

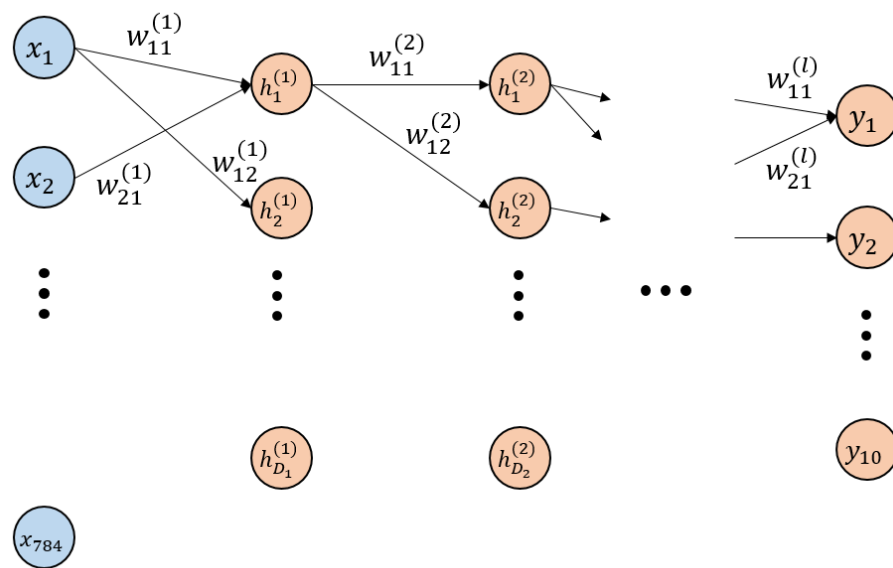
Deep Learning Homework 1 Report

(一) 實驗目的

在本實驗中必須去實作一個 Deep Neural Network (DNN) 來對資料庫進行分類和回歸。

(二) 數學模型

主要分成三的部分：feed forward、back propagation 以及更新權重。首先會先將資料送進如 Figure 1 所示的神經網路中，將得到的結果記錄下來，進行第二部分的計算。Back propagation 算出來的 gradient 將用來更新權重。一直反覆這三個步驟直到權重收斂時即可停止。



▲Figure 1 DNN 示意圖

(a) Feed forward

首先會先將資料送進如 Figure 1 所示的神經網路中，並將得到的結果 y 以及途中各節點的值記錄下來，以便等等進行 back propagation 計算時使用。在計算之前，必須先初始化各個節點的權重，初始化的方式是 He initialization 來隨機亂數產生。

(b) Back propagation

接著要利用 back propagation 算出各權重對目標函式 E 的貢獻。也就是說，我們必須算出每一個權重對目標函式偏微分的值 (gradient)。

(c) 更新權重

算出個權重所對應的 gradient 之後，就可以來進行參數的更新。如此反覆進行 (a)、(b) 和 (c) 的計算，就可訓練出適當的 DNN 模型。

(三) 實驗結果

題目一: regression

特徵的選擇方式為使用全部的特徵或刪去其中的一項特徵來比較其在此回歸中的重要程度，此題目有 8 個特徵所以有 9 種結果。

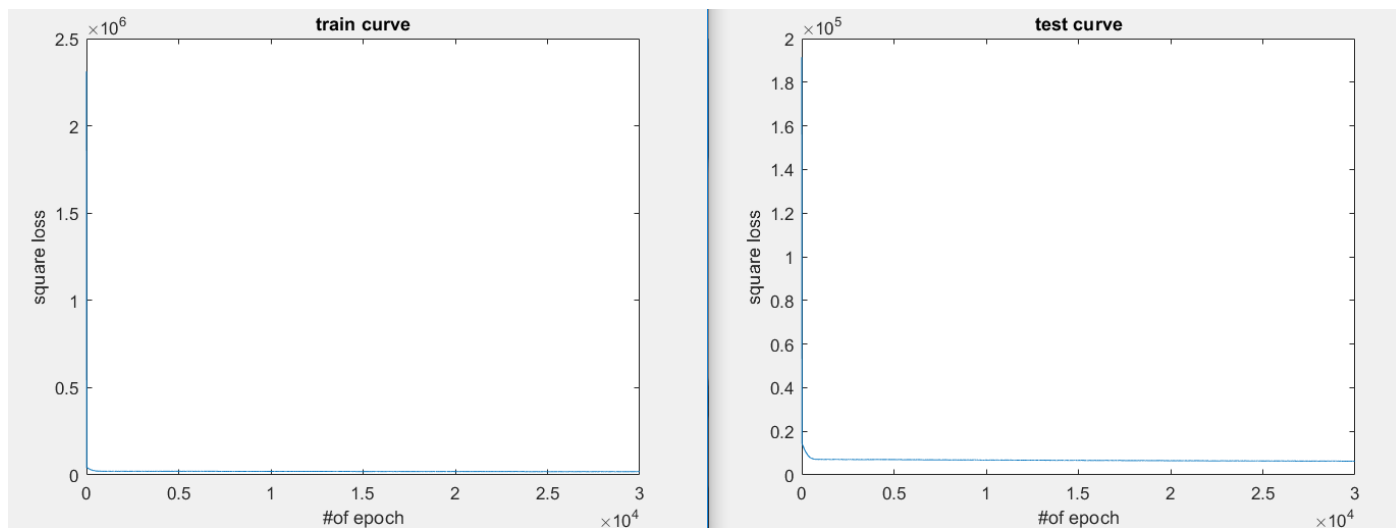
題目一實驗所使用的參數如下表所示：

Number of hidden layer	3
Learning rate	0.0000000022
Number of epoch	30000
Data in a mini-batch	18

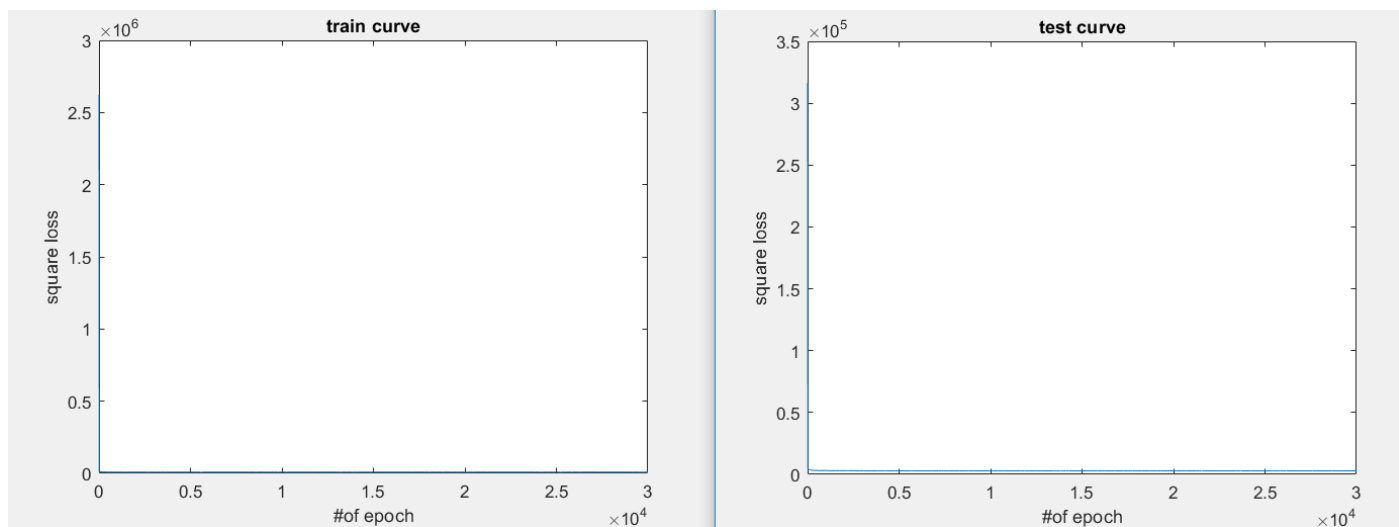
	Hidden units
Hidden layer 1	15
Hidden layer 2	10
Hidden layer 3	10

