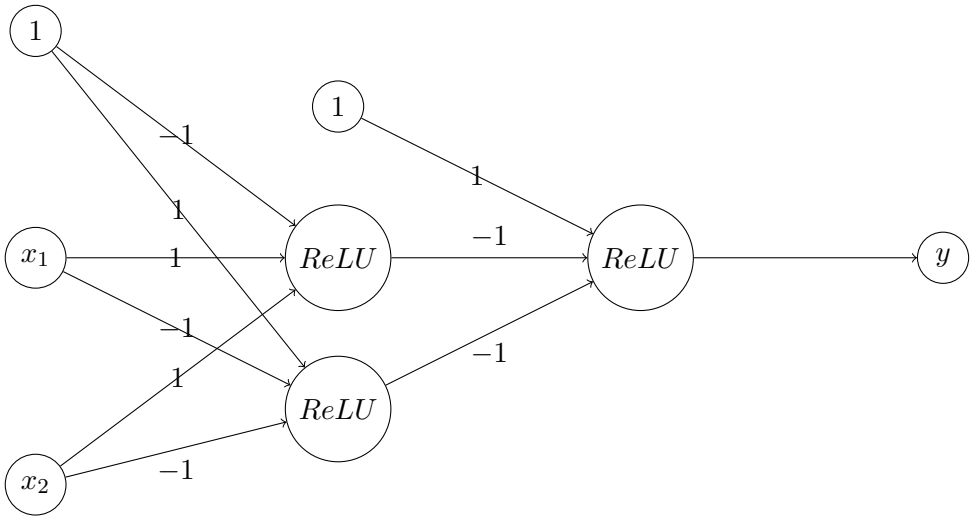
CS241 Homework 8

davidwang200099

November 2020

1. Deeper is not necessarily better. Take ResNet as an example. This article https://arxiv.org/pdf/1512.03385.pdf tells us that deeper ResNet may be worse than shallower one. Deeper network may cause gradient explosion because of the back propagation of loss. Sometimes it is so complicated compared with the real model that it may cause overfitting. In deep learning, some parameters, such as learning rate and loss function need to be carefully designed. For example, for classification problems, cross entropy loss is usually bettern than MSE loss. As the loss reduces, learning rate need to be reduced accordingly. Otherwise, the best solution may be missed. Activation function is also important. It can provide some non-linearity. But improper activation function may cause gradient explosion or vanishing. The number of neurons in hidden layer is also important. This article: https://towardsdatascience.com/neural-network-why-deeper-isnt-always-better-2f862f40e2c4 tells us that neurons in hidden layers can form many areas separated by straight lines, which is important for further classification helped by non-linear functions.

2. The neural network is as follows.



3. Assume the inputs of the neurons in the hidden layer are a_1, a_2, a_3 and a_4 from the top to the bottom. Then

   (a) During the training,

$$a_1 = \text{ReLU}((-1) \times 1 + 0 \times 2 + 0 \times 1) = 0$$

$$a_2 = \text{ReLU}(2 \times 1 + 1 \times 2 + 0 \times 1) = 4$$

$$a_4 = \text{ReLU}(0 \times 1 + 0 \times 2 + 1 \times 1) = 1$$

$$y_1 = \text{ReLU}((-1) \times a_1 + 2 \times a_2 + (-4) \times a_4) = \text{ReLU}((-1) \times 0 + 2 \times 4 + (-4) \times 1) = 4$$

$$y_2 = \text{ReLU}(1 \times a_1 + 0 \times a_2 + (-2) \times a_4) = \text{ReLU}(1 \times 0 + 0 \times 4 + (-2) \times 1) = 0$$

   (b) During the testing,

$$a_1 = \text{ReLU}(0.75 \times ((-1) \times 1 + 2.5 \times 2 + 0 \times 2 + 0 \times 1)) = 3$$

$$a_2 = \text{ReLU}(0.75 \times (2 \times 1 + 0 \times 2 + 1 \times 2 + 0 \times 1)) = 3$$

$$a_3 = \text{ReLU}(0.75 \times (3 \times 1 + (-1) \times 2 + 0 \times 2 + (-2) \times 1)) = 0$$

$$a_4 = \text{ReLU}(0.75 \times (0 \times 1 + 0 \times 2 + 0 \times 2 + 1 \times 1)) = 0.75$$

$$y_1 = \text{ReLU}(0.75 \times ((-1) \times a_1 + 2 \times a_2 + 0 \times a_3 + (-4) \times a_4)) = 0$$

$$y_2 = \text{ReLU}(0.75 \times (1 \times a_1 + 0 \times a_2 + (-1) \times a_3 + (-2) \times a_4)) = 1.125$$