DL Lab1 Backpropagation

0516054 劉雨恩

1. Introduction

在本次實驗目的為實作一個兩層的隱藏層(Hidden Layer)的神經網路，並能瞭解正向傳遞(Forward Pass)和反向傳遞(Backpropagation)的作用和運作原理。實驗規定僅能利用Numpy和python標準函式庫完成，並不能使用如Tensorflow或Pytorch的模板。測試的兩個數據為linear和XOR的分類，神經網路在最後loss會收斂，得到的output為極趨近0或1的數值，且得出的圖片和正解一模一樣。

2. Experiment setups

A. Sigmoid functions

實驗使用的activation function為sigmoid function，可以使網路輸出的結果落於[0,1]的區間，免於結果過大或過小。

$$\sigma(x) = \frac{1}{1-e^{-x}}$$
$$\sigma'(x) = \frac{e^{-x}}{(1+e^{-x})^2} = \sigma(x) \cdot (1 - \sigma(x))$$

B. Neural network

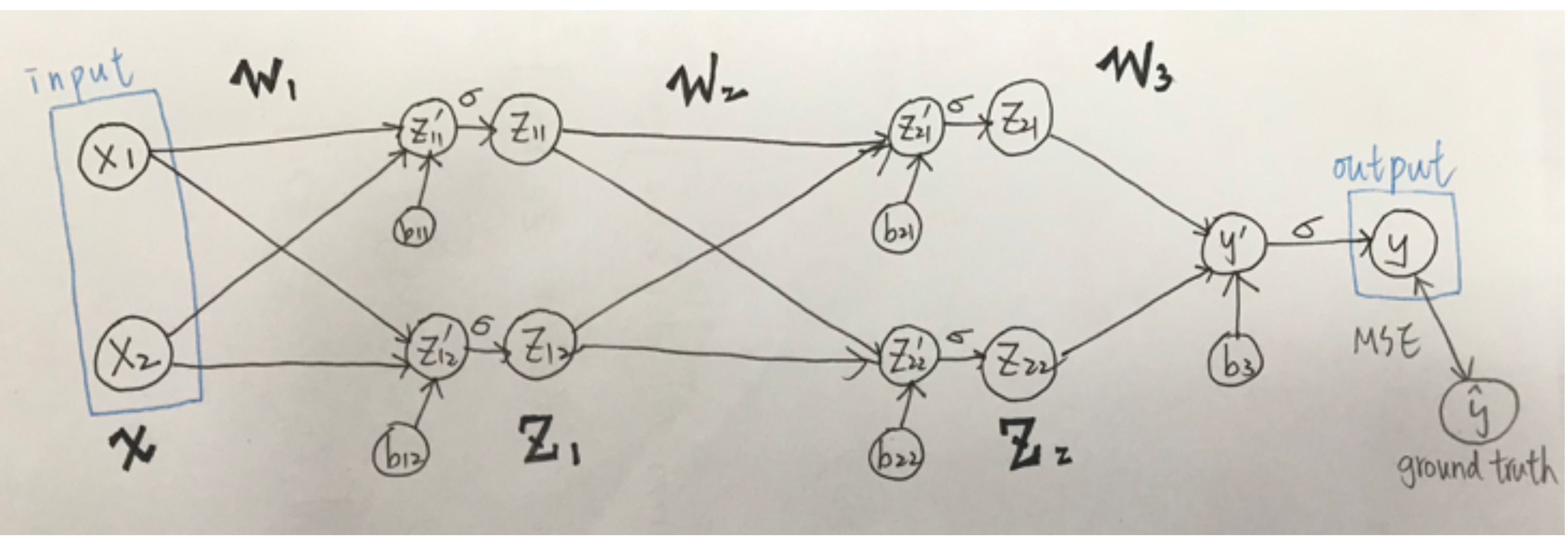

Figure 1. Forward pass

實驗設計的神經網路有兩層hidden layer，而每個layer都分別有兩個hidden units。在訓練的每一個epoch，會將input的點座標一個個forward計算，最後會output出一個接近0或1的數值，然後放入loss function跟正解比較。然後為了使loss function最小，而使用backprapogation來調整網路參數。藉由forward pass和backprapogation的反覆循環迭代，可以使網路獲得最佳訓練結果。
以下為實驗的設定：

1. x_1, x_2: neural network inputs
2. X: [x_1,x_2]

3. y: neural network output
4. ŷ: ground truth
5. W_1, W_2, W_3: weight matrix of network layers
6. b_1, b_2, b_3: bias matrix of network layers
7. Loss Function: Mean Square Error
8. Activation Function: Sigmoid function
9. Learning Rate: (1)linear: 0.1 (2)XOR: 1

X → Layer 1 → Z_1

$$\mathbf{X} \cdot \mathbf{W_1} + \mathbf{b_1} = \mathbf{Z_1'}$$
$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} w_{11} & w_{13} \\ w_{12} & w_{14} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \end{bmatrix} = \begin{bmatrix} z'_{11} & z'_{12} \end{bmatrix}$$
$$\boldsymbol{\sigma}(\mathbf{Z_1'}) = \mathbf{Z_1}$$

Z_1 → Layer 2 → Z_2

$$\mathbf{Z_1} \cdot \mathbf{W_2} + \mathbf{b_2} = \mathbf{Z_2'}$$
$$\begin{bmatrix} z_{11} & z_{12} \end{bmatrix} \begin{bmatrix} w_{21} & w_{23} \\ w_{22} & w_{24} \end{bmatrix} + \begin{bmatrix} b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} z'_{21} & z'_{22} \end{bmatrix}$$
$$\boldsymbol{\sigma}(\mathbf{Z_2'}) = \mathbf{Z_2}$$

Z_2 → Layer 3 → ŷ

$$\mathbf{Z_2} \cdot \mathbf{W_3} + \mathbf{b_3} = \mathbf{\hat{y}'}$$
$$\begin{bmatrix} z_{21} & z_{22} \end{bmatrix} \begin{bmatrix} w_{31} \\ w_{32} \end{bmatrix} + \begin{bmatrix} b_3 \end{bmatrix} = \begin{bmatrix} \hat{y}' \end{bmatrix}$$
$$\boldsymbol{\sigma}(\mathbf{\hat{y}'}) = \mathbf{\hat{y}}$$

C. Backpropagation

為了能夠使網路趨近正確的分類，也就是最小化損失函數，在更新參數的時候計算loss對各個參數的偏微分；偏微分的方向也就是誤差擴大的方向，因此在更新權重的時候將其取反，進而減小誤差。

Loss Function: L = $(\hat{y}-y)^2/2$
Learning Rate: LR

$b_3' \rightarrow b_3 - LR \cdot \nabla_{b_3} L$
$w_3' \rightarrow w_3 - LR \cdot \nabla_{w_3} L$
$b_2' \rightarrow b_2 - LR \cdot \nabla_{b_2} L$
$w_2' \rightarrow w_2 - LR \cdot \nabla_{w_2} L$
$b_1' \rightarrow b_1 - LR \cdot \nabla_{b_1} L$
$w_1' \rightarrow w_1 - LR \cdot \nabla_{w_1} L$

