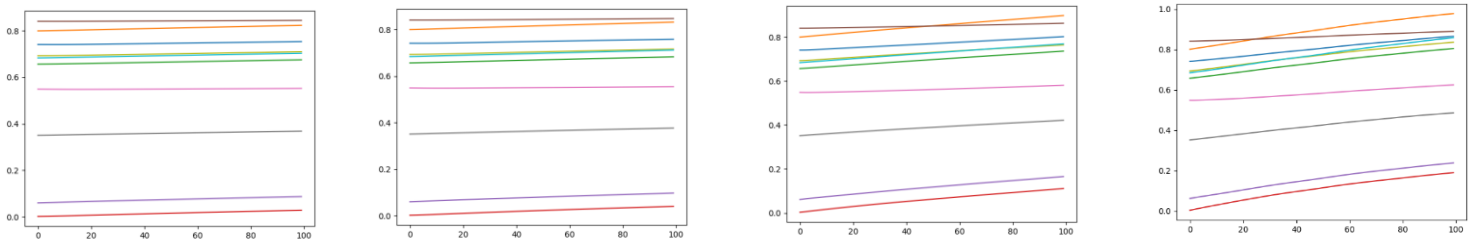


b. Accuracy 變化

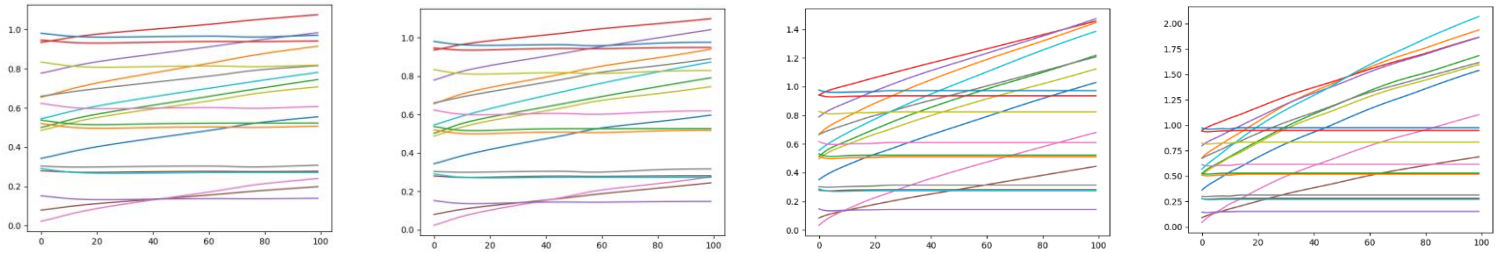


c. Weight 變化

✧ Output layer weights



✧ Hidden layer bias



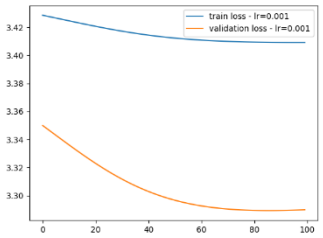
III. Radial Basis Function Network

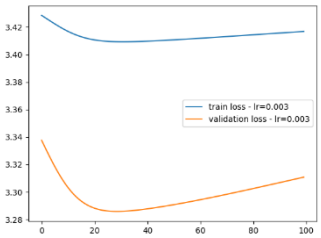
A. 設定 hidden layer 有 10 個 neuron，output size 為 2 (二分類，因此 output layer 的 activation function 固定為 softmax)，loss function 為 cross entropy。

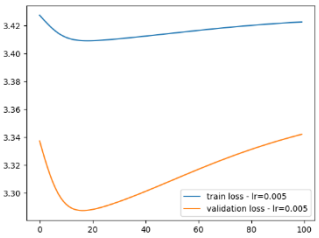
B. 變化 learn rate、radial basis function，訓練 100epochs，觀察 loss、accuracy、weight change。

1. Radial basis function $\varphi(r) = e^{-(\varepsilon r)^2}$，learning rate = 0.001、0.003、0.005。

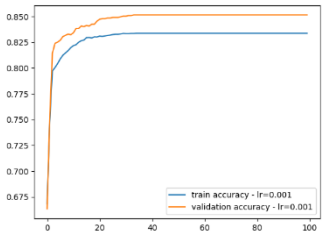
a. Loss 變化

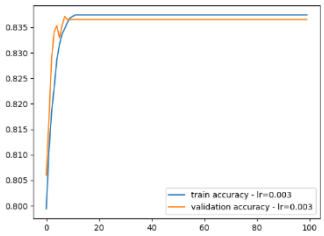


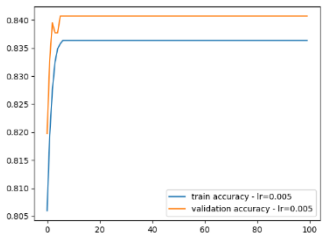




b. Accuracy 變化

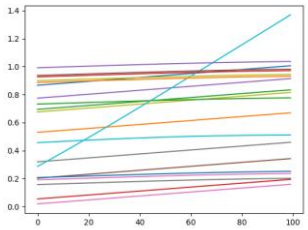
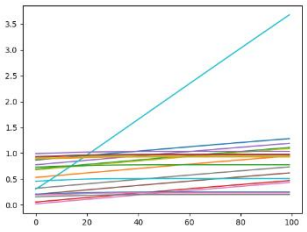
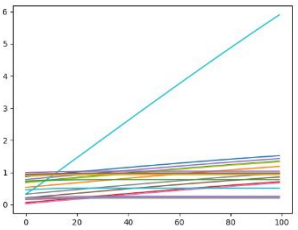






