First Course

Liangjie Cao

VISION@OUC

July 7, 2018

1 What is a neural network?

Neural Network is not a new term, but with the breakthrough of data, computing power and theory, it has only ushered in the spring after the winter. The neural network is mainly composed of neurons (neuron), which is usually a linear combination of multiple inputs and an activation function, where the activation function is often a nonlinear function. Like the human brain, many neurons have powerful capabilities when combined with multiple links. Andrew gave two of the simplest examples from the beginning.

2 Deep learning's widespread

Deep learning can be applied not only to traditional structured data, but also to unstructured data such as sound, images, and text. As a result, the word "structured" was forgotten for a while, and it took a long time to fool a two-dimensional vs. multidimensional, in fact, my intention was structured vs. unstructured. Fortunately, it was just a word, not a knowledge point, and no one noticed it. It can be seen that a good faith book is not as good as a book, especially in the emerging and mixed industry of deep learning.

3 Single neural network

Professor Wu says if we have only one variable in our input, applying a ReLU function on top of it forms a single neuron network. He give the example of the house price. And the full name of the ReLU function is Rectified Linear Unit. Don't be scared by this name. In fact, deep learning is completely a paper tiger. ReLU is actually a max function, which is similar to MARS. Its shape is like this: The sigmoid
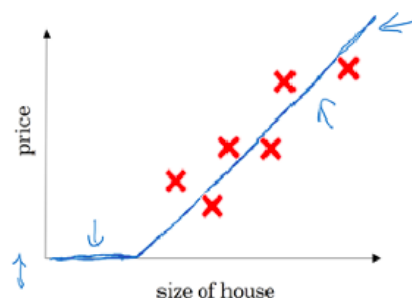
Housing Price Prediction
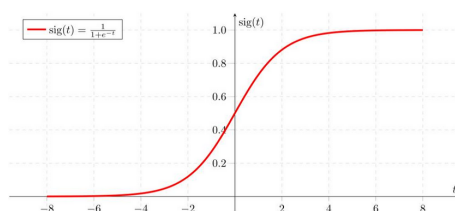

Figure 1: ReLU

Figure 2: Sigmoid

function is not only commonly used in logistic studies, but also in deep learning. When we compare the ReLU function with the sigmoid function, we will find that ReLU replaces part of the data with 0, which makes it faster to calculate large-scale data than sigmoid (Imagine if each layer of ReLU discards 50% of

the useless data, Then the data volume after 4 layers is the original 6%, of course, the real situation is not so simple, and the derivative of the other part of ReLU is 1, it is convenient to calculate the gradient in backpropagation; the sigmoid function is everywhere, but in the absolute Most of the values on the range are very small, which makes it impossible to filter out unwanted data, greatly reduces the learning speed of gradient descent, and there may be a problem of gradient disappearance in backpropagation.

4   Formula of the Sigmoid Function

Professor Wu tell us the function of the sigmoid function very specifically. Where x is input, and y is output. ω and b are parameters. Then we can get the loss function:

$$\ell(\hat{y}, y) = -(ylog(\hat{y}) + (1 - y)) \tag{1}$$

And the cost function is:

$$J(\omega, b) = \frac{1}{m} \sum_{i=1}^{M} \ell(\hat{y}, y) \tag{2}$$

5   Looking at Logistic Regression with the Structure of Neural Network

The logistic regression that everyone is familiar with can be seen as the simplest neural network, where z is a linear combination of input features, and the sigmoid function is a nonlinear transformation to z. Therefore, professor Wu built a basic neural network with Logistic regression as the input and vectorized image as input.

6   Conclusion

In summary, the modeling of neural networks mainly includes the following main steps: First, define the model structure, such as how many layers, what activation function, how many input features, *etc.*; Second, initialize the model parameters. In the case of logistic regression, we only need to initialize ω and b to 0, and we don't need to add some jitter. And we perform a sub-loop include :
Forward Propagation: Calculating the current cost.
Back Propagation: Calculate the current gradient.
Gradient descent: update the parameters.
The updated parameter is $\theta = \theta\alpha d\theta$, where α is the learning rate and $d\theta$ is actually $dJ/d\theta$. Until the end of the iteration, or the gradient is approximately zero and the parameters are no longer changing.