HW1

李承穎
0856708

1.Introduction

- Multilayer perceptron is also a kind of NN with fully connected.
- Activation function is to make the linear model change to non-linear model, it is the main point of NN.
- For training an Neuron Network, we always need to use optimal function and Backpropagation.
- Optimal function is take a good way to minimize the loss function.(Here we use gradient descent.)
- Backpropagation is a way to compute the gradient of loss function to update the bias and weight.

2.Experiment setups

A. Sigmoid functions

1. It is a activation function
2. Formula:

$$S(x) = \frac{1}{1 + e^{-x}}$$

2.Experiment setups

B. Neural network

- Inputs : X is a (n,2).
- First hidden layer : neuron number = 4, $z1 = x * w1$, $a1 = \sigma(z1)$
- Second hidden layer : neuron number = 4, $z2 = a1 * w2$, $a2 = \sigma(z2)$
- Output layer : y_pred, $z3 = a2 * w3$, $ypred = a3 = \sigma\ z3$
- Weight 1 : matrix (2, 4)
- Weight 2 : matrix (4, 4)
- Weigth 3 : matrix (4, 1)

2.Experiment setups

C. Backpropagation

Loss function: $MSE = \frac{\sum_1^n (y_{pred} - y)^2}{n} = L(\theta)$

Gradient of weight1 =

$$\nabla(w_3) = \frac{\partial L}{\partial w_3} = \frac{\partial L}{\partial y_{pred}} \frac{\partial y_{pred}}{\partial z_3} \frac{\partial z_3}{\partial w_3}$$
$$= (y - y_{pred}) * diffsigmoid(z_3) * a_2$$
$$= \frac{\partial L}{\partial z_3} * a_2$$

Update: $w_i = w_i - lr * \nabla(w_i)$

Gradient of weight2 =

$$\nabla(w_2) = \frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial y_{pred}} \frac{\partial y_{pred}}{\partial z_3} \frac{\partial z_3}{\partial a_2} \frac{\partial a_2}{\partial z_2} \frac{\partial z_2}{\partial w_2}$$
$$= \frac{\partial L}{\partial z_3} * w_3 * diffsigmoid(z_2) * a_1$$
$$= \frac{\partial L}{\partial z_2} * a_1$$

$diffsigmoid(z_i) = sigmoid(z_i) * (1 - sigmoid(z_i))$

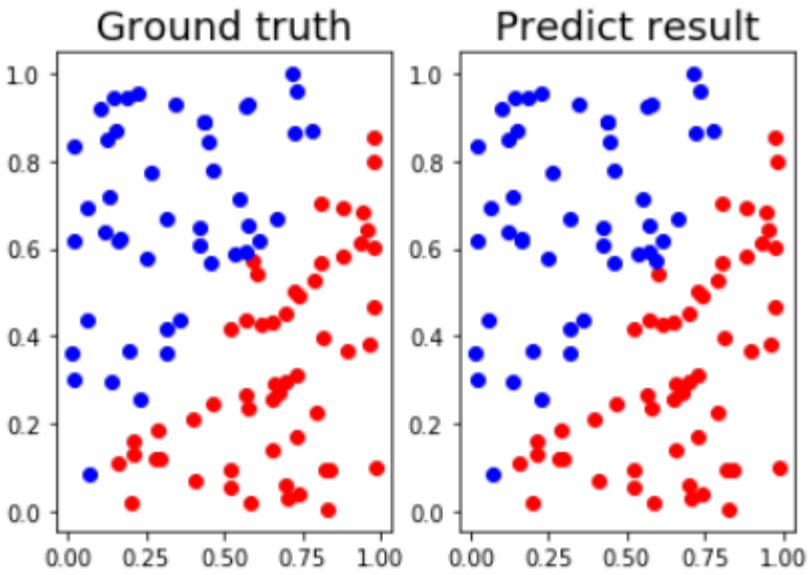
Gradient of weight1 =

$$\nabla(w_1) = \frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial z_2} \frac{\partial z_2}{\partial a_1} \frac{\partial a_1}{\partial z_1} \frac{\partial z_1}{\partial w_1}$$
$$= \frac{\partial L}{\partial z_2} * w_2 * diffsigmoid(z_1) * a_1$$

