

## 1. Introduction (20%)

我使用 class 來實作 neural network，可調的超參數為：data、hidden unit 數量、epoch、learning\_rate，data 為 0 代表使用了 generate\_linear 資料，data 為 1 代表使用 XOR 資料。

## 2. Experiment setups (30%)

### A. Sigmoid functions

使用了助教提供的程式碼。

```
def sigmoid(x):  
    return 1.0 / (1.0 + np.exp(-x))  
  
def derivative_sigmoid(x):  
    return np.multiply(x, 1.0-x)
```

### B. Neural network

我使用 class 來實作，以下為詳細解釋：

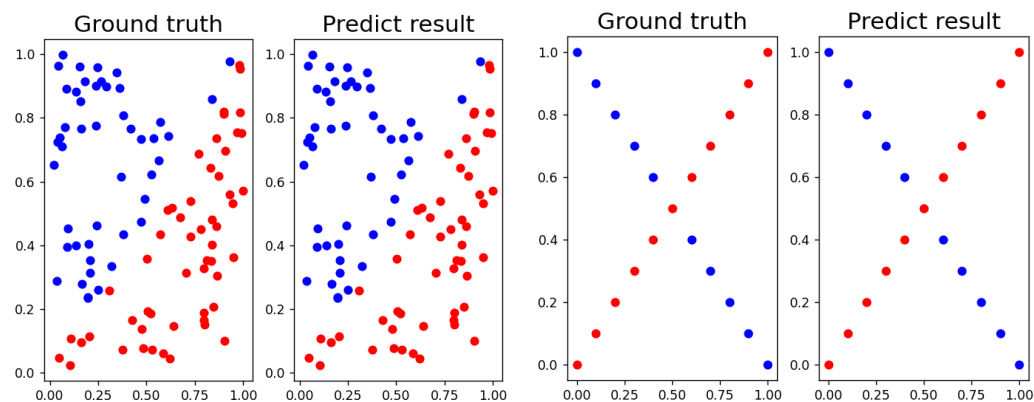
1. `__init__(self, hsize1, hsize2)`：hsize 為 hidden units 的數量。使用 `np.random.rand` 初始化 weight 與 bias，如果使用 `np.full` 塞滿 0.1 初始值的話，在學習 XOR 時好像會掉進 local minimal，導致最後 accuracy 並非 100%。
2. `forward(self, x)`  
進行矩陣乘法，乘上 weight 與加上 bias，再用 sigmoid 函式，得出下一層的輸入，重複 3 次後得到最終輸出。
3. `backward(self, x, y, learning_rate)`  
chain rule 告訴我們有兩個部分要算，一部分就是與 weight 相乘的輸入，另一部分可由後往前算得到，所以首先先算出 error，乘上 `derivative_sigmoid` 後得出誤差，再繼續往回算，最終分別跟輸入相乘得出 gradient，進行更新學習。
4. `train(self, x, y, epochs, learning_rate)`  
使用 forward 與 backward 來進行學習，同時輸出 loss 與紀錄學習曲線。
5. `predict(self, x)`  
輸入 input，輸出預測結果。