Layer 3: b_3, W_3

$$\left[\frac{dL}{db_3}\right] = \left[(\hat{y} - y) \cdot y(1 - y)\right]$$

$$\begin{bmatrix} \frac{dL}{dw_{31}} \\ \frac{dL}{dw_{32}} \end{bmatrix} = \left[\frac{dL}{db_3}\right] \odot \begin{bmatrix} z_{21} & z_{22} \end{bmatrix}^T$$

$$\frac{dL}{db_3} = \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)} \frac{d(\hat{y}-y)}{d\hat{y}} \frac{d\hat{y}}{d\hat{y}'} \frac{d\hat{y}'}{db_3} = (\hat{y} - y) \cdot y(1 - y)$$

$$\frac{dL}{dw_{31}} = \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)} \frac{d(\hat{y}-y)}{d\hat{y}} \frac{d\hat{y}}{d\hat{y}'} \frac{d\hat{y}'}{dw_{31}} = (\hat{y} - y) \cdot y(1 - y) \cdot z_{21}$$

$$\frac{dL}{dw_{32}} = \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)} \frac{d(\hat{y}-y)}{d\hat{y}} \frac{d\hat{y}}{d\hat{y}'} \frac{d\hat{y}'}{dw_{32}} = (\hat{y} - y) \cdot y(1 - y) \cdot z_{22}$$

Layer 2: b_2, W_2

$$\begin{bmatrix} \frac{dL}{db_{21}} & \frac{dL}{db_{22}} \end{bmatrix} = \left[\frac{dL}{db_3}\right] \odot \begin{bmatrix} w_{31} \\ w_{32} \end{bmatrix}^T \odot \begin{bmatrix} z_{21}(1 - z_{21}) & z_{22}(1 - z_{22}) \end{bmatrix}$$

$$\begin{bmatrix} \frac{dL}{dw_{21}} & \frac{dL}{dw_{23}} \\ \frac{dL}{dw_{22}} & \frac{dL}{dw_{24}} \end{bmatrix} = \begin{bmatrix} \frac{dL}{db_{21}} & \frac{dL}{db_{22}} \\ \frac{dL}{db_{21}} & \frac{dL}{db_{22}} \end{bmatrix} \odot \begin{bmatrix} z_{11} & z_{11} \\ z_{12} & z_{12} \end{bmatrix}$$

$$\frac{dL}{db_{21}} = \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)} \frac{d(\hat{y}-y)}{d\hat{y}} \frac{d\hat{y}}{d\hat{y}'} \frac{d\hat{y}'}{dz_{21}} \frac{dz_{21}}{dz'_{21}} \frac{dz'_{21}}{db_{21}} = (\hat{y} - y) \cdot y(1 - y) \cdot w_{31} \cdot z_{21}(1 - z_{21})$$

$$\frac{dL}{dw_{21}} = \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)} \frac{d(\hat{y}-y)}{d\hat{y}} \frac{d\hat{y}}{d\hat{y}'} \frac{d\hat{y}'}{dz_{21}} \frac{dz_{21}}{dz'_{21}} \frac{dz'_{21}}{dw_{21}} = (\hat{y} - y) \cdot y(1 - y) \cdot w_{31} \cdot z_{21}(1 - z_{21}) \cdot z_{11}$$

$$\frac{dL}{dw_{22}} = \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)} \frac{d(\hat{y}-y)}{d\hat{y}} \frac{d\hat{y}}{d\hat{y}'} \frac{d\hat{y}'}{dz_{21}} \frac{dz_{21}}{dz'_{21}} \frac{dz'_{21}}{dw_{22}} = (\hat{y} - y) \cdot y(1 - y) \cdot w_{31} \cdot z_{21}(1 - z_{21}) \cdot z_{12}$$

$$\frac{dL}{db_{22}} = \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)} \frac{d(\hat{y}-y)}{d\hat{y}} \frac{d\hat{y}}{d\hat{y}'} \frac{d\hat{y}'}{dz_{22}} \frac{dz_{22}}{dz'_{22}} \frac{dz'_{22}}{db_{22}} = (\hat{y} - y) \cdot y(1 - y) \cdot w_{32} \cdot z_{22}(1 - z_{22})$$

$$\frac{dL}{dw_{23}} = \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)} \frac{d(\hat{y}-y)}{d\hat{y}} \frac{d\hat{y}}{d\hat{y}'} \frac{d\hat{y}'}{dz_{22}} \frac{dz_{22}}{dz'_{22}} \frac{dz'_{22}}{dw_{23}} = (\hat{y} - y) \cdot y(1 - y) \cdot w_{32} \cdot z_{22}(1 - z_{22}) \cdot z_{11}$$

$$\frac{dL}{dw_{24}} = \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)} \frac{d(\hat{y}-y)}{d\hat{y}} \frac{d\hat{y}}{d\hat{y}'} \frac{d\hat{y}'}{dz_{22}} \frac{dz_{22}}{dz'_{22}} \frac{dz'_{22}}{dw_{24}} = (\hat{y} - y) \cdot y(1 - y) \cdot w_{32} \cdot z_{22}(1 - z_{22}) \cdot z_{12}$$

Layer 1: b_1, W_1

$$\begin{bmatrix} \frac{dL}{db_{11}} & \frac{dL}{db_{12}} \end{bmatrix} = \begin{bmatrix} \frac{dL}{db_{21}} & \frac{dL}{db_{22}} \end{bmatrix} \begin{bmatrix} w_{21} & w_{23} \\ w_{22} & w_{24} \end{bmatrix} \odot \begin{bmatrix} z_{11}(1 - z_{11}) & z_{12}(1 - z_{12}) \end{bmatrix}$$

$$\begin{bmatrix} \frac{dL}{dw_{11}} & \frac{dL}{dw_{13}} \\ \frac{dL}{dw_{12}} & \frac{dL}{dw_{14}} \end{bmatrix} = \begin{bmatrix} \frac{dL}{db_{11}} & \frac{dL}{db_{12}} \\ \frac{dL}{db_{11}} & \frac{dL}{db_{12}} \end{bmatrix} \odot \begin{bmatrix} x_1 & x_1 \\ x_2 & x_2 \end{bmatrix}$$

$$\begin{aligned} \frac{dL}{db_{11}} &= \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)} \frac{d(\hat{y}-y)}{d\hat{y}} \frac{d\hat{y}}{d\hat{y}'} \frac{d\hat{y}'}{dz_{21}} \frac{dz_{21}}{dz'_{21}} \frac{dz'_{21}}{dz_{11}} \frac{dz_{11}}{dz'_{11}} \frac{dz'_{11}}{db_{11}} + \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)} \frac{d(\hat{y}-y)}{d\hat{y}} \frac{d\hat{y}}{d\hat{y}'} \frac{d\hat{y}'}{dz_{22}} \frac{dz_{22}}{dz'_{22}} \frac{dz'_{22}}{dz_{11}} \frac{dz_{11}}{dz'_{11}} \frac{dz'_{11}}{db_{11}} \\ &= (\hat{y} - y) \cdot y(1 - y) \cdot [w_{31} \cdot z_{21}(1 - z_{21}) \cdot w_{21} + w_{32} \cdot z_{22}(1 - z_{22}) \cdot w_{23}] \cdot z_{11}(1 - z_{11}) \end{aligned}$$