學習曲線如下所示：

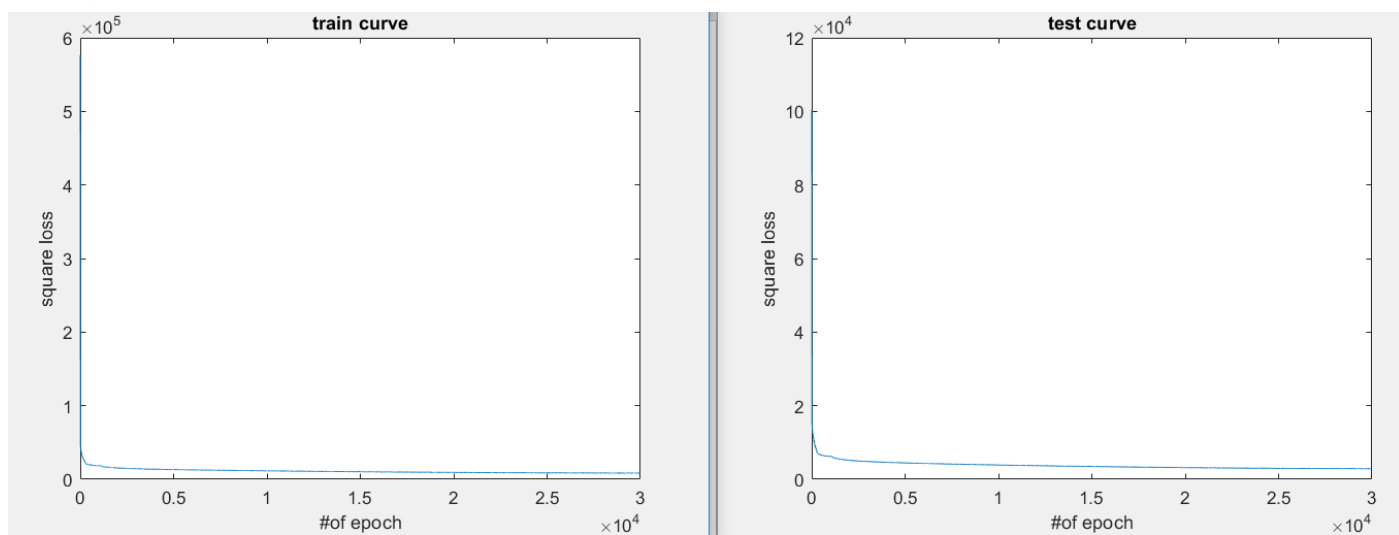
刪去第一個特徵：



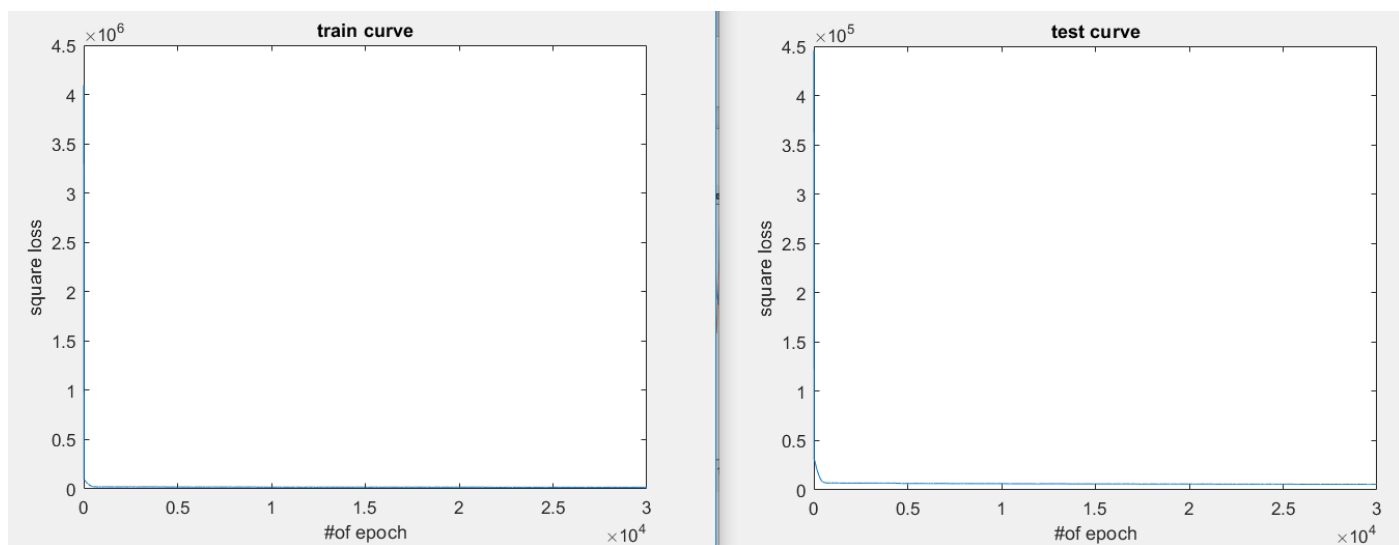
刪去第二個特徵：



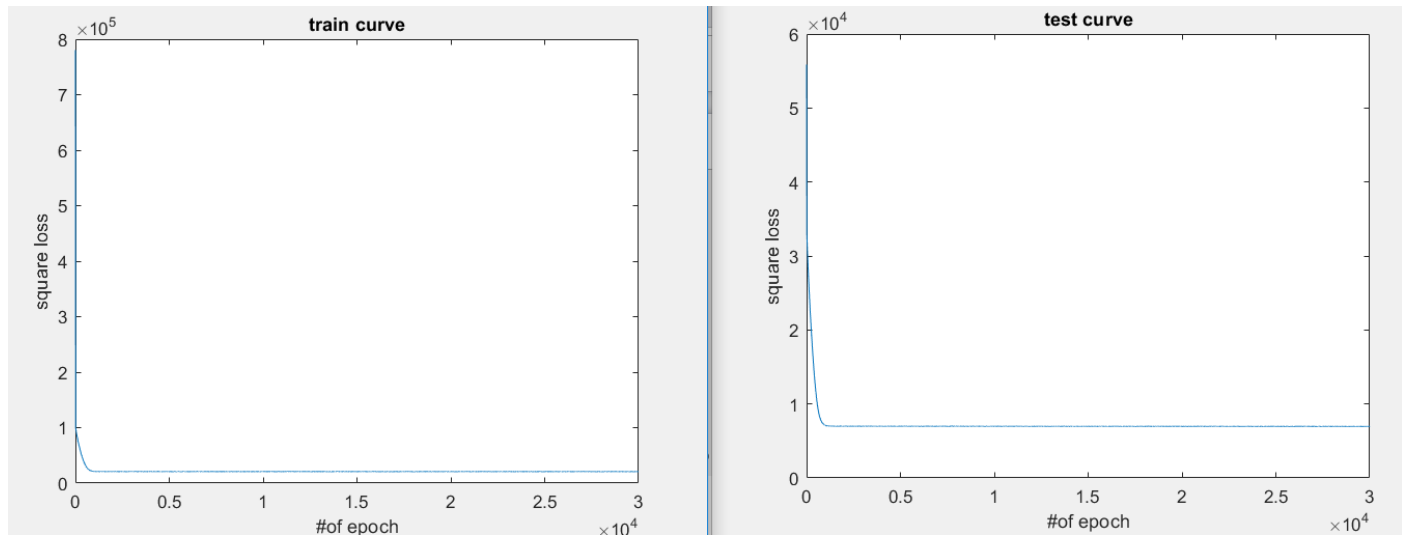
刪去第三個特徵：



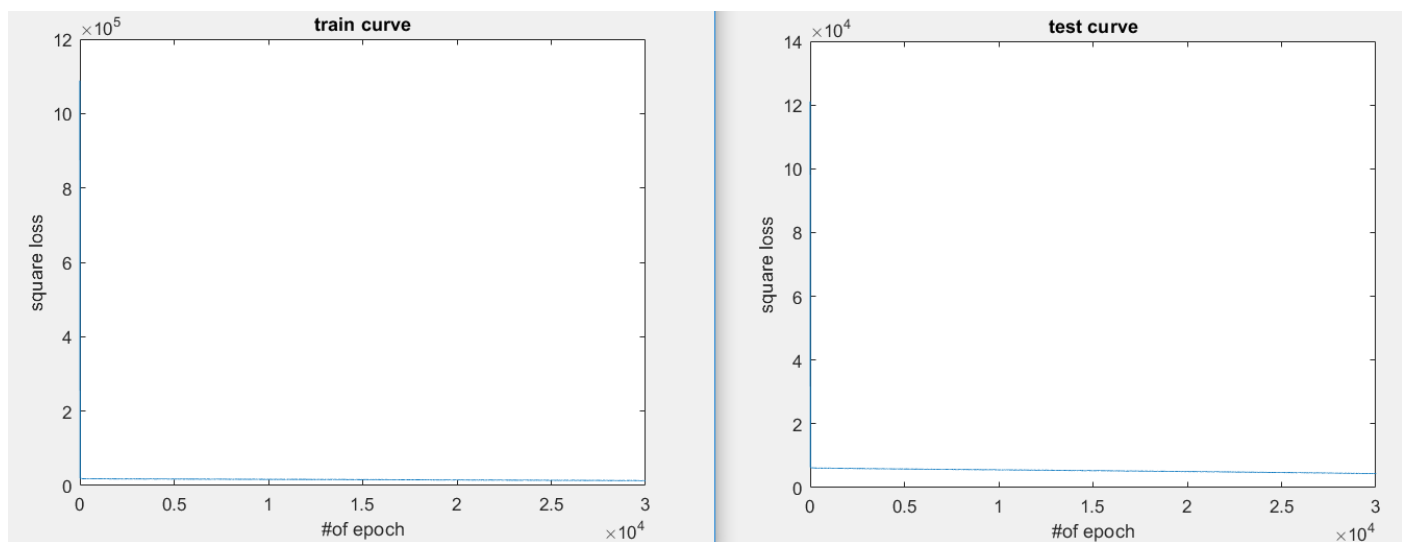
