

HW 8

余北辰 519030910245

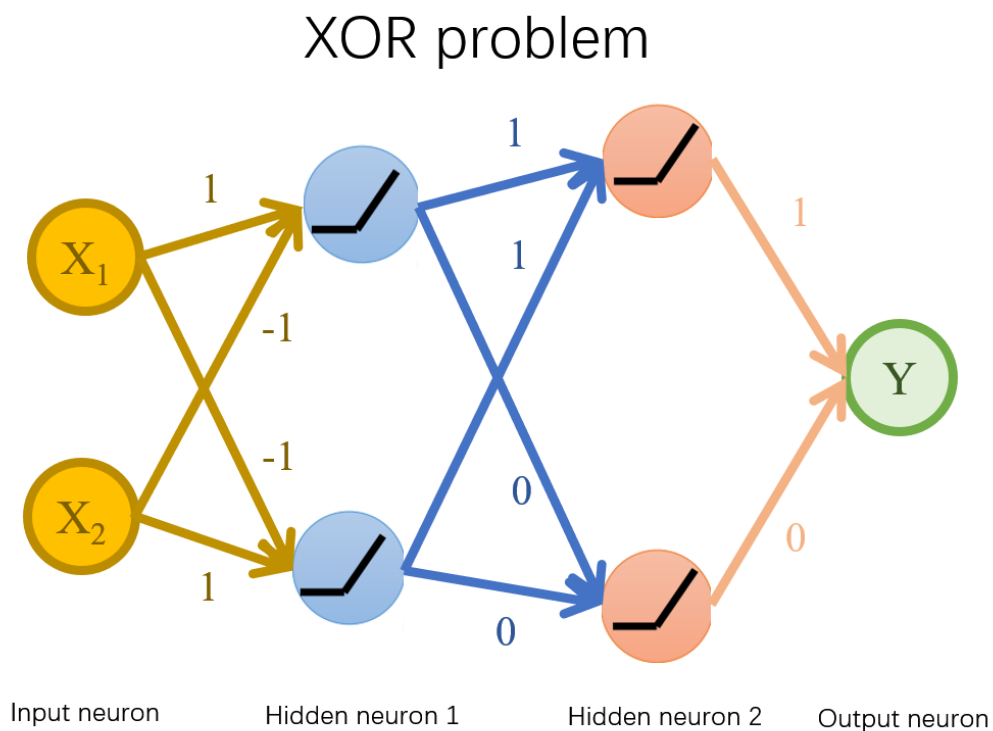
1.

Usually, deeper is better.

If there are more parameters in the neural network, and usually we can get a more accurate results. And from the example in our lecture, the deep neural network behaves much better than the wide neural network, so the deep learning can give us a more accurate results. In fact, the neural network with only one hidden neurons can work as well but it is really hard to train.

The deep neural network prone to overfitting and it is easy to fall into local minimum. It is necessary to take some methods to avoid these disadvantages.

2.



3.

When learning:

$$t_1 = \max\{-1, 0\} = 0;$$

$$t_2 = \max\{2 + 1 \times 2, 0\} = 4;$$

t_3 is omitted;

$$t_4 = \max\{1, 0\} = 1;$$

$$y_1 = \max\{-t_1 + 2t_2 - 4t_3, 0\} = \max\{4, 0\} = 4;$$

$$y_2 = \max\{t_1 - 2t_4, 0\} = \max\{-2, 0\} = 0.$$

When testing:

$$t_1 = 0.75 \times \max\{-1 + 2.5 \times 2, 0\} = 3;$$

$$t_2 = 0.75 \times \max\{2 + 2, 0\} = 3;$$

$$t_3 = 0.75 \times \max\{3 - 1 \times 2 - 2, 0\} = 0;$$

$$t_4 = 0.75 \times \max\{1, 0\} = 0.75;$$

$$y_1 = 0.75 \times \max\{-t_1 + 2t_2 - 4t_3, 0\} = 0.75 \times \max\{0, 0\} = 0;$$

$$y_2 = 0.75 \times \max\{t_1 - t_3 - 2t_4, 0\} = 0.75 \times \max\{1.125, 0\} = 1.125.$$