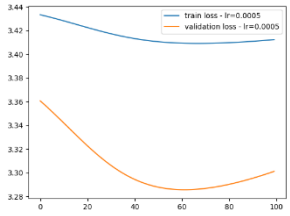
c. Weight 變化

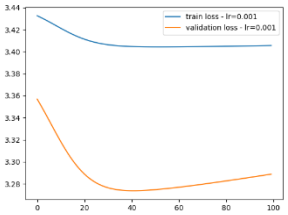
✧ Output layer weights

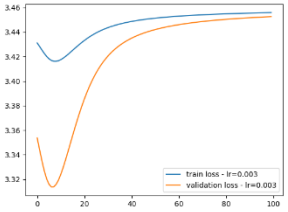




2. Radial basis function $\varphi(r) = \frac{1}{\sqrt{1+(\varepsilon r)^2}}$，learning rate = 0.0005、0.001、0.003。

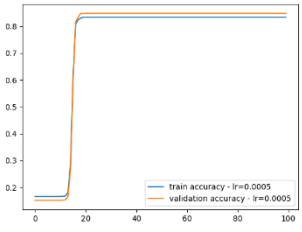
a. Loss 變化

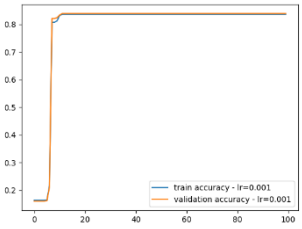


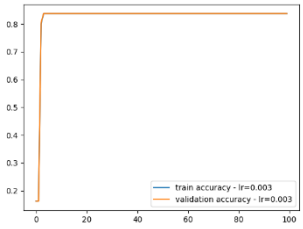




b. Accuracy 變化

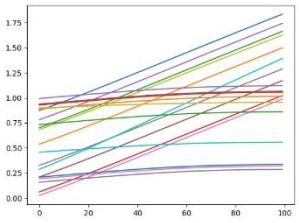
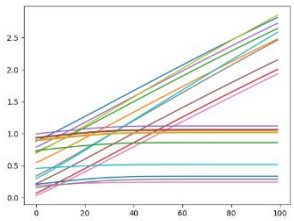
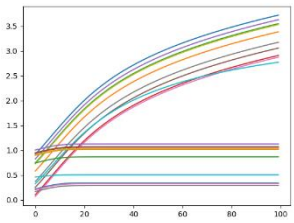






c. Weight 變化

✧ Output layer weights

IV. 比較與討論

A. Data preprocessing

1. Training data 有 shuffle 對於 validation data 的 accuracy 提升有幫助。
2. 資料標準化後訓練時會較順利，即較容易找到訓練得起來的 hyperparameters。
3. 做 undersampling 較不會只學到單一種資料的資訊，對於準確度提升亦有幫助。

B. Neural network

1. 比較相同 activation function，不同 learning rate
   a. learning rate 越大，loss、accuracy、weight 收斂越快。
2. 比較 relu 與 sigmoid 兩個 activation function
   a. Sigmoid 收斂的曲線較平滑，也較不易有 overfitting 的情形。
   b. Relu 只要 learning rate 稍微過大，很容易就有嚴重的 overfitting，甚至比訓練前的結果還糟。

C. Radial Basis Function Network

1. 比較相同 radial basis function，不同 learning rate
   a. learning rate 越大，loss、accuracy、weight 收斂越快。
   b. learning rate 稍大 loss 就容易 overfitting。
2. 比較 2 個 radial basis function
   a. $\varphi(r) = e^{-(\varepsilon r)^2}$ 需要較大的 learning rate。
   b. $\varphi(r) = \frac{1}{\sqrt{1+(\varepsilon r)^2}}$ accuracy 的部分在一開始會有停滯期，不像 $\varphi(r) = e^{-(\varepsilon r)^2}$ 一下子就有明顯成長。

D. NN 與 RBFN 比較

1. NN 可以訓練的參數較多，搭配足夠多的數據，較容易找到 hyperparameter 並訓練起來。
2. RBFN 較不容易訓練成功，需多試幾組 hyperparameters。且訓練起來後 loss 非常容易一下子就 overfitting。
3. 使用 NN，並用 sigmoid 當作 hidden layer 的 activation function，最適合本次的資料集。