刪去第四個特徵：



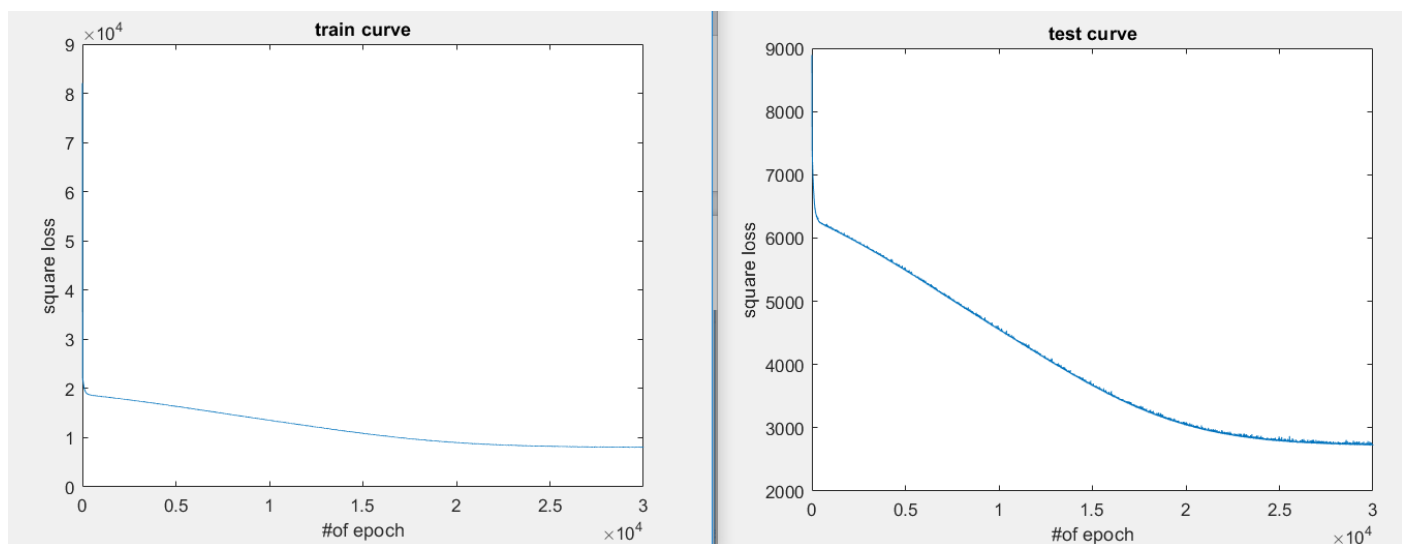
刪去第五個特徵：



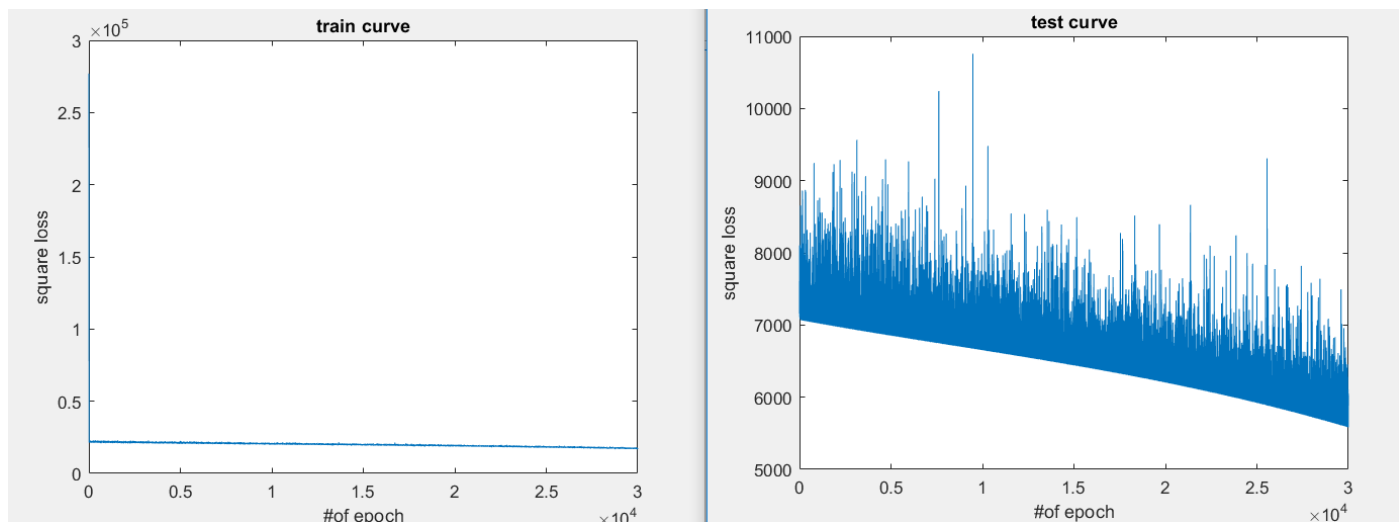
刪去第六個特徵：



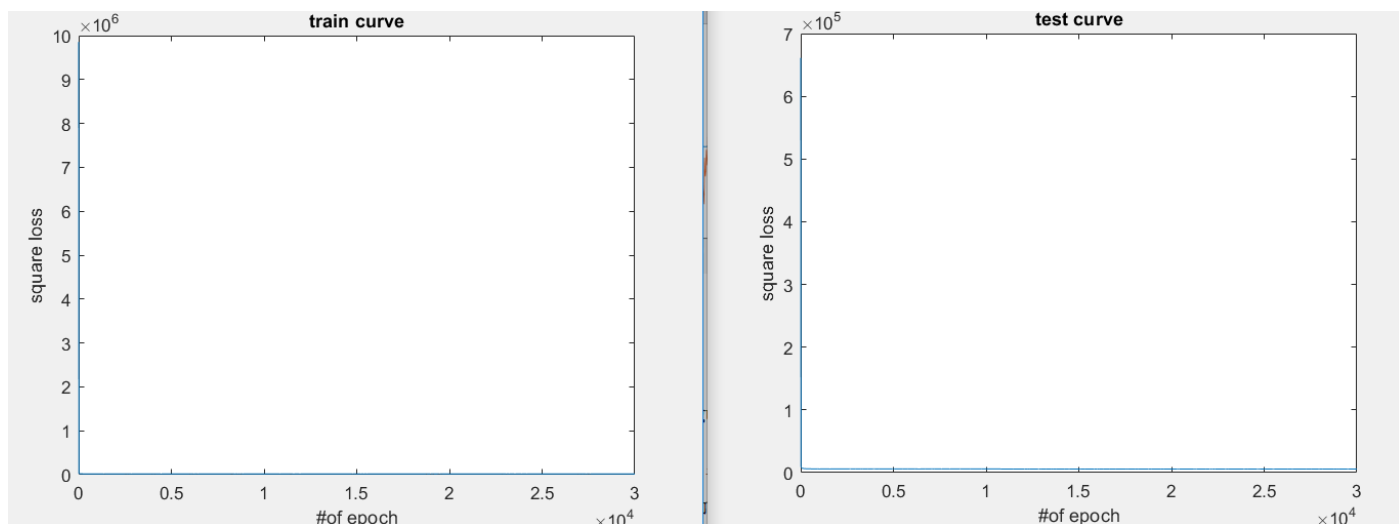
刪去第七個特徵：



刪去第八個特徵：



使用所有特徵：