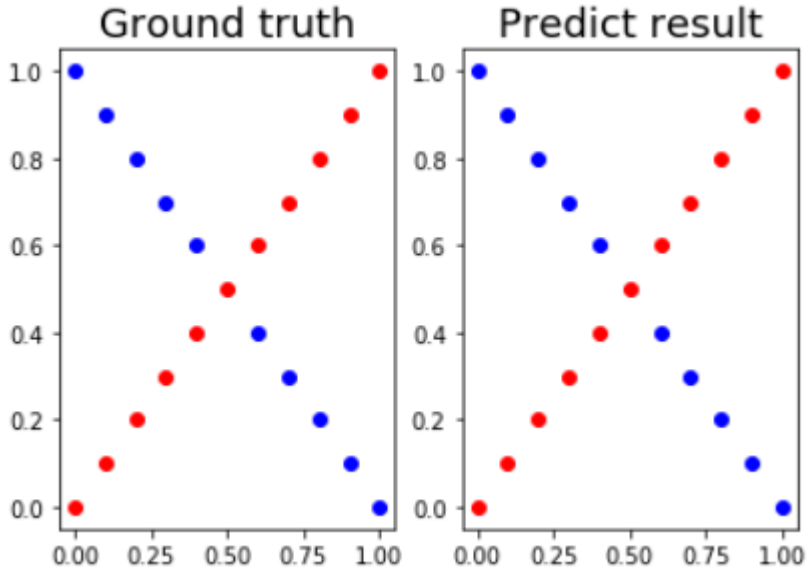
3.Results of your testing

A. Screenshot and comparison figure

linear



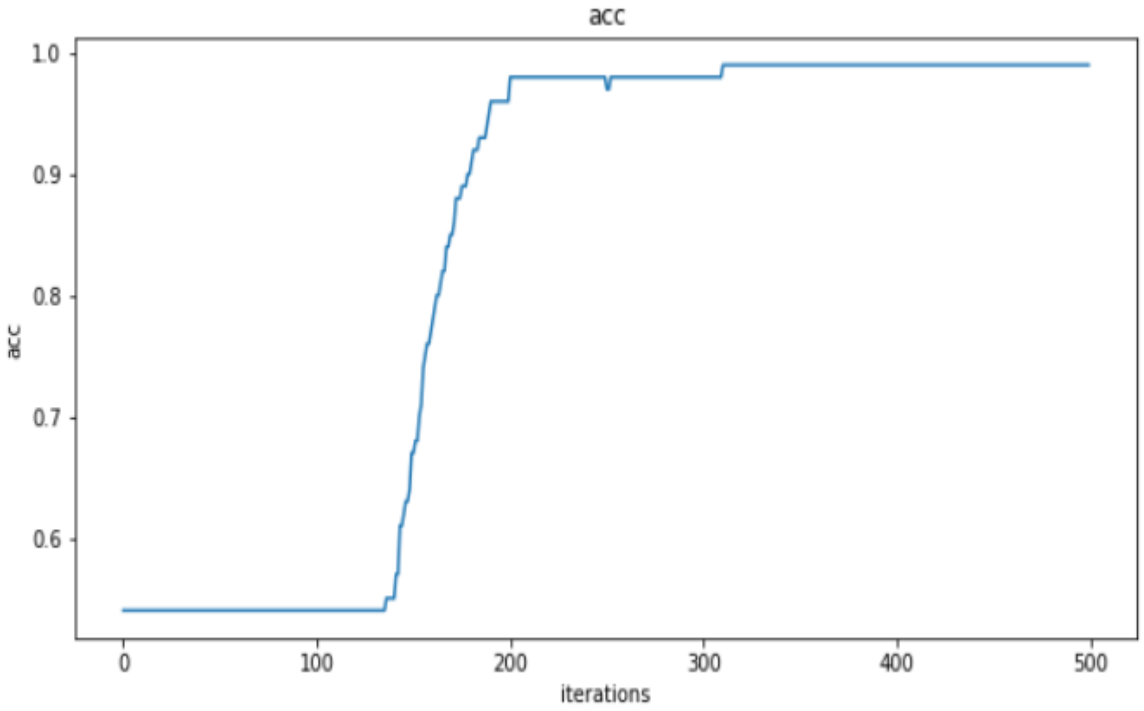
XOR



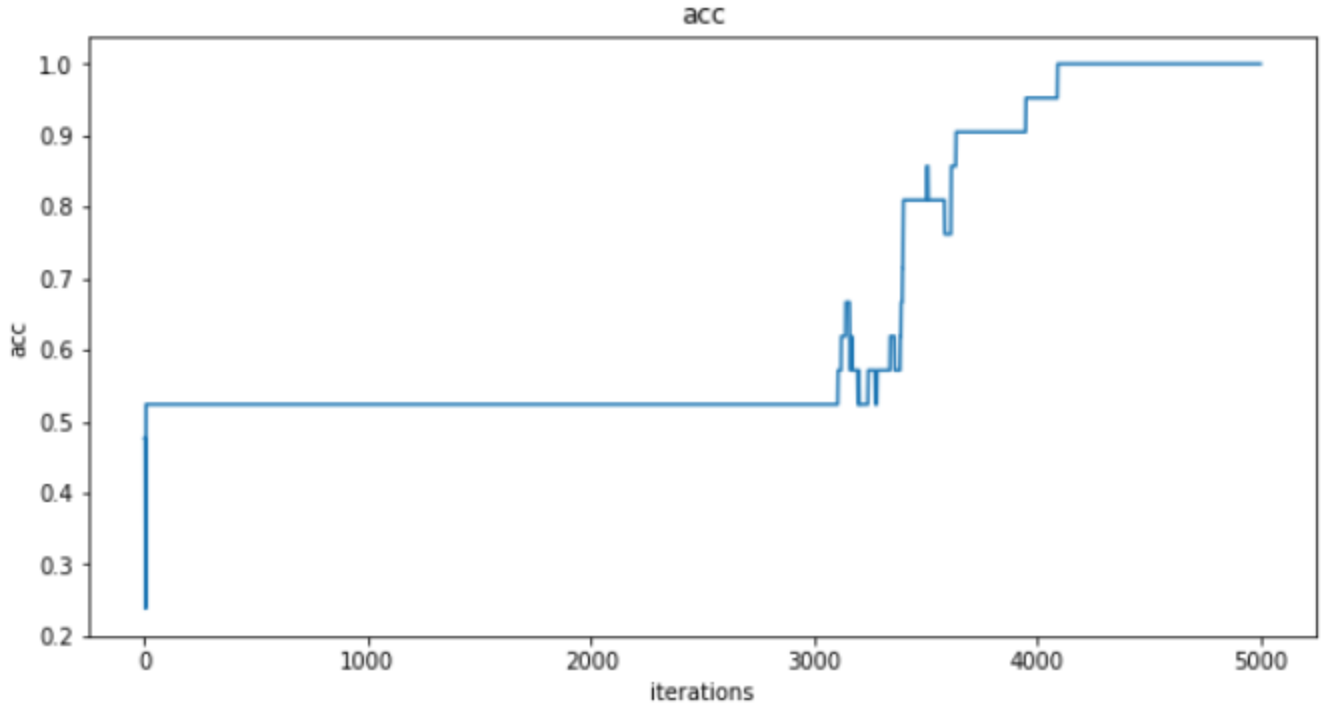
3.Results of your testing

