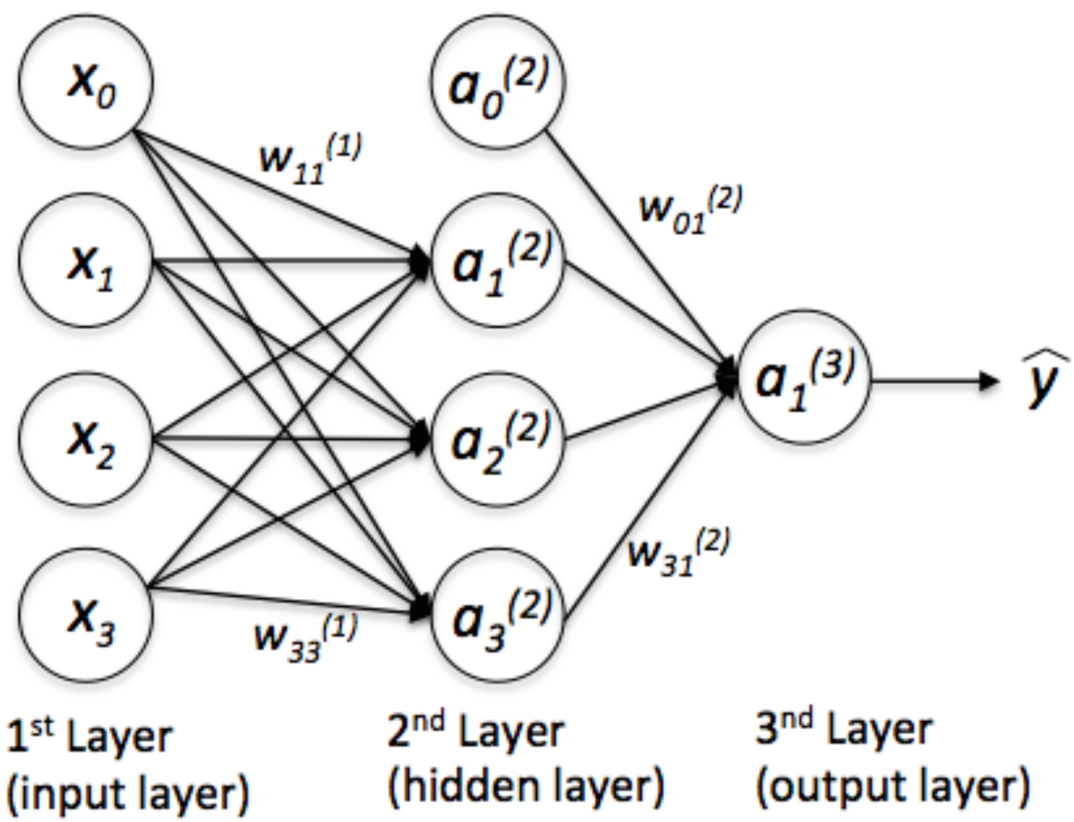
Neural Networks
Gordon on 2016.09.13

This blog talks about the basic knowledge of Neural Networks. Hope in the future I can get deeper in this topic.

Neural Networks are a pretty old algorithm of having machines that can mimic the brain. "In machine learning and cognitive science, an artificial neural network (ANN) is a network inspired by biological neural networks (the central nervous systems of animals, in particular the brain) which are used to estimate or approximate functions that can depend on a large number of inputs that are generally unknown."

What does the basic Neuron Network look like? It has an input layer, one or more hidden layers and one output layer. Based on different algorithms, the links(weights) among layers may be different.



Using Logistic Regression Algorithm in a Neural Network:

a_i^j is the "activation" of unit i in layer j. w^j or θ^j is the matrix of weights controlling function mapping from layer j to layer j+1. x_0, a_0^2 are the bias units.

The Logistic Activation Function:

$$a_1^2 = g\left(\theta_{10}^1 x_0 + \theta_{11}^1 x_1 + \theta_{12}^1 x_2 + \theta_{13}^1 x_3\right)$$
$$a_2^2 = g\left(\theta_{20}^1 x_0 + \theta_{21}^1 x_1 + \theta_{22}^1 x_2 + \theta_{23}^1 x_3\right)$$
$$a_3^2 = g\left(\theta_{30}^1 x_0 + \theta_{31}^1 x_1 + \theta_{32}^1 x_2 + \theta_{33}^1 x_3\right)$$

The output bases on the previous layer:

$$a_1^3 = h_\theta(x) = g\left(\theta_{10}^2 a_0^2 + \theta_{11}^2 a_1^2 + \theta_{12}^2 a_2^2 + \theta_{13}^2 a_3^2\right)$$

How to calculate the weight's dimension:

If the network has s_j units in layer j, s_{j+1} units in layer j + 1, then θ^j will be of dimension $s_{j+1} \cdot (s_j + 1)$.

We can vectorized the above functions:

$z^j = \theta^{j-1} x,\ a^j = g(z^j)$ x is the value vector of the previous layer.

Multiclass Classification

For multi-classes, we want to distinguish the objects based on the output layer y. Instead of making y = {1, 2, 3, 4}, we want to use the combinations of its values to classify the objects.

When it is a cat, $h_\theta(x) \approx \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$.

When it is a dog, $h_\theta(x) \approx \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$.

When it is a bird, $h_\theta(x) \approx \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

With different algorithms, the neural network will have different performances. In the next blog, I will discuss one of the powerful algorithms, Backpropagation Algorithm, and how to use it to automatically train our neural network.