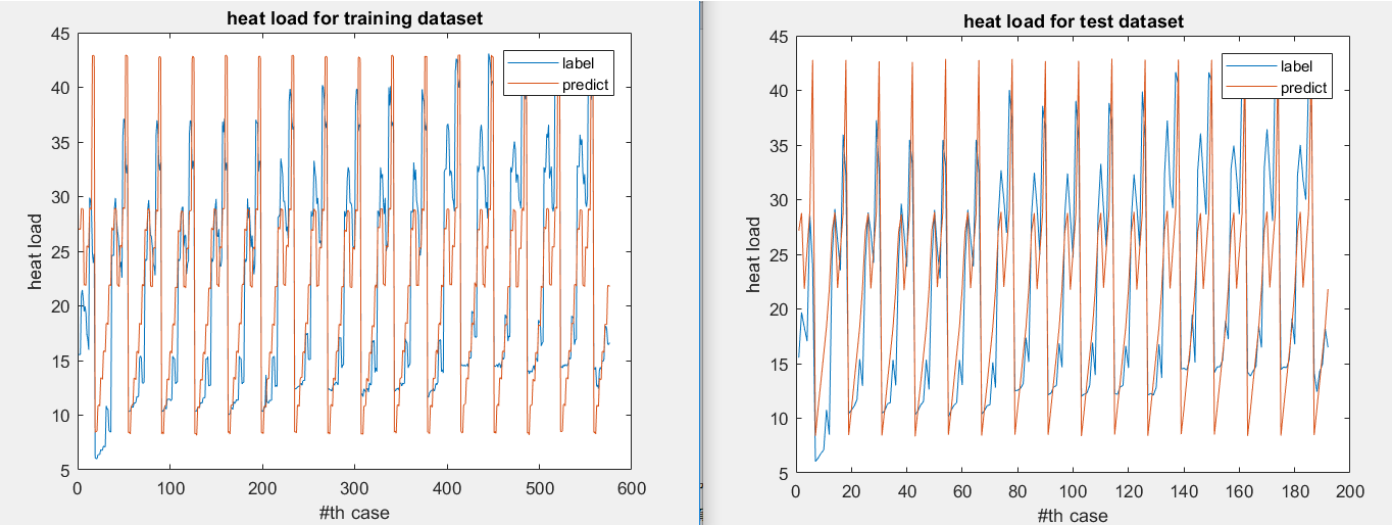


▲ 訓練和測試資料的學習曲線

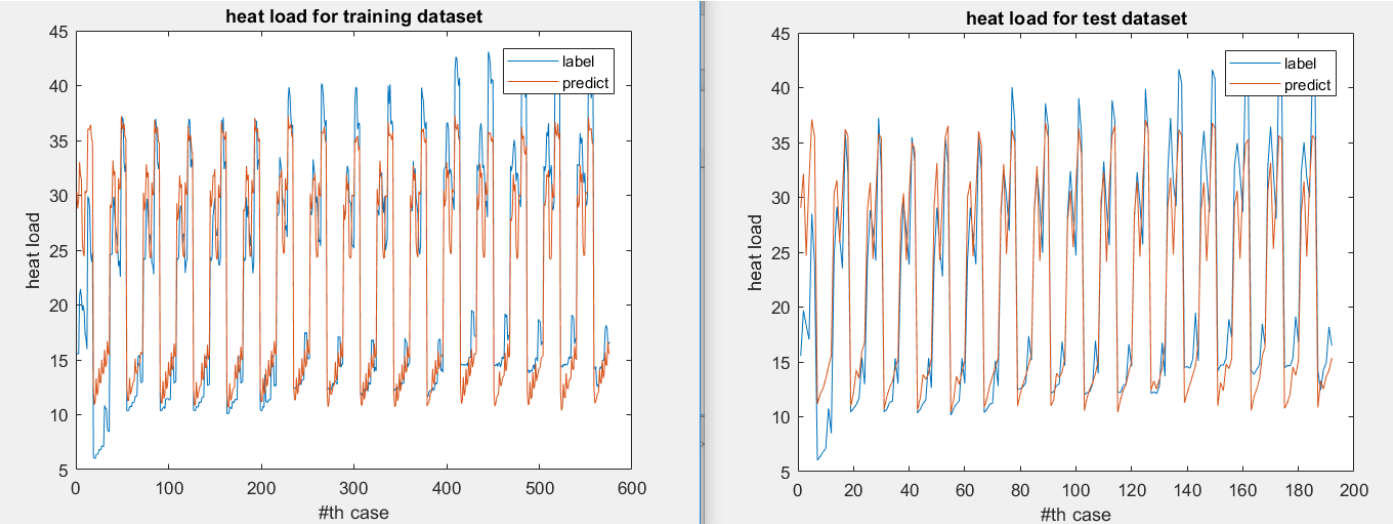
由所有特徵放入訓練的結果可看出，其中訓練資料表現最低的 error 收斂在 5.41；而測試資料表現最低的 error 為 5.4135。也就是說，訓練資料的表現還是比測試資料好，但差異不大，是我們期待的結果。