B. Show the accuracy of your prediction

linear



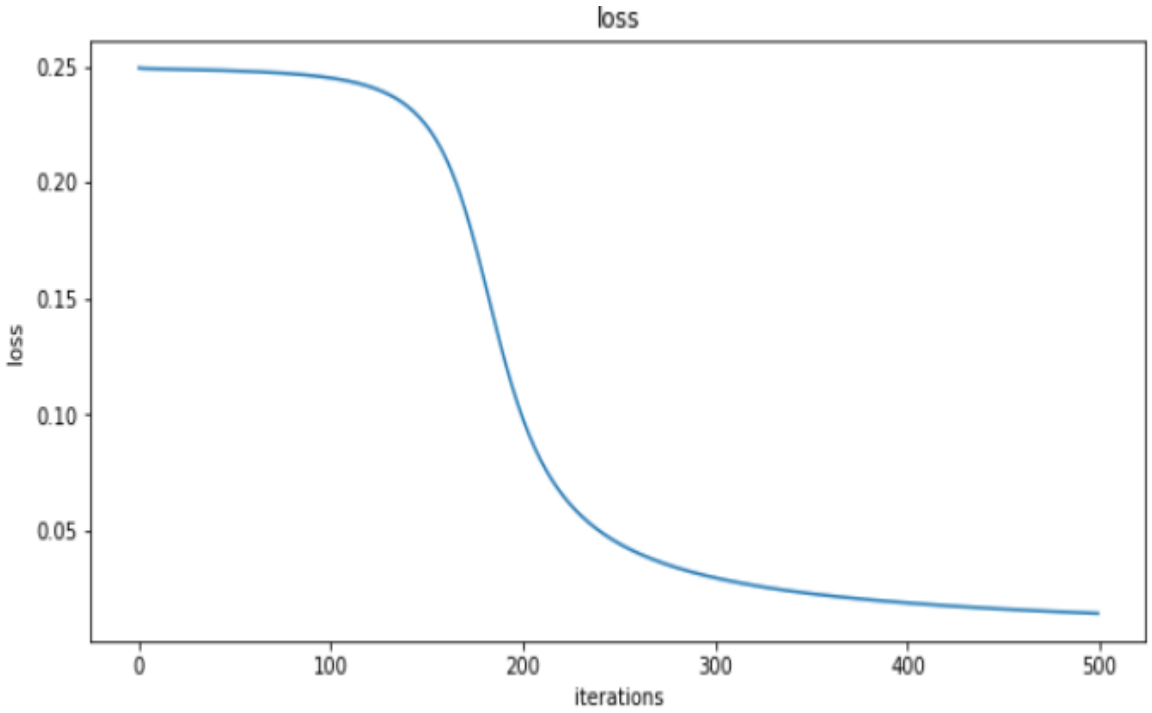
XOR



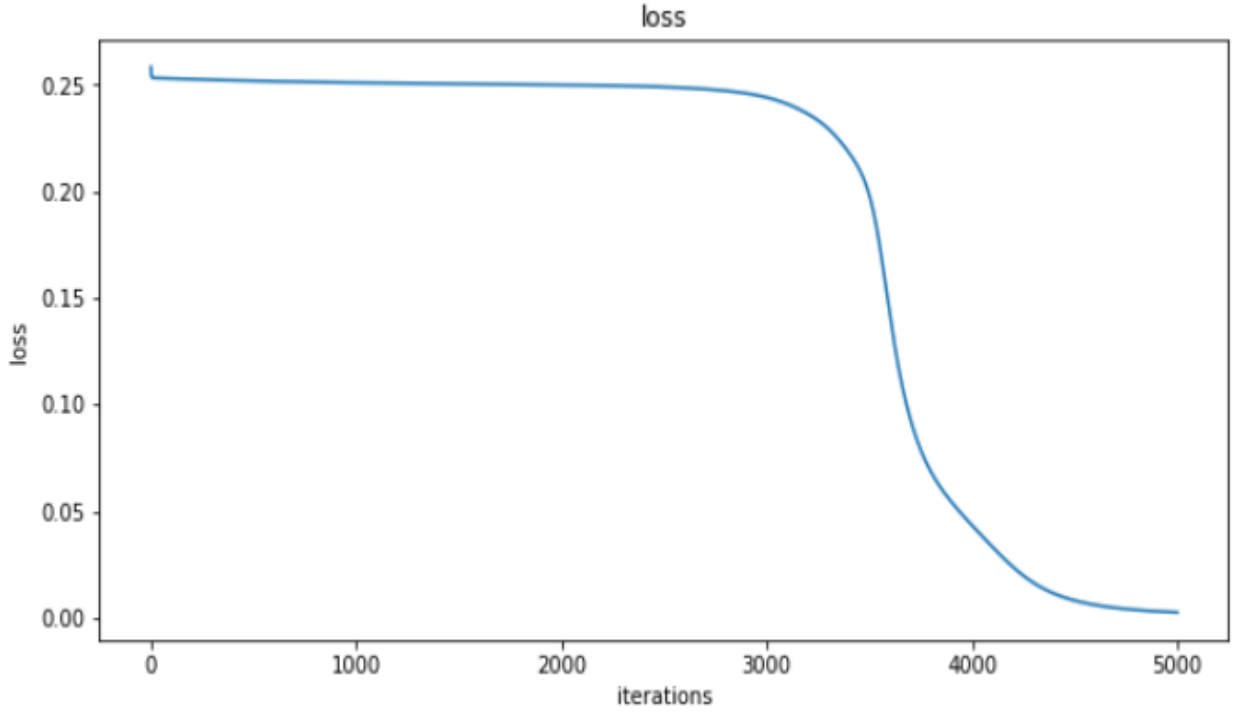
3.Results of your testing

C. Learning curve (loss, epoch curve)

linear



XOR