### C. Backpropagation

如 B.3.。

3. Results of your testing (20%)

A. Screenshot and comparison figure

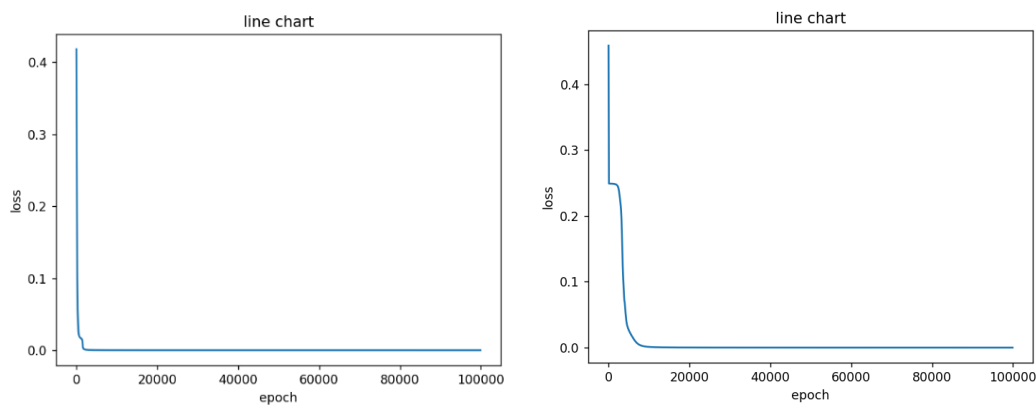


B. Show the accuracy of your prediction

Iter89	Ground truth: 0	prediction: 0.00002
Iter90	Ground truth: 0	prediction: 0.00002
Iter91	Ground truth: 1	prediction: 0.99996
Iter92	Ground truth: 1	prediction: 0.99998
Iter93	Ground truth: 1	prediction: 0.99998
Iter94	Ground truth: 1	prediction: 0.99998
Iter95	Ground truth: 1	prediction: 0.99998
Iter96	Ground truth: 1	prediction: 0.99998
Iter97	Ground truth: 1	prediction: 0.99139
Iter98	Ground truth: 0	prediction: 0.00015
Iter99	Ground truth: 0	prediction: 0.00003
loss=0.00000 accuracy=100.0%		

Iter10	Ground truth: 0	prediction: 0.01173
Iter11	Ground truth: 0	prediction: 0.00070
Iter12	Ground truth: 1	prediction: 0.99142
Iter13	Ground truth: 0	prediction: 0.00009
Iter14	Ground truth: 1	prediction: 1.00000
Iter15	Ground truth: 0	prediction: 0.00003
Iter16	Ground truth: 1	prediction: 1.00000
Iter17	Ground truth: 0	prediction: 0.00002
Iter18	Ground truth: 1	prediction: 1.00000
Iter19	Ground truth: 0	prediction: 0.00002
Iter20	Ground truth: 1	prediction: 1.00000
loss=0.00001 accuracy=100.0%		

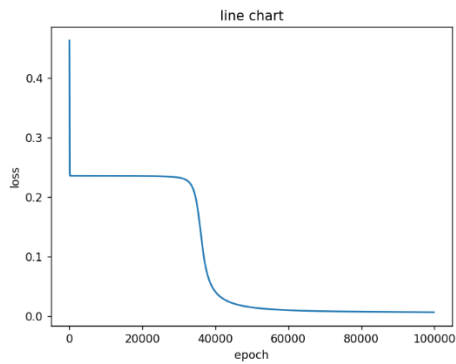
C. Learning curve (loss, epoch curve)



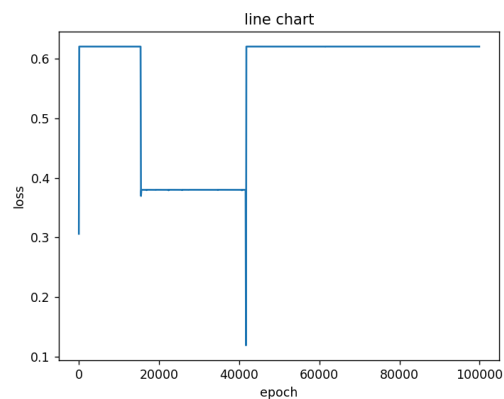
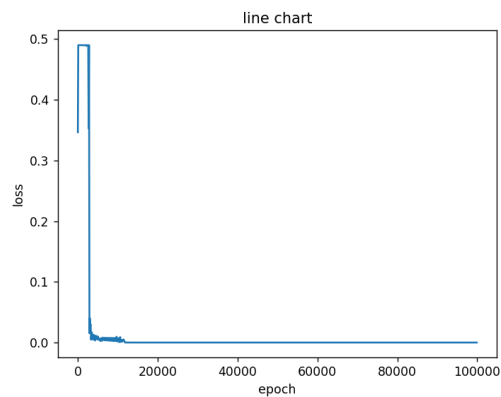
#### 4. Discussion (30%)

##### A. Try different learning rates

當 learning rates 設成 0.001 時，learning curve 曲線漸緩，且有可能掉進 local minimal 跑不出來。



當 learning rates 設成 0.9 以上時，會開始不穩定。



##### B. Try different numbers of hidden units

就算數量皆設為 1，準確率還是能到達 100%。

##### C. Try without activation functions

會發生數字 overflow 的問題。