回歸結果如 Figure 3 所示：

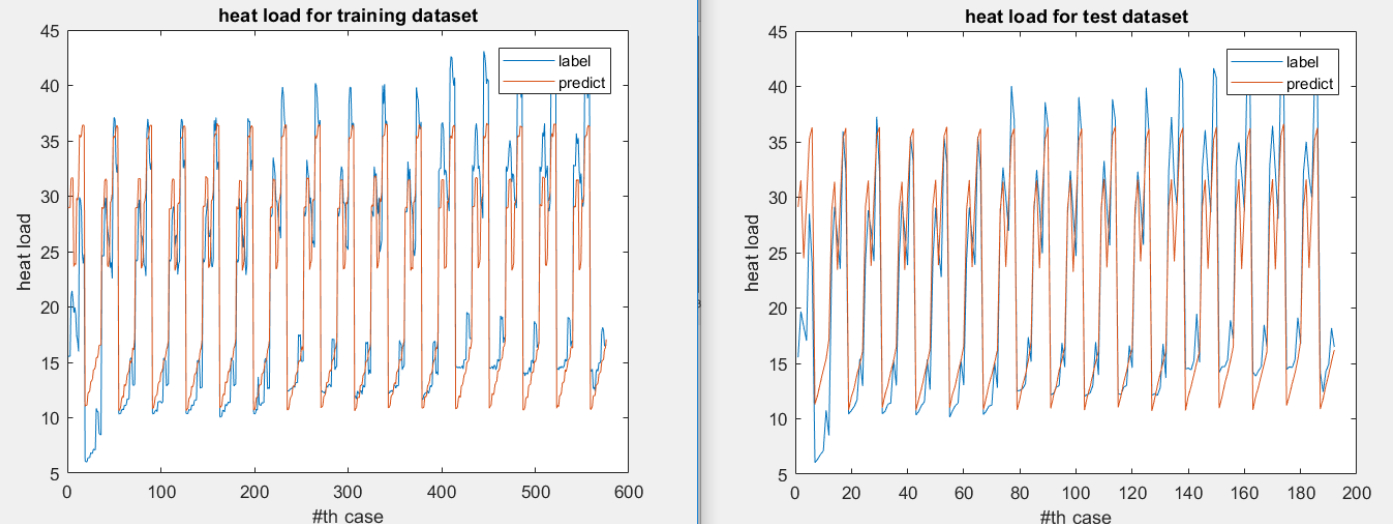
刪去第一個特徵：



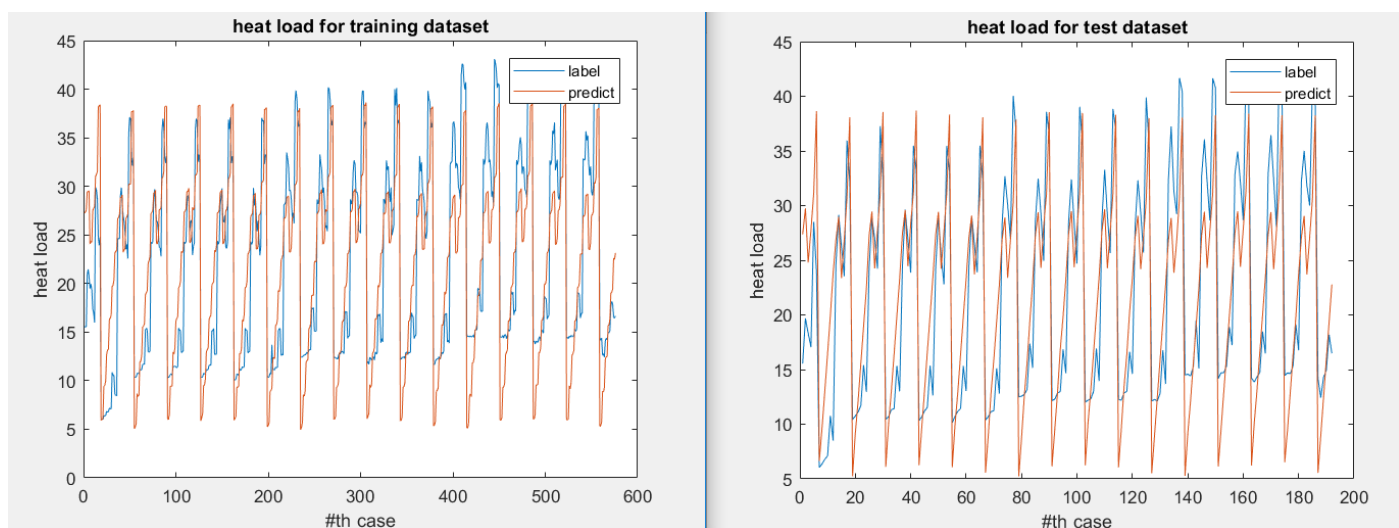
刪去第二個特徵：



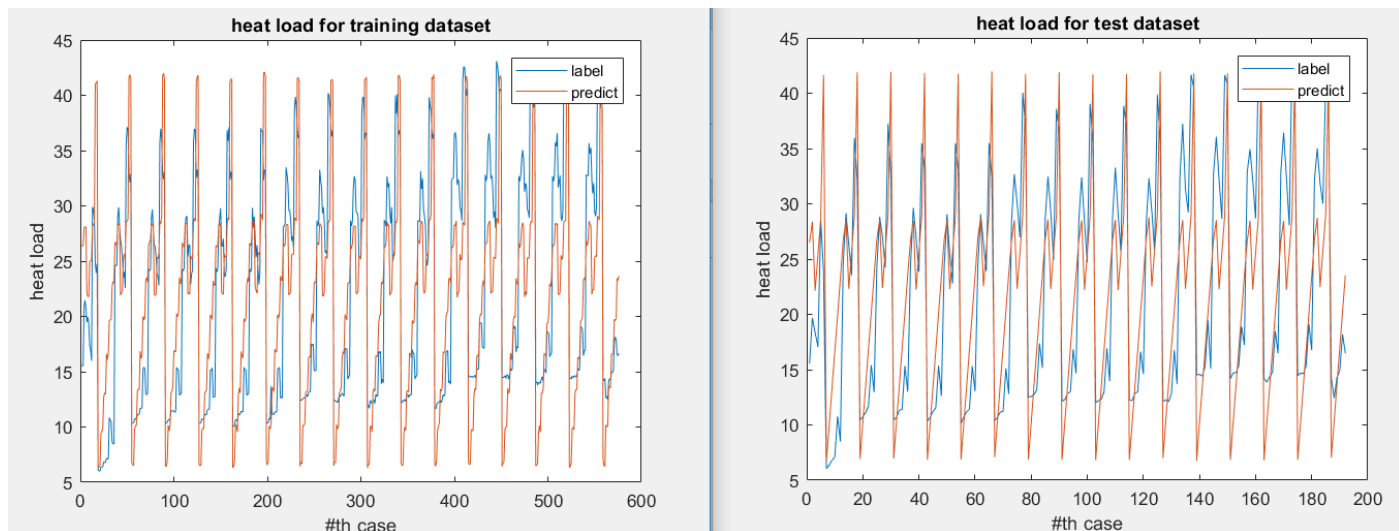