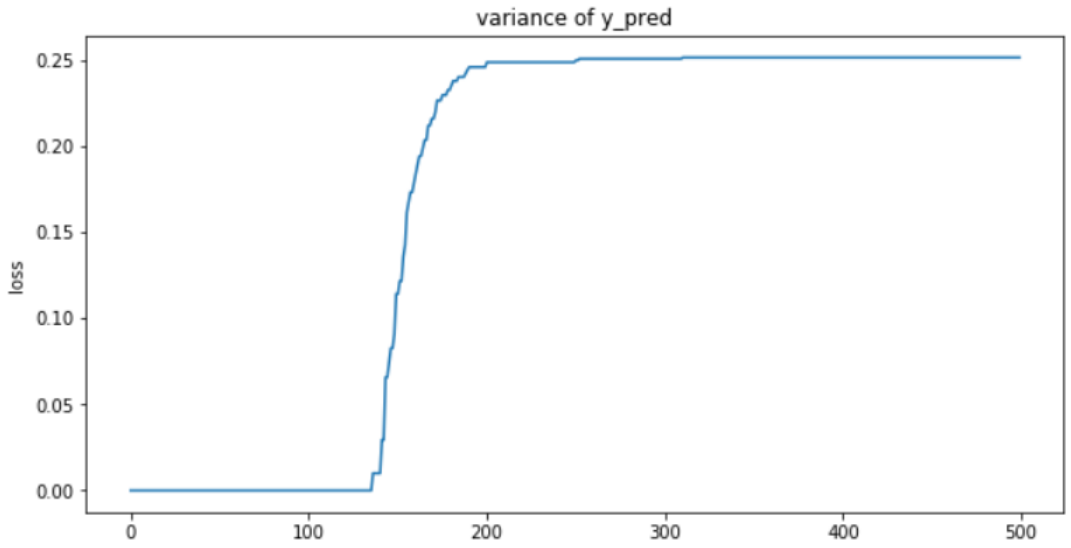


3.Results of your testing

D. anything you want to present

Variance of y_pred

linear



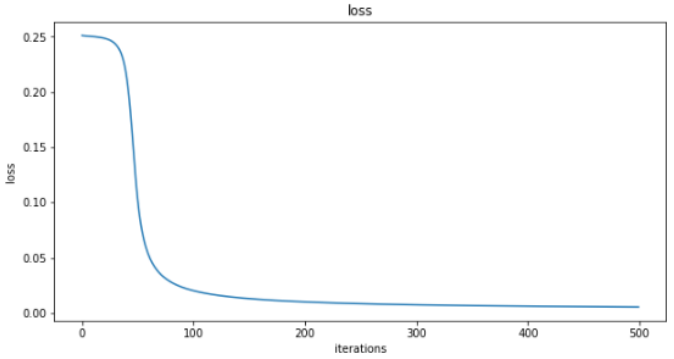
XOR

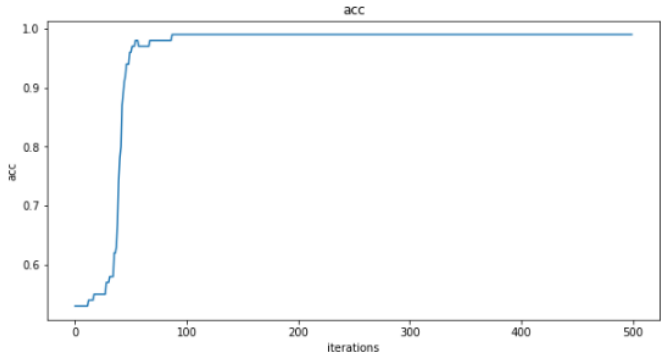
4. Discussion