$$\frac{dL}{dw_{11}} = \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)}\frac{d(\hat{y}-y)}{d\hat{y}}\frac{d\hat{y}}{d\hat{y}'}\frac{d\hat{y}'}{dz_{21}}\frac{dz_{21}}{dz'_{21}}\frac{dz'_{21}}{dz_{11}}\frac{dz_{11}}{dz'_{11}}\frac{dz'_{11}}{dw_{11}} + \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)}\frac{d(\hat{y}-y)}{d\hat{y}}\frac{d\hat{y}}{d\hat{y}'}\frac{d\hat{y}'}{dz_{22}}\frac{dz_{22}}{dz'_{22}}\frac{dz'_{22}}{dz_{11}}\frac{dz_{11}}{dz'_{11}}\frac{dz'_{11}}{dw_{11}}$$

$$= (\hat{y} - y) \cdot y(1 - y) \cdot [w_{31} \cdot z_{21}(1 - z_{21}) \cdot w_{21} + w_{32} \cdot z_{22}(1 - z_{22}) \cdot w_{23}] \cdot z_{11}(1 - z_{11}) \cdot$$

$$\frac{dL}{dw_{12}} = \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)}\frac{d(\hat{y}-y)}{d\hat{y}}\frac{d\hat{y}}{d\hat{y}'}\frac{d\hat{y}'}{dz_{21}}\frac{dz_{21}}{dz'_{21}}\frac{dz'_{21}}{dz_{11}}\frac{dz_{11}}{dz'_{11}}\frac{dz'_{11}}{dw_{12}} + \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)}\frac{d(\hat{y}-y)}{d\hat{y}}\frac{d\hat{y}}{d\hat{y}'}\frac{d\hat{y}'}{dz_{22}}\frac{dz_{22}}{dz'_{22}}\frac{dz'_{22}}{dz_{11}}\frac{dz_{11}}{dz'_{11}}\frac{dz'_{11}}{dw_{12}}$$

$$= (\hat{y} - y) \cdot y(1 - y) \cdot [w_{31} \cdot z_{21}(1 - z_{21}) \cdot w_{21} + w_{32} \cdot z_{22}(1 - z_{22}) \cdot w_{23}] \cdot z_{11}(1 - z_{11}) \cdot$$

$$\frac{dL}{db_{12}} = \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)}\frac{d(\hat{y}-y)}{d\hat{y}}\frac{d\hat{y}}{d\hat{y}'}\frac{d\hat{y}'}{dz_{21}}\frac{dz_{21}}{dz'_{21}}\frac{dz'_{21}}{dz_{12}}\frac{dz_{12}}{dz'_{12}}\frac{dz'_{12}}{db_{12}} + \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)}\frac{d(\hat{y}-y)}{d\hat{y}}\frac{d\hat{y}}{d\hat{y}'}\frac{d\hat{y}'}{dz_{22}}\frac{dz_{22}}{dz'_{22}}\frac{dz'_{22}}{dz_{12}}\frac{dz_{12}}{dz'_{12}}\frac{dz'_{12}}{db_{12}}$$

$$= (\hat{y} - y) \cdot y(1 - y) \cdot [w_{31} \cdot z_{21}(1 - z_{21}) \cdot w_{22} + w_{32} \cdot z_{22}(1 - z_{22}) \cdot w_{24}] \cdot z_{12}(1 - z_{12})$$

$$\frac{dL}{dw_{13}} = \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)}\frac{d(\hat{y}-y)}{d\hat{y}}\frac{d\hat{y}}{d\hat{y}'}\frac{d\hat{y}'}{dz_{21}}\frac{dz_{21}}{dz'_{21}}\frac{dz'_{21}}{dz_{12}}\frac{dz_{12}}{dz'_{12}}\frac{dz'_{12}}{dw_{13}} + \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)}\frac{d(\hat{y}-y)}{d\hat{y}}\frac{d\hat{y}}{d\hat{y}'}\frac{d\hat{y}'}{dz_{22}}\frac{dz_{22}}{dz'_{22}}\frac{dz'_{22}}{dz_{12}}\frac{dz_{12}}{dz'_{12}}\frac{dz'_{12}}{dw_{13}}$$

$$= (\hat{y} - y) \cdot y(1 - y) \cdot [w_{31} \cdot z_{21}(1 - z_{21}) \cdot w_{22} + w_{32} \cdot z_{22}(1 - z_{22}) \cdot w_{24}] \cdot z_{12}(1 - z_{12}) \cdot x$$

$$\frac{dL}{dw_{14}} = \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)}\frac{d(\hat{y}-y)}{d\hat{y}}\frac{d\hat{y}}{d\hat{y}'}\frac{d\hat{y}'}{dz_{21}}\frac{dz_{21}}{dz'_{21}}\frac{dz'_{21}}{dz_{12}}\frac{dz_{12}}{dz'_{12}}\frac{dz'_{12}}{dw_{14}} + \frac{d(\hat{y}-y)^2/2}{d(\hat{y}-y)}\frac{d(\hat{y}-y)}{d\hat{y}}\frac{d\hat{y}}{d\hat{y}'}\frac{d\hat{y}'}{dz_{22}}\frac{dz_{22}}{dz'_{22}}\frac{dz'_{22}}{dz_{12}}\frac{dz_{12}}{dz'_{12}}\frac{dz'_{12}}{dw_{14}}$$

$$= (\hat{y} - y) \cdot y(1 - y) \cdot [w_{31} \cdot z_{21}(1 - z_{21}) \cdot w_{22} + w_{32} \cdot z_{22}(1 - z_{22}) \cdot w_{24}] \cdot z_{12}(1 - z_{12}) \cdot x$$

