Week 4

Arjun Sarao

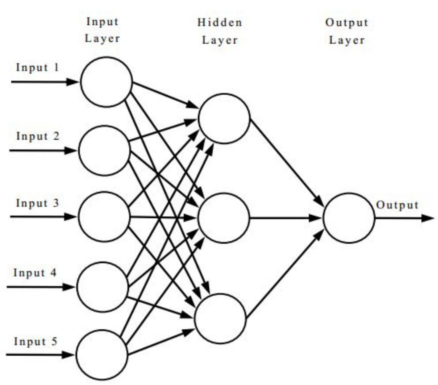
April 3, 2021

1 Non-Linear Hypotheses

Sometimes we want to solve a classificatin problem where the hypotheses function is non-linear. A good estimation on the number of features is: ${}_nC_r$ where n is the number of features and r is the degree of polynomial.

2 Neural Networks

Neurons are a type of cell found in the brain, they consist of *Dendrites* and *Axons*. We can think of Dendrites as inputs and Axons as outputs. In neural networks we have the first layer known as the input layer which takes in inputs of $x_1, x_2, \ldots, x_n$ and outputs the result of out hypothesis function. We can also have a x_0 input node, which is called the *Bias Unit* and is always equal to 1. In neural networks we use the same logistic function for classification except now we call it the *Sigmoid* ***Activation*** *Function*. We can also call our θ parameters *Weights*.



Our input nodes are connected to another layer called the *Hidden Layer*, which is connected to the *Output Layer*. We denote the hidden layers as $a_i^{(j)}$, where j is the layer number and i is the unit number,they can also be reffered to as *Activatoin Units*.

We also have a matrix that dictates the connections between the different layers, the notation is $\Theta^{(j)}$, which controls the mapping from layer j to layer $j+1$. Basically if the Neural Network has s_j units in layer j and s_{j+1} units in layer $j+1$, then $\Theta^{(j)}$ will be $s_{j+1} \times (s_j + 1)$. The $+1$ comes from the *Bias Nodes* x_0 and $\Theta_0^{(j)}$.

The following is an example of a neural net:

$$a_1^{(2)} = g(\Theta_{10}^{(1)} x_0 + \Theta_{11}^{(1)} x_1 + \Theta_{12}^{(1)} x_2 + \Theta_{13}^{(1)} x_3) \tag{1}$$

$$a_2^{(2)} = g(\Theta_{20}^{(1)} x_0 + \Theta_{21}^{(1)} x_1 + \Theta_{22}^{(1)} x_2 + \Theta_{23}^{(1)} x_3) \tag{2}$$

$$a_3^{(2)} = g(\Theta_{10}^{(1)} x_0 + \Theta_{31}^{(1)} x_1 + \Theta_{32}^{(1)} x_2 + \Theta_{33}^{(1)} x_3) \tag{3}$$

$$h_\Theta(x) = a_1^{(3)} = g(\Theta_{10}^{(2)} a_0^{(2)} + \Theta_{11}^{(2)} a_1^{(2)} + \Theta_{12}^{(2)} a_2^{(2)} + \Theta_{13}^{(2)} a_3^{(2)}) \tag{4}$$

Instead of writing all of this down, we can simplify to:

$$a_1^{(1)} = g(z_1^{(2)}) \tag{5}$$

$$a_2^{(1)} = g(z_2^{(2)}) \tag{6}$$

$$a_3^{(1)} = g(z_3^{(2)}) \tag{7}$$

By manipulating the values for the theta's we can actually simulate logical operations.