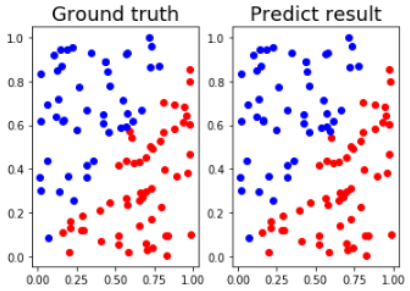
A. Try different learning rates

SET LR = 0.2

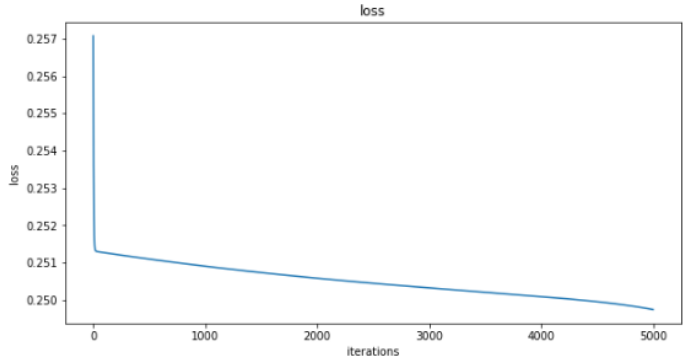
linear

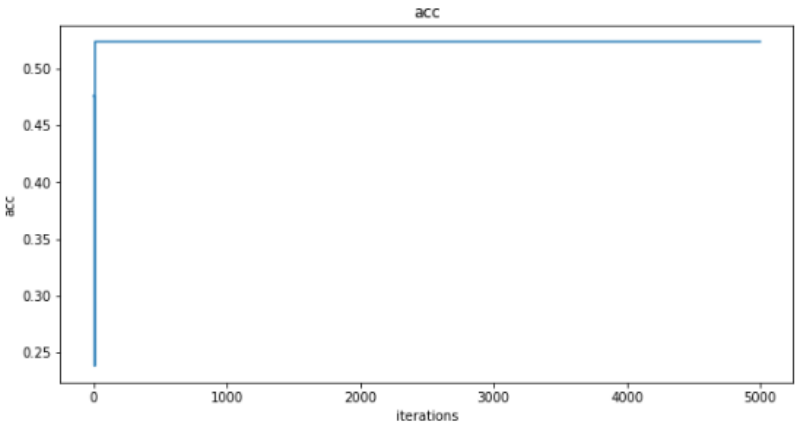


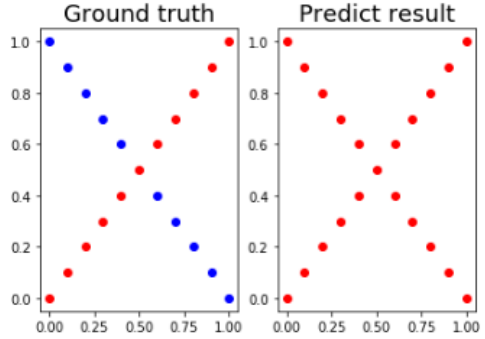




XOR