3. Results of your testing

A. Screenshot and comparison figure

兩種不同的數據經由神經網路得出的圖和正解相同。

Linear

n=100

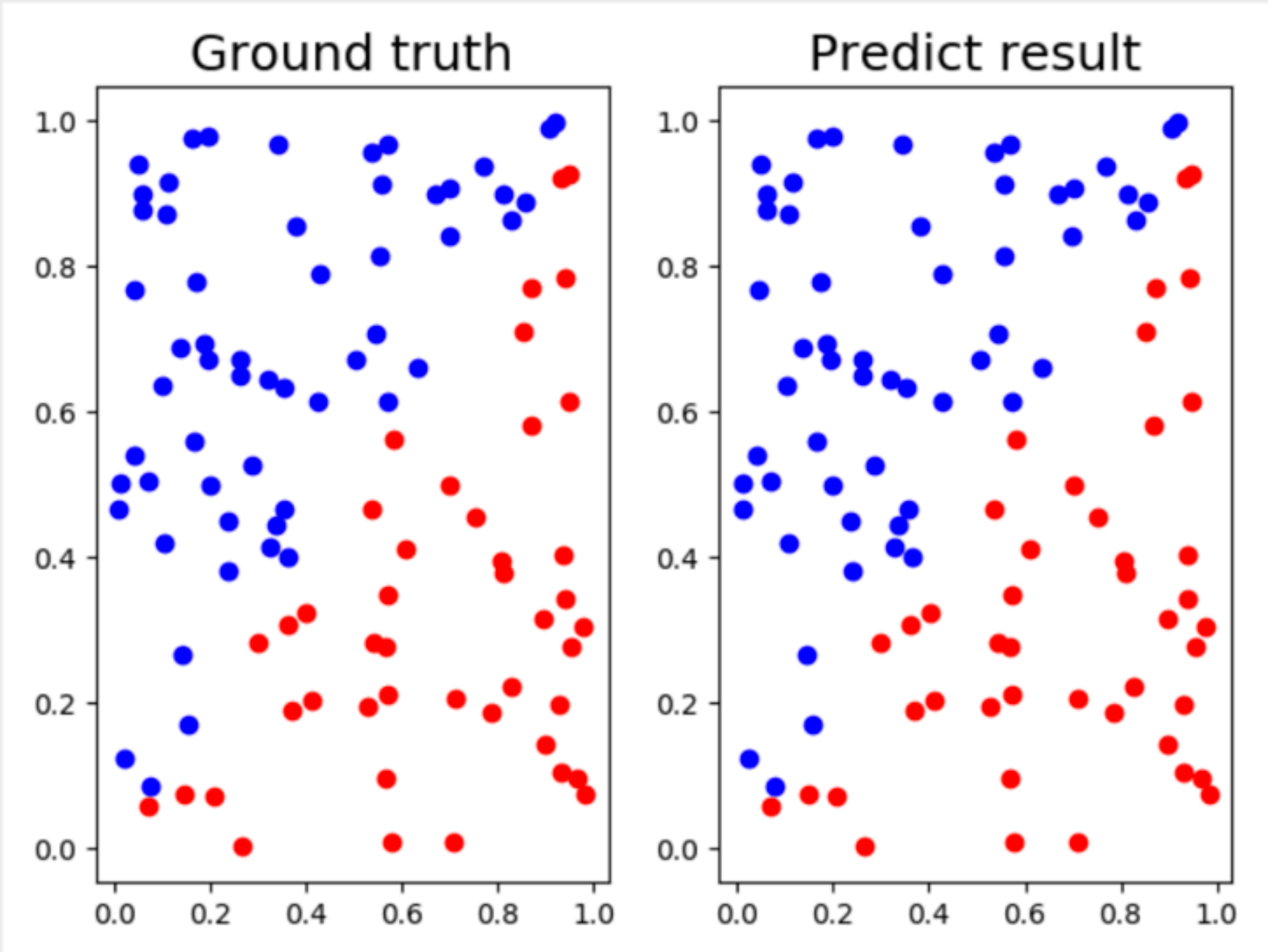

Figure 2. Linear(n=100) comparison figure

n=200

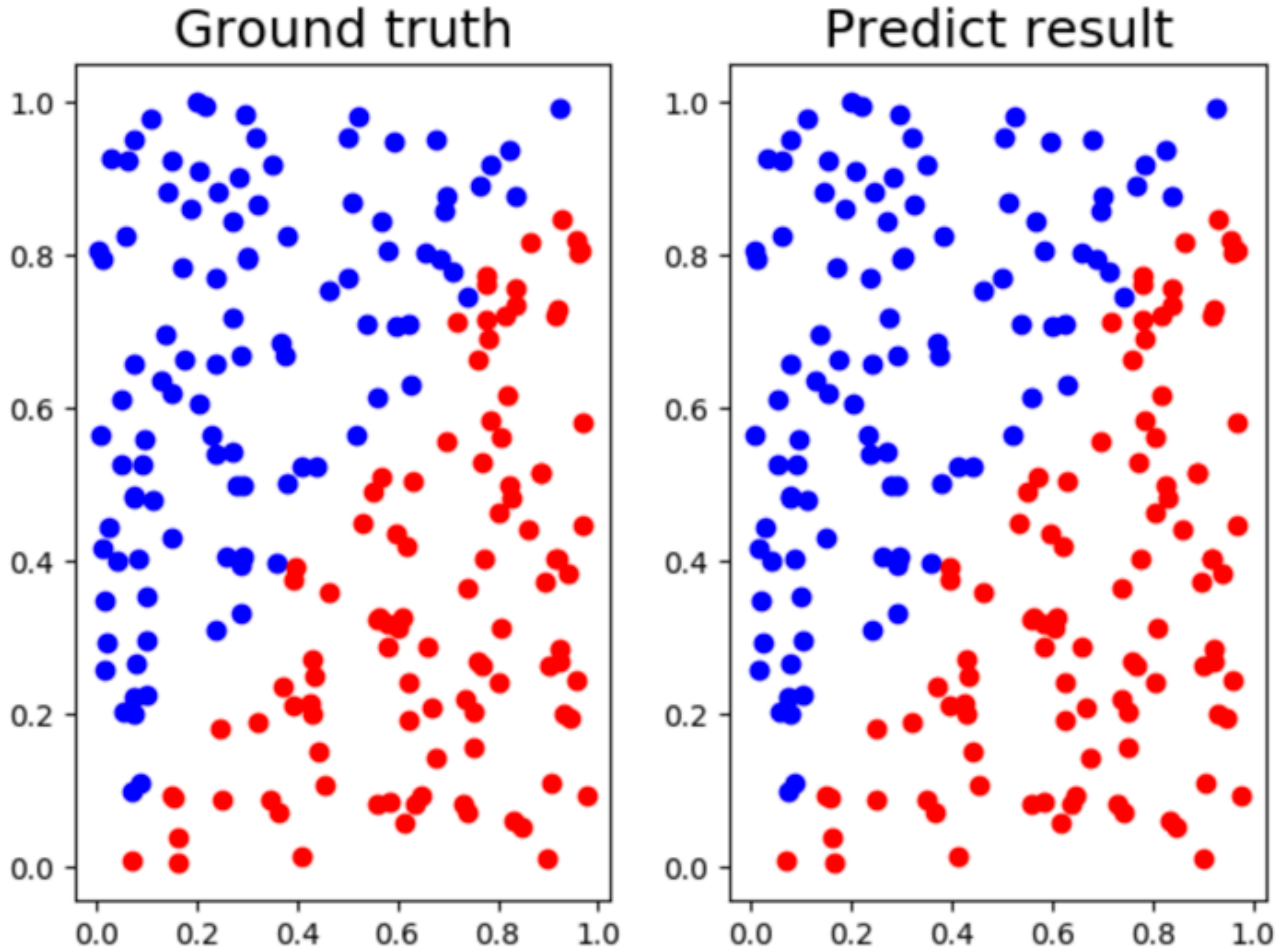

Figure 3. Linear(n=200) comparison figure

n=300

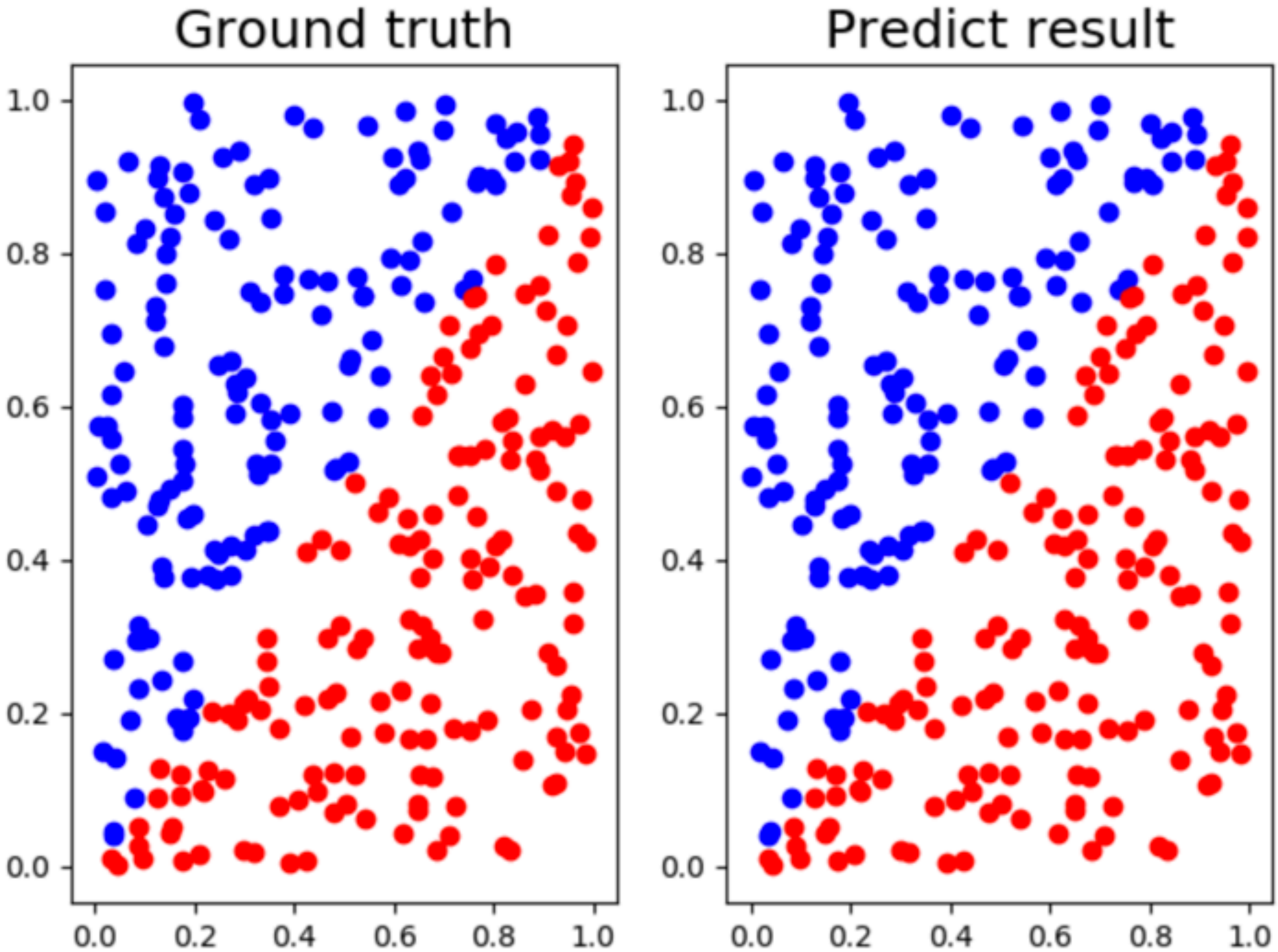

Figure 4. Linear(n=300) comparison figure

XOR easy

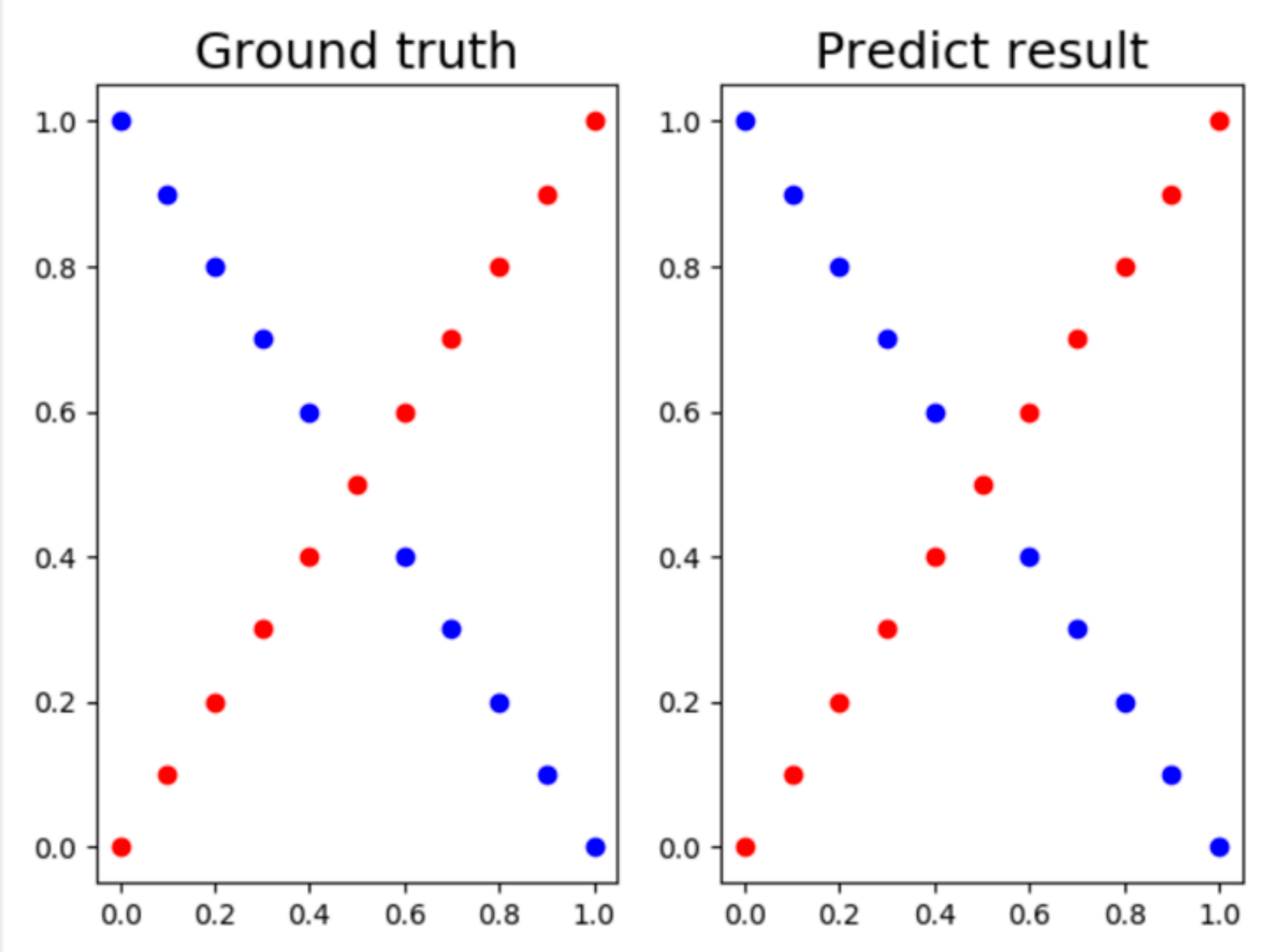

Figure 5. XOR comparison figure

B. anything you want to present

loss都有逐漸收斂，且能在五分鐘內得到accurancy為1的結果。

Linear

n=100

```
epoch  100 , loss: 0.08547551570216562 , accuracy: 0.95
epoch  200 , loss: 0.04584292443123324 , accuracy: 0.93
epoch  300 , loss: 0.031990660080388186 , accuracy: 0.94
epoch  400 , loss: 0.026038416716969 , accuracy: 0.96
epoch  500 , loss: 0.02279215932047285 , accuracy: 0.97
epoch  600 , loss: 0.020685318319860008 , accuracy: 0.97
epoch  700 , loss: 0.019243170794214875 , accuracy: 0.97
epoch  800 , loss: 0.014468983120083774 , accuracy: 0.98
epoch  900 , loss: 0.010442449131257405 , accuracy: 1.0
```

Figure 6. Linear(n=100) every 100 epoch

```
[0.985895  ]
[0.99448441]
[0.99476133]
```

```
[[0.57944896]
 [0.00326583]
 [0.98996394]
 [0.00447334]
 [0.99482258]
 [0.54961646]
 [0.00305579]
 [0.99364033]
 [0.99462867]
 [0.99482088]
 [0.9934419 ]
 [0.9946459 ]