刪去第三個特徵：



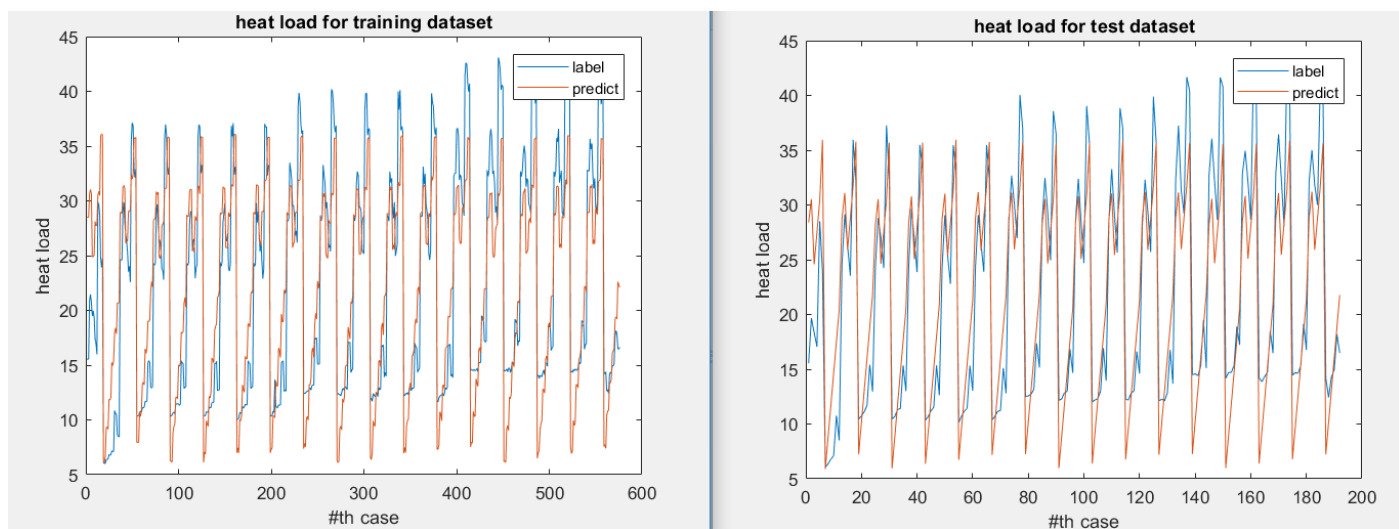
刪去第四個特徵：



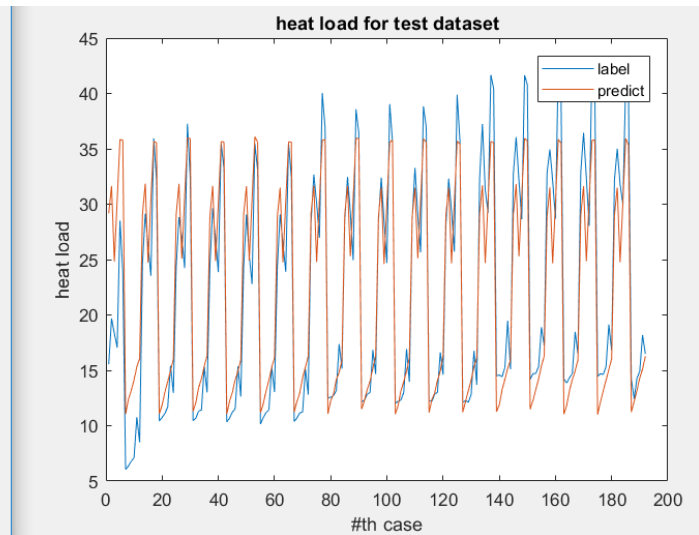
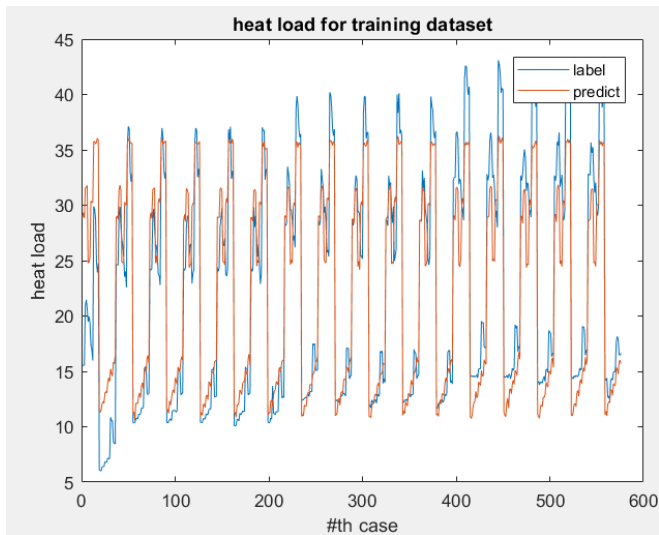
刪去第五個特徵：