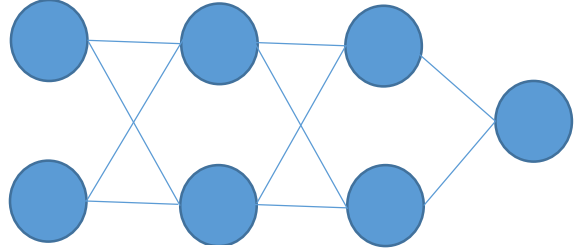




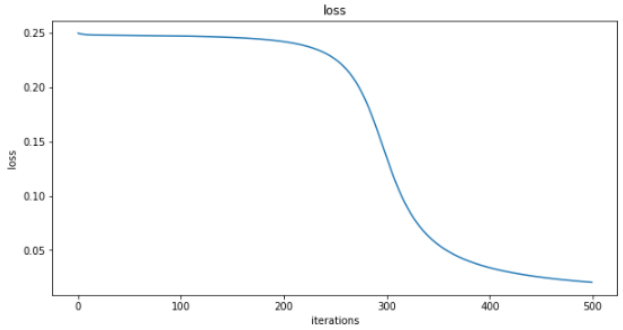
4. Discussion

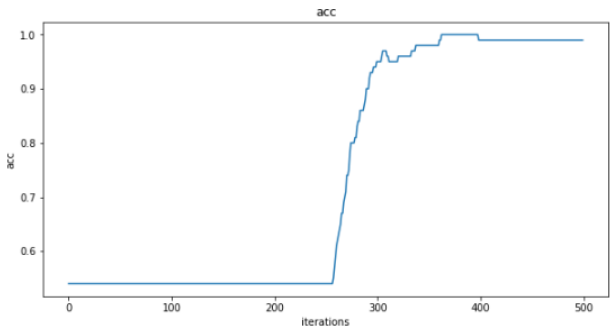
B. Try different numbers of hidden units

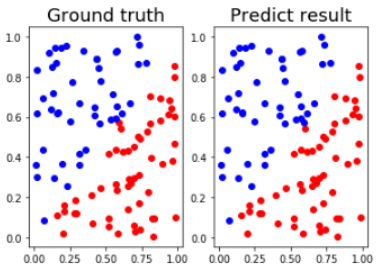
Model structure:



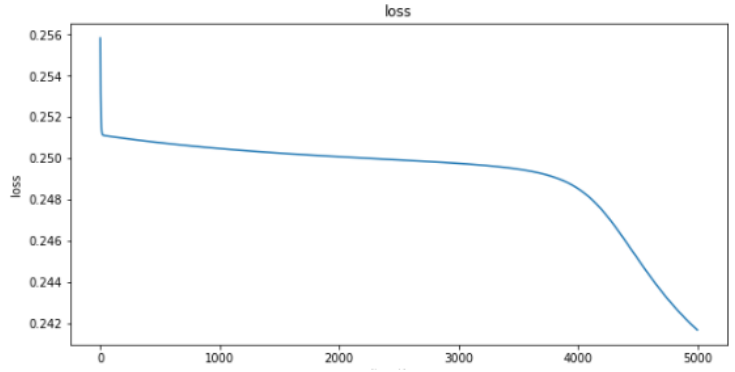
linear

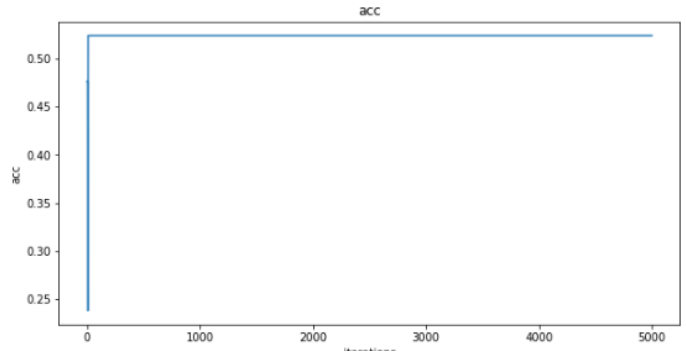


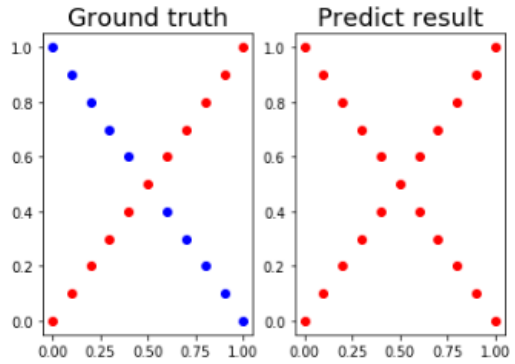




XOR



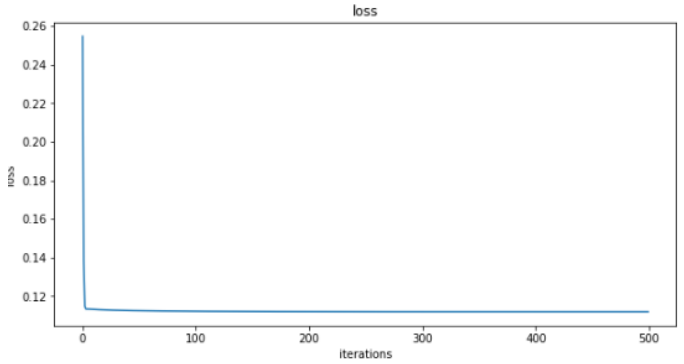


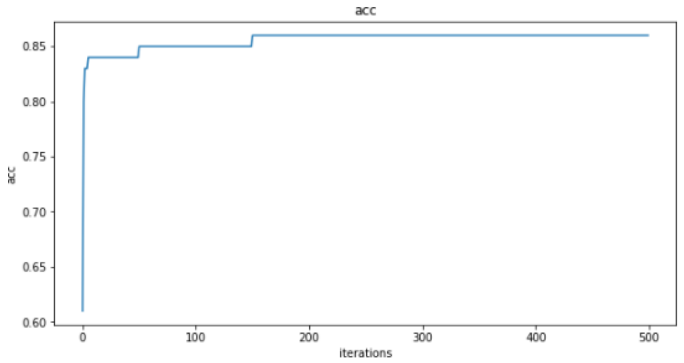


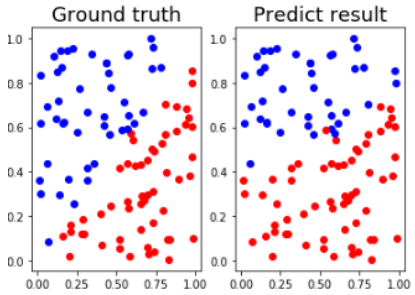
4. Discussion

C. Try without sigmoid function

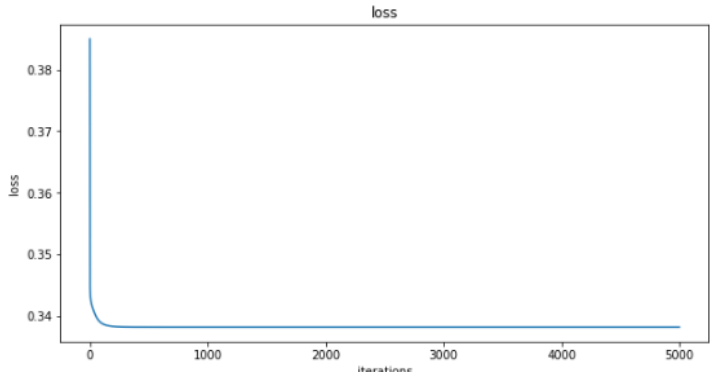