 [0.03009271]
 [0.99458925]
 [0.99482446]
 [0.80142879]
 [0.99455495]
 [0.00371178]
 [0.00699817]
 [0.9805997 ]
 [0.03008235]
 [0.99429515]
 [0.99308512]
 [0.18810642]
 [0.99465878]
 [0.00348263]
 [0.99136112]
 [0.0050832 ]
 [0.01284434]
 [0.00318809]
 [0.00303342]
 [0.99402188]
 [0.00335178]
 [0.99431153]
 [0.00303148]
 [0.96970327]
 [0.00368404]
 [0.95155446]
 [0.00411079]
 [0.02592289]
 [0.96766643]
 [0.99481625]
 [0.00304417]
 [0.99481871]
 [0.00456552]
 [0.00306641]
 [0.00796165]
 [0.92116277]
 [0.99471765]
 [0.98611909]
 [0.9946984 ]
 [0.99440285]
 [0.98551865]
 [0.99089942]
 [0.7049236 ]
 [0.99482208]
 [0.00333486]
 [0.1967381 ]
 [0.99438374]
 [0.00351488]
 [0.99477264]
 [0.99459505]
 [0.14812622]
 [0.00437636]
 [0.00304164]
 [0.00353382]
 [0.97067375]
 [0.00307487]
 [0.93195056]
 [0.99407309]
 [0.00309477]
 [0.05334752]
 [0.98242724]
 [0.00310182]
 [0.99246072]
 [0.00581135]
 [0.77135073]
 [0.00306866]
 [0.97772857]
 [0.99373881]
 [0.98880515]
 [0.00303589]
 [0.67101004]
 [0.00307153]
 [0.00304523]
 [0.28208787]
 [0.99427314]
 [0.99203194]
 [0.99480918]
 [0.65182459]
 [0.93880434]
 [0.99482763]
 [0.00306609]
 [0.22737728]
 [0.00316201]
 [0.00305396]
 [0.00311947]]
```

Figure 7. Linear(n=100) predictions

n=200

```
epoch  100 , loss: 0.08890036836013131 , accuracy: 0.865