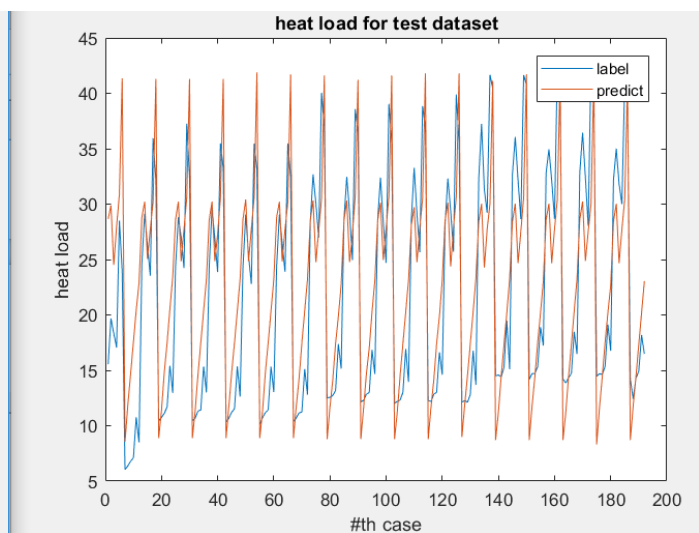
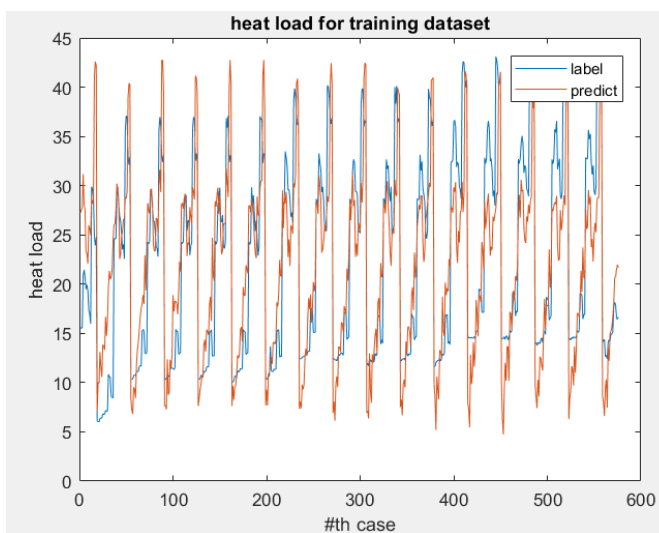
刪去第六個特徵：



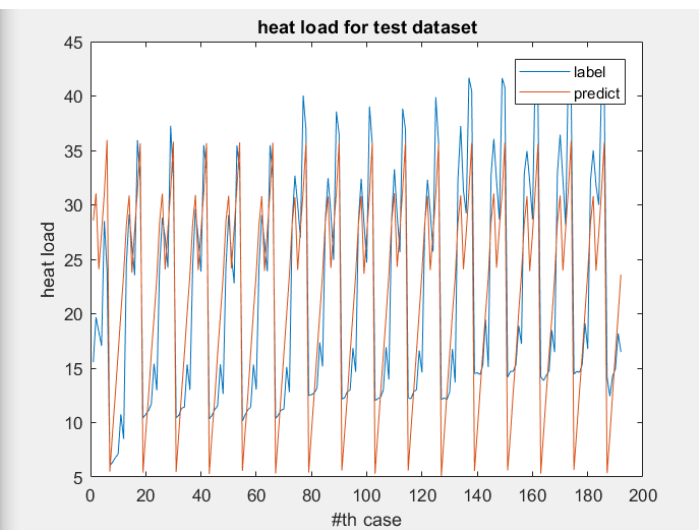
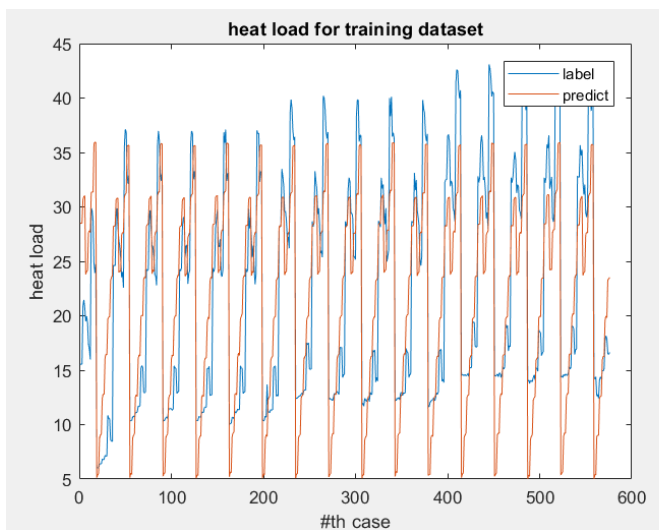
刪去第七個特徵：



刪去第八個特徵：



使用所有特徵：



▲訓練和測試資料的回歸結果

由結果可以看出預測值跟標籤值十分接近，在一開始打亂資料集的時候看到的結果是相對雜亂的，解決方法是打亂資料集後記住其位置，在輸出時使用原本的位置輸出，即能看到漂亮的結果，且能將資料集打亂使每次的訓練結果不同。

特徵重要性的討論如下：

Network architecture	15-10-10-1	15-10-10-1	15-10-10-1	15-10-10-1	15-10-10-1	15-10-10-1	15-10-10-1	15-10-10-1	15-10-10-1
Selected features	2-8	1, 3-8	1, 2, 4-8	1-3, 5-8	1-4, 6-8	1-5, 7-8	1-6, 8	1-7	1-8
Training ERMS	5.73138 2541	3.79917 5155	3.82770 9893	5.42854 8577	6.00754 9867	4.77149 9182	3.73750 4238	5.52565 9435	5.41041 3993
Test ERMS	5.73676 292	3.84891 9143	3.87112 8026	5.45637 9052	6.02660 1557	4.78141 0043	3.77050 4904	5.53552 5196	5.43541 7473
編號	1	2	3	4	5	6	7	8	9

我們從 ERMS 去討論特徵的重要性，可以看出 error 最大的是編號 5 跟 1(少掉第一個特徵與第五個特徵的情況)，代表這兩個特徵是比較重要的，因為少了該特徵造成我們預測誤差很大，error 最小的是編號 2 和 7(少掉第二個特徵與第七個特徵的情況)，且 error 比全部的特徵丟入訓練還要小，代表這兩特徵較不重要，甚至降低了我們預測的表現。

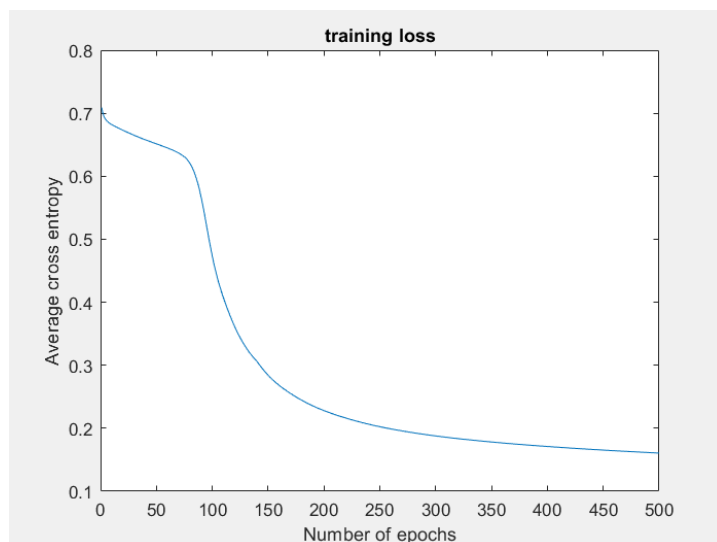
題目二:classification

題目二實驗所使用的參數如下表所示：

Number of hidden layer	3
Learning rate	0.00002
Number of epoch	500
Data in a mini-batch	100