linear

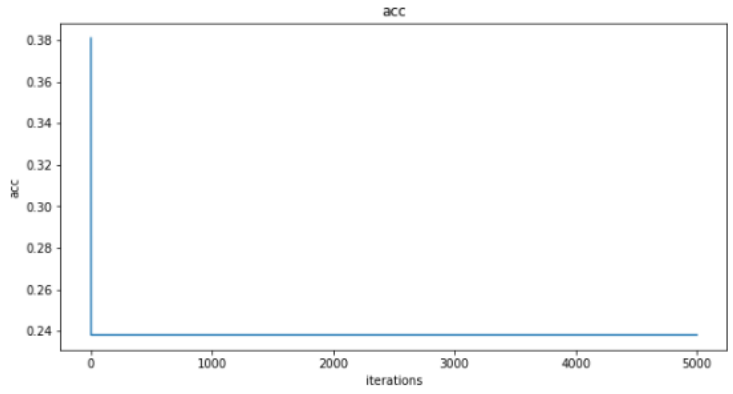


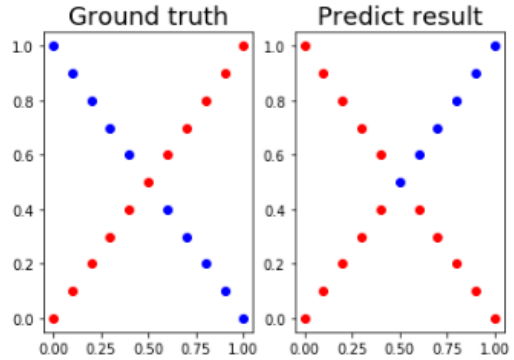




XOR



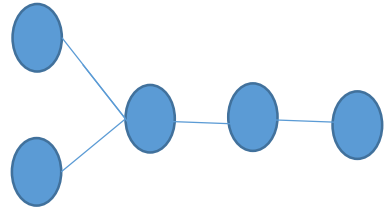




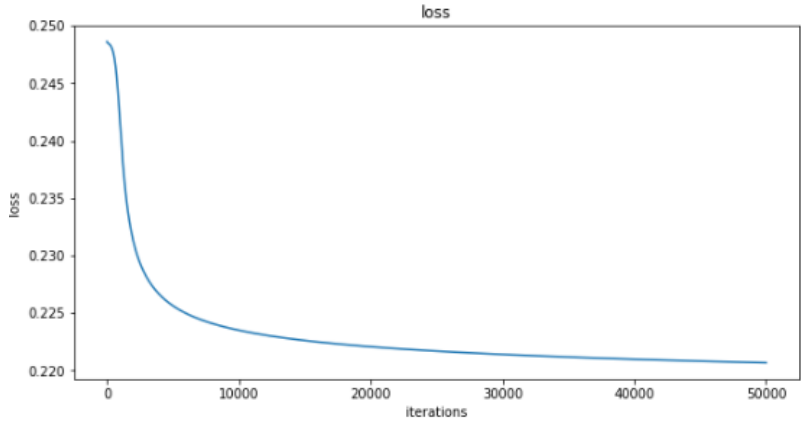
4. Discussion

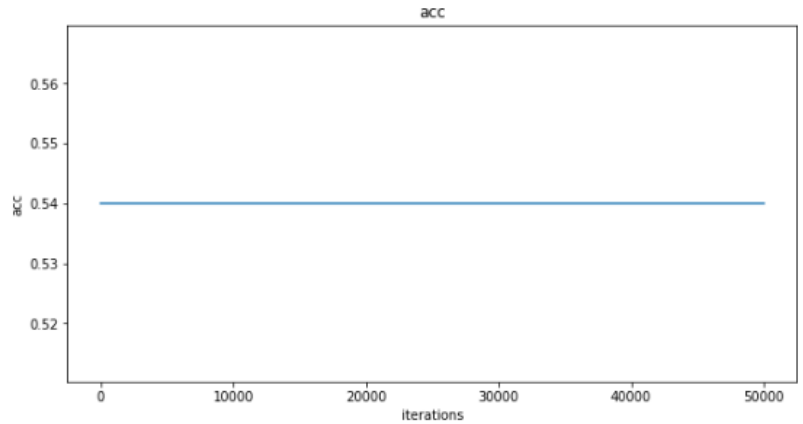
D. Anything you want to share

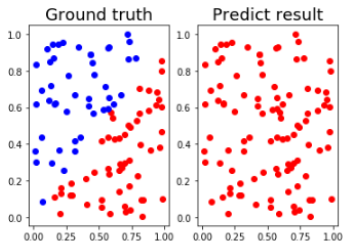
Try Model structure:



linear







XOR