epoch  200 , loss: 0.026992096001925564 , accuracy: 0.95
epoch  300 , loss: 0.014888744725014589 , accuracy: 0.985
epoch  400 , loss: 0.013158035634809311 , accuracy: 0.985
epoch  500 , loss: 0.012012270233177023 , accuracy: 0.985
epoch  600 , loss: 0.010637313632673959 , accuracy: 0.99
epoch  700 , loss: 0.009635029980840638 , accuracy: 0.99
epoch  800 , loss: 0.009006091039373502 , accuracy: 0.99
epoch  900 , loss: 0.008587480745759098 , accuracy: 0.99
epoch  1000 , loss: 0.008291947849294346 , accuracy: 0.99
epoch  1100 , loss: 0.008074219840476808 , accuracy: 0.99
epoch  1200 , loss: 0.007908465443846468 , accuracy: 0.99
epoch  1300 , loss: 0.00777879226354557 , accuracy: 0.99
epoch  1400 , loss: 0.007604511059969395 , accuracy: 0.99
epoch  1500 , loss: 0.0056124273218498685 , accuracy: 0.99
epoch  1600 , loss: 0.005107483116779407 , accuracy: 0.99
epoch  1700 , loss: 0.0048013425006074315 , accuracy: 0.99
epoch  1800 , loss: 0.012227897644666544 , accuracy: 0.985
epoch  1900 , loss: 0.005811896081299477 , accuracy: 0.985
epoch  2000 , loss: 0.008896703946869361 , accuracy: 0.99
epoch  2100 , loss: 0.00519470220660048 , accuracy: 0.99
epoch  2200 , loss: 0.0043086886045021885 , accuracy: 0.99
epoch  2300 , loss: 0.006037273793492568 , accuracy: 0.99
epoch  2400 , loss: 0.003619681064834265 , accuracy: 1.0
```

