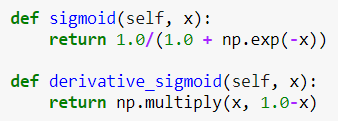
Lab1 : back-propagation

0516069 翁英傑

1. Introduction

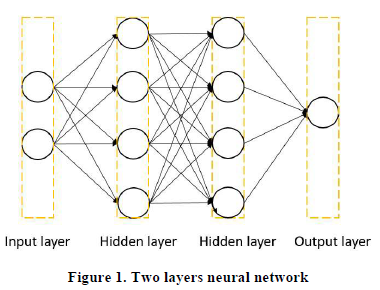
In the training stage of a neural network, it relies on a technic called gradient decent. While calculating the gradient of the loss regarding to each weight, using back propagation will ease the computation, which the gradient is calculate from the output layer to the input layer.

1. Experiment setups
2. Sigmoid functions

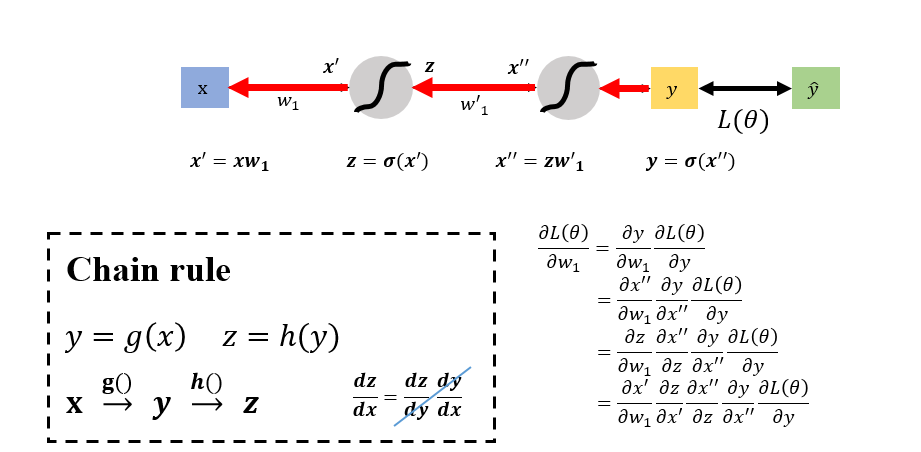
Sigmoid function is used for making each layer none linear so that it’s meaningful for to create a multi-hidden layer network.

B. Neural network



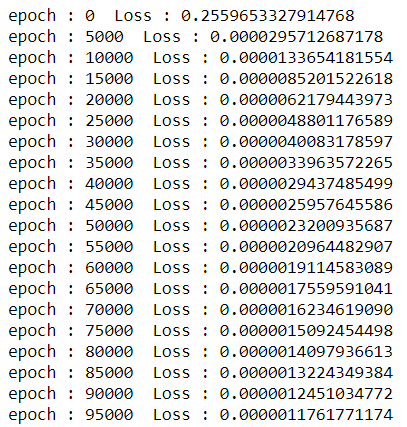
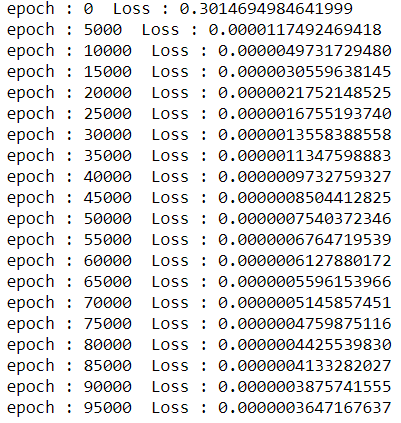
A picture of a network with two hidden layers, which each layer contains four nodes.

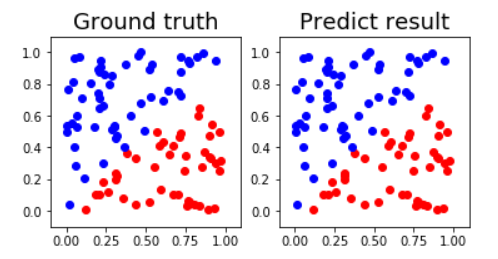
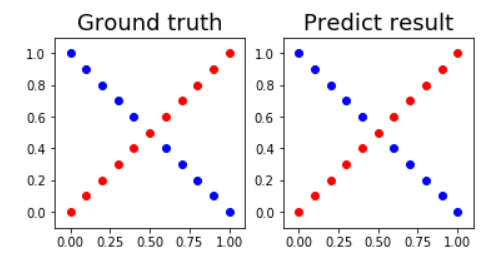
C. Backpropagation



With the chain rule, if we calculate the gradient beginning from the output layer, the previous gradient will be used in the computation of the next gradient.

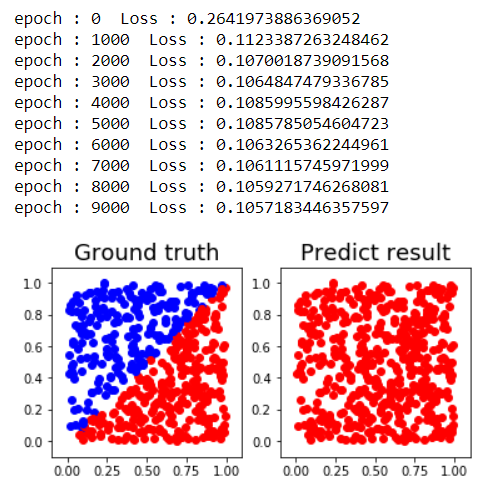
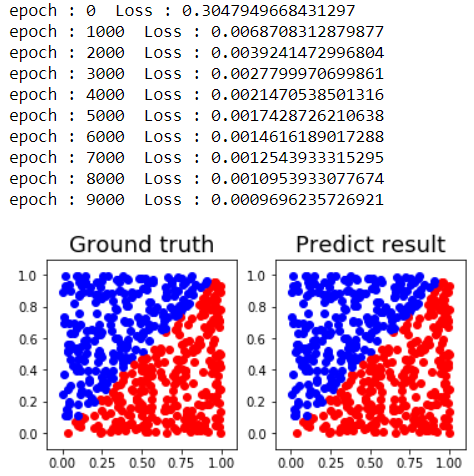
1. Results of your testing
2. Linear distribution 2. XOR distribution



1. Discussion

A common tip in training is to add a learning rate to avoid the overshooting problem. In the basic setting of this LAB, which generates 100 data samples by default setting, there is no problem with gradient decent. But if we generate over 200 data samples, the gradient will be too large that it crosses the lowest point and the loss does not decrease. Adding a learning rate solves the problem.

Learnig rate = 1 Learning rate = 0.01