Reading Circle #02

3.1 From Perceptron to Neural Network
3.2 Activation function

2018/05/21

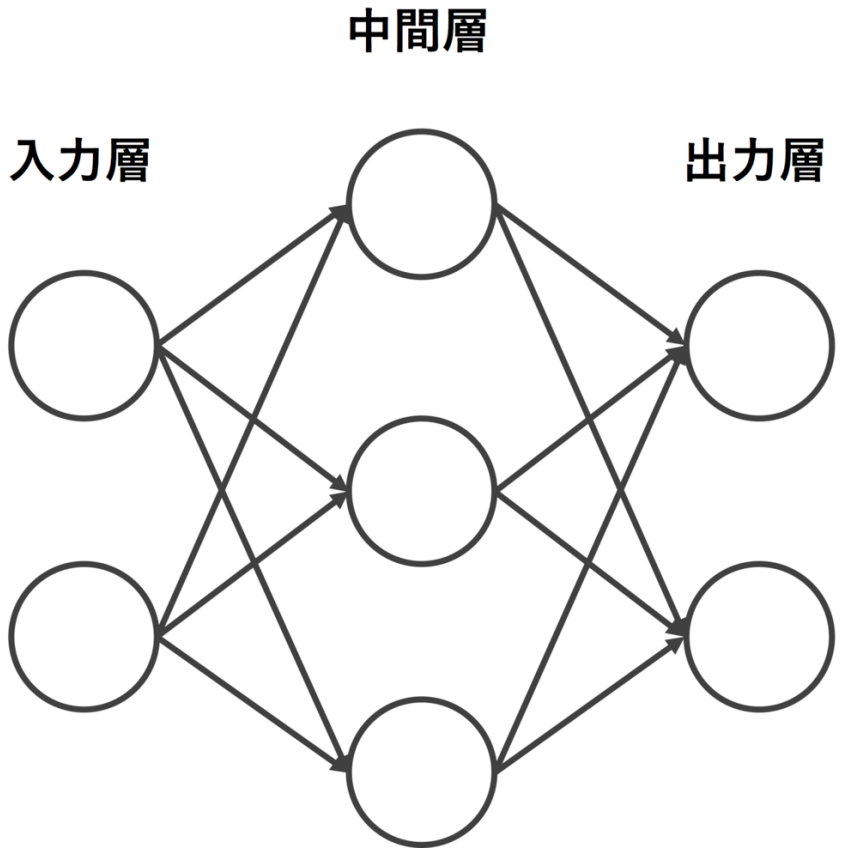
B4 Yasutaka MIYAMOTO

Index

Neural Network

- input layer
- hidden layer
- output layer

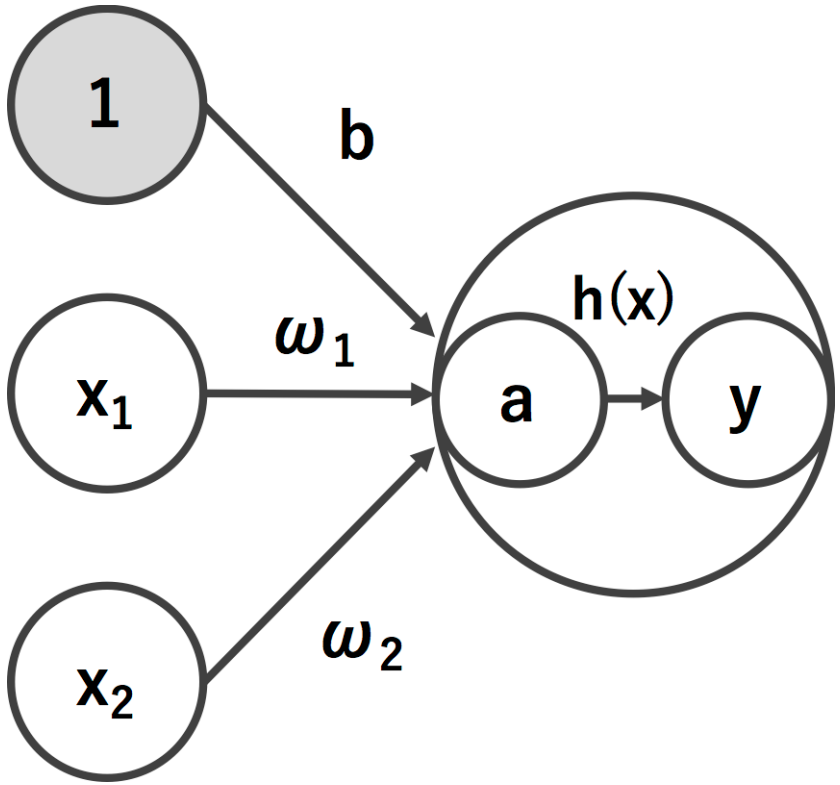
- Automatically learn weights from data



Signal transmission method

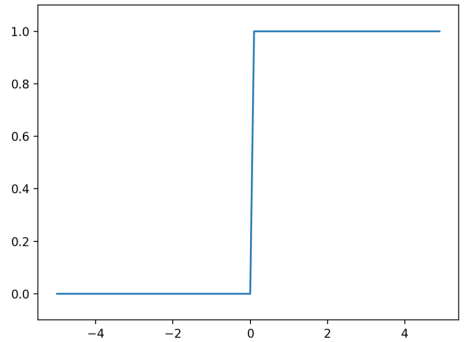
- $y = h(a)$
- $a = b + w_1x_1 + w_2x_2$

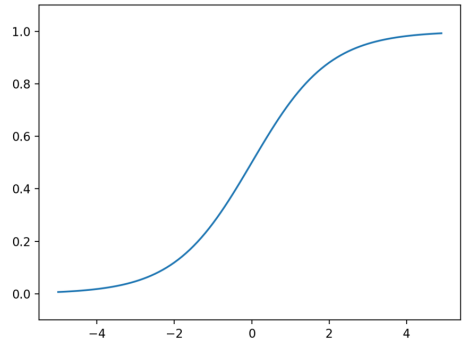
- $h(a)$: Activation funcntion
  - Determine
    how the sum of input signals fires



Difference from perceptron

- Type of activation function