Figure 8. Linear(n=200) every 100 epoch

```
[[0.99960414]
 [0.00185044]
 [0.00187194]
 [0.00185148]
 [0.0018941 ]
 [0.00321763]
 [0.99776443]
 [0.99959091]
 [0.00197385]
 [0.00720667]
 [0.00186328]
 [0.00190968]
 [0.99948986]
 [0.00777061]
 [0.00273008]
 [0.00359316]
```

```
[0.00184892]
[0.99960354]
[0.99943866]
[0.00184633]
[0.99955   ]
[0.00196653]
[0.00184114]
[0.99959688]
[0.99957091]
[0.00183823]
[0.00202538]
[0.99953867]
[0.00462192]
[0.99959914]
[0.99873665]
[0.00850586]
```

```
[0.99958155]
[0.00225087]
[0.99960388]
[0.99960538]
[0.99959273]
[0.14273038]
[0.99934534]
[0.99751241]
[0.00184986]
[0.00191276]
[0.99960474]
[0.00184123]
[0.00184582]
[0.00205273]
[0.00193781]
[0.99960017]
```

```
[0.99960191]
[0.99919743]
[0.00932567]
[0.00205019]
[0.99959112]
[0.99960376]
[0.00222847]
[0.99958244]
[0.99065629]
[0.99958896]
[0.99960072]
[0.99953242]
[0.99349788]
[0.99959932]
[0.64539629]
[0.00187485]
[0.47193194]
[0.99960537]
[0.0018845 ]
[0.00184582]
[0.99926643]
[0.99960411]
[0.99960279]
[0.64929785]
[0.00247967]
[0.00187642]
[0.00223941]
```

```
[0.97675953]
[0.99924897]
[0.99922518]
[0.99960276]
[0.00187521]
[0.99960526]
[0.99958186]
[0.00191084]
[0.99945728]
[0.00183909]
[0.99959628]
[0.99959912]
[0.99959341]
[0.99958396]
[0.00256835]
[0.3530673 ]
[0.99960129]
[0.99960533]
[0.00188236]
[0.00186614]
[0.11369065]
[0.99959944]
[0.0022619 ]
[0.00185539]
[0.0098149 ]
[0.99960413]
[0.99957605]
```

```
[0.99960259]
[0.99935992]
[0.00187543]
[0.99936612]
[0.00183703]
[0.00187113]
[0.00343263]
[0.00184042]
[0.99920595]
[0.99958147]
[0.99959991]
[0.9996038 ]
[0.00453787]
[0.99955802]
[0.99960162]
[0.97043933]
[0.00188817]
[0.99879133]
[0.0018379 ]
[0.99960539]
[0.99958654]
[0.00197004]
[0.99938838]
[0.00184557]
[0.00723229]
[0.00183947]
[0.99960544]
[0.99960129]
```

```
[0.00243234]
[0.00424071]
[0.00193987]
[0.00185276]
[0.00184909]
[0.99960513]
[0.9995066 ]
[0.00185993]
[0.00184972]
[0.00250675]
[0.00198448]
[0.99625744]
[0.00191358]
[0.99933819]
[0.00197946]
```

```
[0.00185192]
[0.00207579]
[0.99960452]
[0.99951468]
[0.00184189]
[0.99956214]
```

```
[0.99958696]
[0.99960176]
[0.99960497]
[0.99960558]
[0.00263641]
[0.00183869]
[0.00184352]
[0.00211419]
[0.9994562 ]
[0.0018464 ]
[0.99960543]
[0.98674944]
[0.01280576]
[0.99960197]
[0.30255286]
[0.99960388]
[0.00187466]
[0.99959776]
[0.9937958 ]
[0.99960557]
[0.00209025]
[0.01302411]
[0.00183791]
[0.99960461]
[0.99960556]
[0.99959582]
[0.00205932]
[0.9981817 ]
[0.00185042]
[0.99926964]
[0.99958321]
[0.99960302]
[0.99959599]
[0.9996056 ]
[0.00183699]
[0.00211717]
[0.00184183]
[0.99944697]
[0.00184738]
[0.00184568]
[0.00226325]
[0.99960213]
[0.00373127]
[0.99960474]
[0.00192116]
[0.99960506]
[0.00314322]
[0.00227149]
[0.00184948]]
```

Figure 9. Linear(n=200) predictions

n=300

```
epoch  100 , loss: 0.13084881786847166 , accuracy: 0.8333333333333334
epoch  200 , loss: 0.05366768997613396 , accuracy: 0.93
epoch  300 , loss: 0.03441418573499303 , accuracy: 0.9466666666666667
epoch  400 , loss: 0.028501755628121378 , accuracy: 0.95
epoch  500 , loss: 0.025992080413497844 , accuracy: 0.9533333333333334
epoch  600 , loss: 0.02416754716475541 , accuracy: 0.9566666666666667
epoch  700 , loss: 0.02243977189687639 , accuracy: 0.9666666666666667
epoch  800 , loss: 0.020676838678029875 , accuracy: 0.9766666666666667
epoch  900 , loss: 0.01884772129251868 , accuracy: 0.9766666666666667
epoch  1000 , loss: 0.017265415586333172 , accuracy: 0.9766666666666667
epoch  1100 , loss: 0.01613334888118305 , accuracy: 0.9766666666666667
epoch  1200 , loss: 0.01527826960568187 , accuracy: 0.9766666666666667
epoch  1300 , loss: 0.014544652628895894 , accuracy: 0.9766666666666667
epoch  1400 , loss: 0.013860101104121829 , accuracy: 0.9766666666666667
epoch  1500 , loss: 0.01320618517866626 , accuracy: 0.9766666666666667
epoch  1600 , loss: 0.012584237212191589 , accuracy: 0.9766666666666667
epoch  1700 , loss: 0.011992214047200278 , accuracy: 0.9766666666666667
epoch  1800 , loss: 0.01141805317267302 , accuracy: 0.98
epoch  1900 , loss: 0.010834915447091085 , accuracy: 0.9833333333333333
epoch  2000 , loss: 0.008903004919638543 , accuracy: 0.99
epoch  2100 , loss: 0.0024420313563401717 , accuracy: 1.0
```