	Hidden units
Hidden layer 1	150
Hidden layer 2	100
Hidden layer 3	30

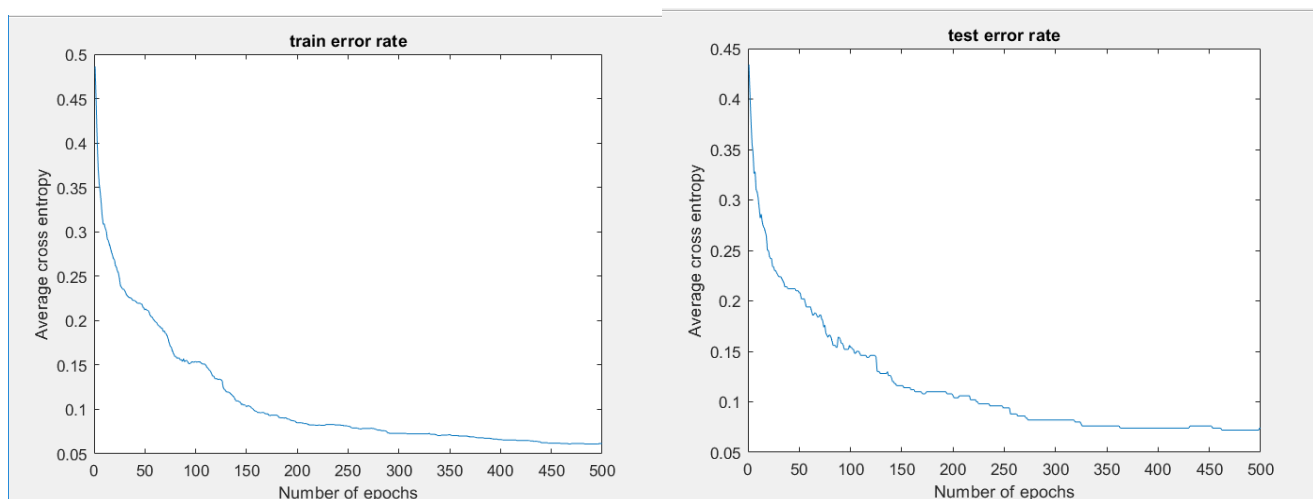
學習曲線如 Figure 4 所示：



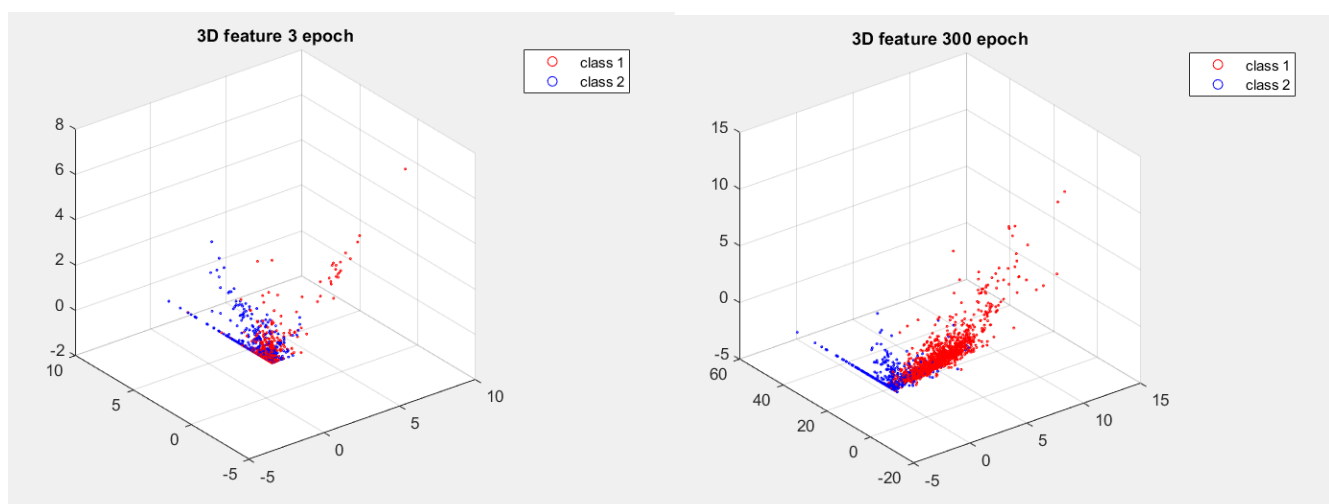
▲Figure 4 訓練的學習曲線

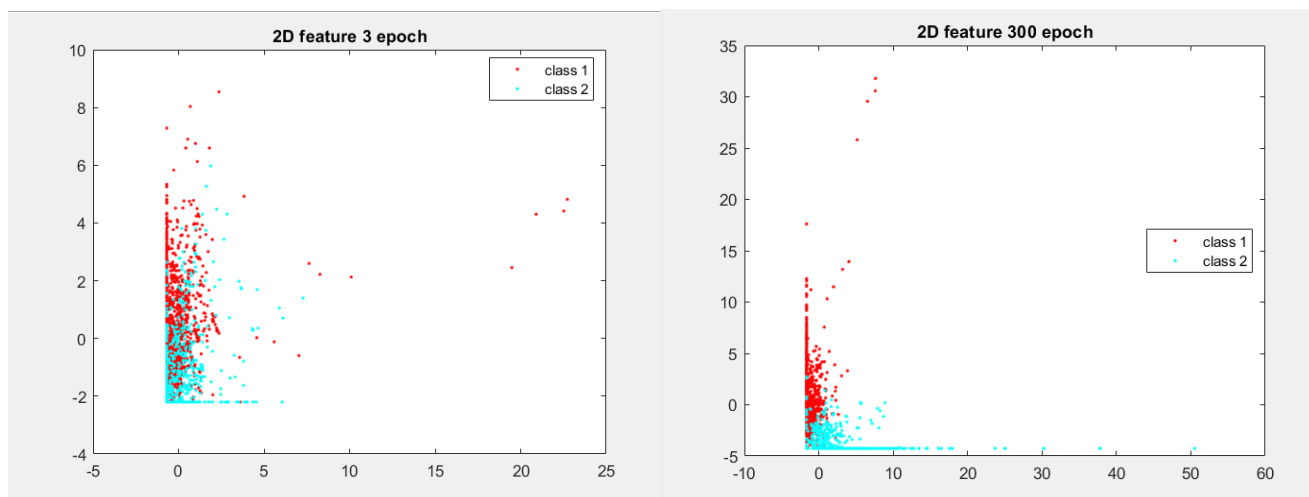
最後 average cross entropy 收斂在 0.16 左右

Error rate 結果如 Figure 5 所示：



在最後 epoch，其中訓練資料表現最低的 error rate 收斂在 0.06；而測試資料表現最低的 error 為 0.074。也就是說，訓練資料的表現還是比測試資料好，但差異不大，是我們期待的結果。





由圖可以看出，兩種不同種類的 feature 距離越來越遠，可以得知隨著我們訓練次數變多，分類越來越好，feature 的分布也越來越集中且明確分成兩個不同的群集。