  - Perceptron : Step function
  - NN : Sigmoid fucntion, ReLU, etc.





- Continuous nonlinear function
  - It's important to determine weight value(in Backpropagation)
    - Continuous(smooth) : Differentiable
    - Nonlinear : Improve expression
      - A neural network using a linear activation function is equivalent to a neural network without a hidden layer[1][2]

[1]線形の活性化関数はなぜだめか/ニューラルネットワーク - たれぱんのびぼーろく : https://tarepan.hatenablog.com/entry/2015/10/06/183036

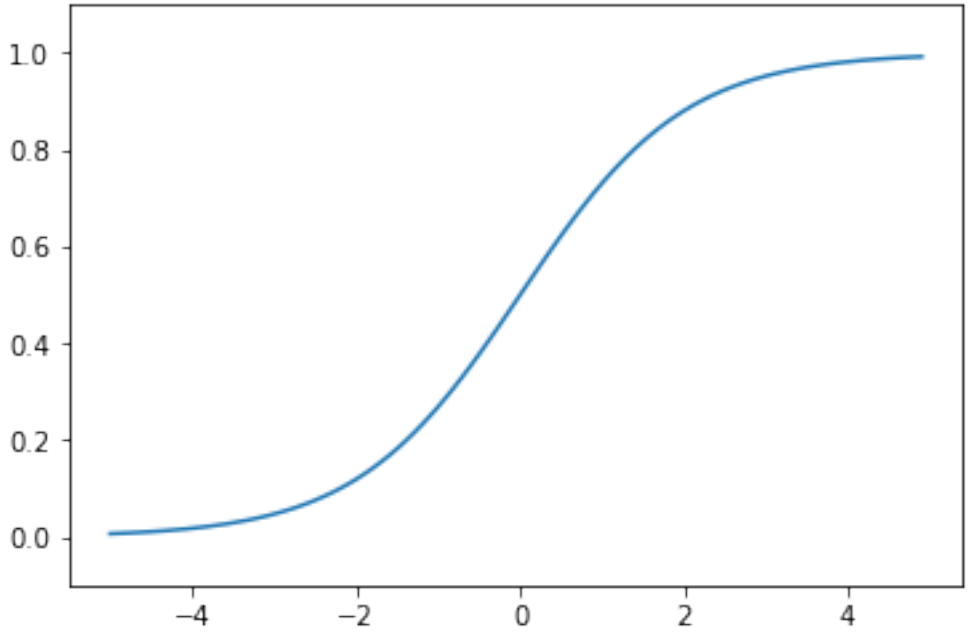
[2]ゼロから始めるDeepLearning_その1_ニューラルネットとは - 分からんこと多すぎ: http://rishida.hatenablog.com/entry/2014/02/25/110643