Figure 10. Linear(n=300) every 100 epoch

```
[[0.998677  ]
 [0.00287783]
 [0.0028599 ]
 [0.00285878]
 [0.00289228]
 [0.00322454]
 [0.99859688]
 [0.99867698]
 [0.99867596]
 [0.00285807]
 [0.00340038]
 [0.00285854]
 [0.00676713]
 [0.99867677]
 [0.00287141]
 [0.99865589]
 [0.00286399]
 [0.7461853 ]
 [0.99867589]
 [0.00285902]
 [0.99867701]
 [0.0028581 ]
 [0.9986753 ]
 [0.99867304]
 [0.00285846]
 [0.0028644 ]
 [0.00286351]
 [0.00285811]
 [0.99863678]
 [0.99867663]
 [0.99867695]
 [0.99867701]
 [0.99866647]
 [0.0028582 ]
 [0.99864045]
 [0.00285807]
 [0.00896676]
 [0.00285847]
 [0.00285826]
 [0.00285881]
 [0.00285807]
 [0.00291206]
 [0.00288054]
 [0.00326369]
 [0.99848031]
```

```
 [0.99867697]
 [0.00286017]
 [0.98828796]
 [0.9986606 ]
 [0.99737952]
 [0.99867698]
 [0.0028588 ]
 [0.99867243]
 [0.00285809]
 [0.01547998]
 [0.00285807]
 [0.00285893]
 [0.00285868]
 [0.00309859]
 [0.99867701]
 [0.00325918]
 [0.00294575]
 [0.99867701]
 [0.99867701]
 [0.99867693]
 [0.99862491]
 [0.99863629]
 [0.998562  ]
 [0.00286326]
 [0.00285904]
 [0.00285873]
 [0.9255348 ]
 [0.01379405]
 [0.00286202]
 [0.00286262]
 [0.99867698]
 [0.00339273]
 [0.00285852]
 [0.00285947]
 [0.00285808]
 [0.99814547]
 [0.99867358]
 [0.00306578]
 [0.00286482]
 [0.00287526]
 [0.0031708 ]
 [0.99865016]
 [0.00285808]
 [0.99867699]
 [0.00285948]
```

```
[[0.998677  ]
 [0.00287783]
 [0.0028599 ]
 [0.00285878]
 [0.00289228]
 [0.00322454]
 [0.99859688]
 [0.99867698]
 [0.99867596]
 [0.00285807]
 [0.00340038]
 [0.00285854]
 [0.00676713]
 [0.99867677]
 [0.00287141]
 [0.99865589]
 [0.00286399]
 [0.7461853 ]
 [0.99867589]
 [0.00285902]
 [0.99867701]
 [0.0028581 ]
 [0.9986753 ]
 [0.99867304]
 [0.00285846]
 [0.0028644 ]
 [0.00286351]
 [0.00285811]
 [0.99863678]
 [0.99867663]
 [0.99867695]
 [0.99867701]
 [0.99866647]
 [0.0028582 ]
 [0.99864045]
 [0.00285807]
 [0.00896676]
 [0.00285847]
 [0.00285826]
 [0.00285881]
 [0.00285807]
 [0.00291206]
 [0.00288054]
 [0.00326369]
 [0.99848031]
```

```
 [0.00309053]
 [0.00285932]
 [0.99856357]
 [0.0214775 ]
 [0.99838475]
 [0.99867433]
 [0.99864658]
 [0.00285817]
 [0.99867698]
 [0.9986736 ]
 [0.00468166]
 [0.99867679]
 [0.99867698]
 [0.99867701]
 [0.99867351]
 [0.99866546]
 [0.65813665]
 [0.99867605]
 [0.00285833]
 [0.00285907]
 [0.998677  ]
 [0.99861342]
 [0.00286769]
 [0.99867398]
 [0.00322278]
 [0.01180017]
 [0.00285893]
 [0.00994759]
 [0.99867701]
 [0.00285807]
 [0.00286067]
 [0.00285811]
 [0.0028582 ]
```

```
[0.9985073 ]   [0.99867672]   [0.9985073 ]   [0.998677   ]
[0.99817028]   [0.96990124]   [0.99817028]   [0.99867659]
[0.00285807]   [0.99867592]   [0.00285807]   [0.00329805]
[0.99865691]   [0.00507271]   [0.99865691]   [0.00285849]
[0.99867697]   [0.00285857]   [0.99867697]   [0.00303394]
[0.99859837]   [0.9986764 ]   [0.99859837]   [0.00316579]
[0.99862585]   [0.00285978]   [0.99862585]   [0.00292711]
[0.0028585 ]   [0.00286664]   [0.0028585 ]   [0.0028819 ]
[0.99867685]   [0.99866503]   [0.99867685]   [0.99867669]
[0.00285807]   [0.00285809]   [0.00285807]   [0.99858925]
[0.92900068]   [0.99867612]   [0.92900068]   [0.0028656 ]
[0.998677  ]   [0.99863761]   [0.998677  ]   [0.998677   ]
[0.99846548]   [0.99867699]   [0.99846548]   [0.00787714]
[0.00285814]   [0.00285807]   [0.00285814]   [0.99857926]
[0.99867699]   [0.60378426]   [0.99867699]   [0.00285811]
[0.99867631]   [0.89250016]   [0.99867631]   [0.998677   ]
[0.99802668]   [0.00285808]   [0.99802668]   [0.00286511]
[0.99867696]   [0.00311846]   [0.99867696]   [0.07757526]
[0.00647291]   [0.0028729 ]   [0.00647291]   [0.002868   ]
[0.99866113]   [0.00286146]   [0.99866113]   [0.99866707]
[0.99867692]   [0.0028592 ]   [0.99867692]   [0.99867701]
[0.00286031]   [0.99860104]   [0.00286031]   [0.99864829]
[0.99866334]   [0.00285807]   [0.99866334]   [0.99867588]
[0.99867492]   [0.00286056]   [0.99867492]   [0.0028596 ]]
[0.99228369]   [0.0061976 ]   [0.99228369]
[0.99866923]   [0.99867591]   [0.99866923]
[0.99867499]   [0.00285817]   [0.99867499]
[0.99867598]   [0.99867651]   [0.99867598]
[0.99867582]   [0.00299789]   [0.99867582]
[0.77930607]   [0.01544525]   [0.77930607]
[0.99867566]   [0.00285839]   [0.99867566]
[0.99867472]   [0.00285807]   [0.99867472]
[0.00286443]   [0.67927482]   [0.00286443]
[0.99855785]   [0.99866559]   [0.99855785]
[0.00285941]   [0.00286958]   [0.00285941]
[0.99849385]   [0.00285824]   [0.99849385]
```

Figure 11. Linear(n=300) predictions

XOR easy

```
epoch  100 , loss: 0.24939993235324523 , accuracy: 0.5238095238095238
epoch  200 , loss: 0.2493497517157231 , accuracy: 0.5238095238095238
epoch  300 , loss: 0.24911392410382793 , accuracy: 0.5238095238095238
epoch  400 , loss: 0.24447481357267853 , accuracy: 0.6666666666666666
epoch  500 , loss: 0.19962872540485876 , accuracy: 0.7142857142857143
epoch  600 , loss: 0.17716046289929985 , accuracy: 0.7619047619047619
epoch  700 , loss: 0.16878560978907692 , accuracy: 0.7619047619047619
epoch  800 , loss: 0.011324921259608847 , accuracy: 1.0
```
Figure 12. XOR every 100 epoch

```
[[0.09160782]
 [0.96898961]
 [0.08977842]
 [0.96898595]
 [0.0880281 ]
 [0.9689505 ]
 [0.08635308]
 [0.96851306]
 [0.08474974]
 [0.93176274]
 [0.08321464]
 [0.08174451]
 [0.62736796]
 [0.08033624]
 [0.92231779]
 [0.07898687]
 [0.9471187 ]
 [0.07769361]
 [0.95155104]
 [0.07645379]
 [0.95274033]]
```
Figure 13. XOR predictions

4. Discussion

A. Anything you want to share

影響收斂速度的因素

- 初始的參數：一開始我是將所有的參數取[0,1]的random數，但會發現每次loss收斂速度相差極大，而不易去了解和調整其他可能會造成收斂速度的變數。故後來使用RandomState來排除這個不確定性。

- Batch：比起一次forward就更新一次參數，一次epoch再更新一次參數能夠提升收斂速度。其原因可能為若每一次forward就更新會使得參數偏向某個數據，而容易產生震盪。
- Learing rate：兩個數據的LR設定不同，能夠造成對其較佳的結果。

程式會需要注意的錯誤

- Matrix shape：每個matrix都要好好想清楚和仔細檢查打出來的shape，否則會發生loss越來越大的慘劇。