Activation Function

- Decide that the sum of the input signals fires
- In order to improve the performance, various functions have been proposed
  - Sigmoid function
  - Rectified Linear function(ReLU)

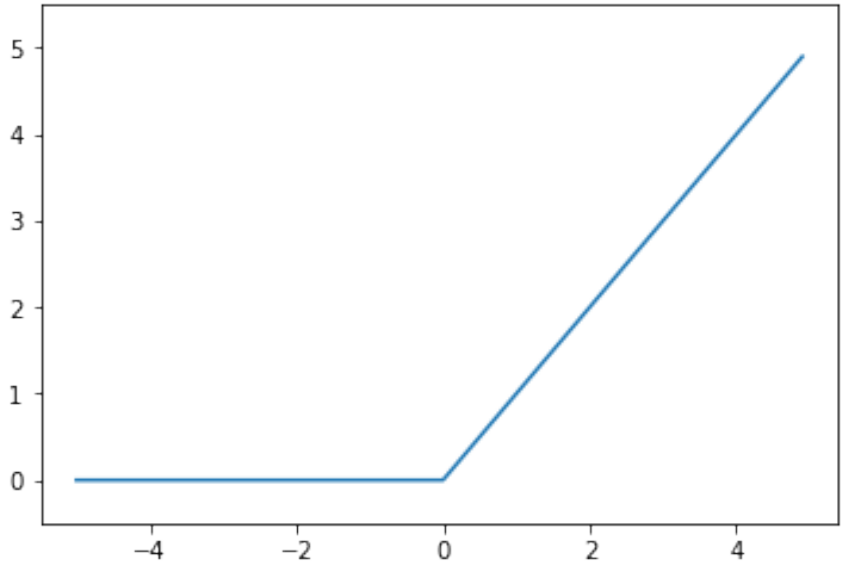
Sigmoid function

- used for a long time in research on neural networks
- $h(x) = \frac{1}{1+\exp(-x)}$



Rectified Linear function(ReLU)

- Recently mainly used
- $h(x) = \begin{cases} x\ (x > 0) \\ 0\ (x \leq 0) \end{cases}$
- differentiable?
  - non-differentiable(x = 0)
  - Since the gradient is constant, it is possible to deal with the gradient loss problem occurring in learning[3] (Chapter6)



[3]ニューラルネットワーク / ディープラーニング - 星の本棚 : http://yagami12.hatenablog.com/entry/2017/09/17/111935

Summary

- The difference between neural network and perceptron is the activation function
- It is important that the activation function is a continuous nonlinear function in order to determine the weight value
- In order to improve learning accuracy, various activation functions have been proposed

What’s “Vanishing gradient problem”

- > The problem is that in some cases, the gradient will be vanishingly small, effectively preventing the weight from changing its value. In the worst case, this may completely stop the neural network from further training.[4]

[4] Vanishing gradient problem - Wikipedia: https://en.wikipedia.org/wiki/Vanishing_gradient_problem

machine learning - How does rectilinear activation function solve the vanishing gradient problem in neural networks? - Cross Validated : https://stats.stackexchange.com/questions/176794/how-does-rectilinear-activation-function-solve-the-vanishing-gradient-problem-in

Reasons to use ReLU function

- It is empirically known that not only the calculation amount is effective but also the learning optimization can be performed more quickly and better.[5]

- At present it is not fully understood. I would like to understand as I go through with the book.

[5] 深層学習について 岡谷貴之 日本ロボット学会誌 : https://www.jstage.jst.go.jp/article/jrsj/33/2/33_33_92/_pdf