**Day 13**

**What to do?**

Learn about vanishing gradient problem.

**The problem:**

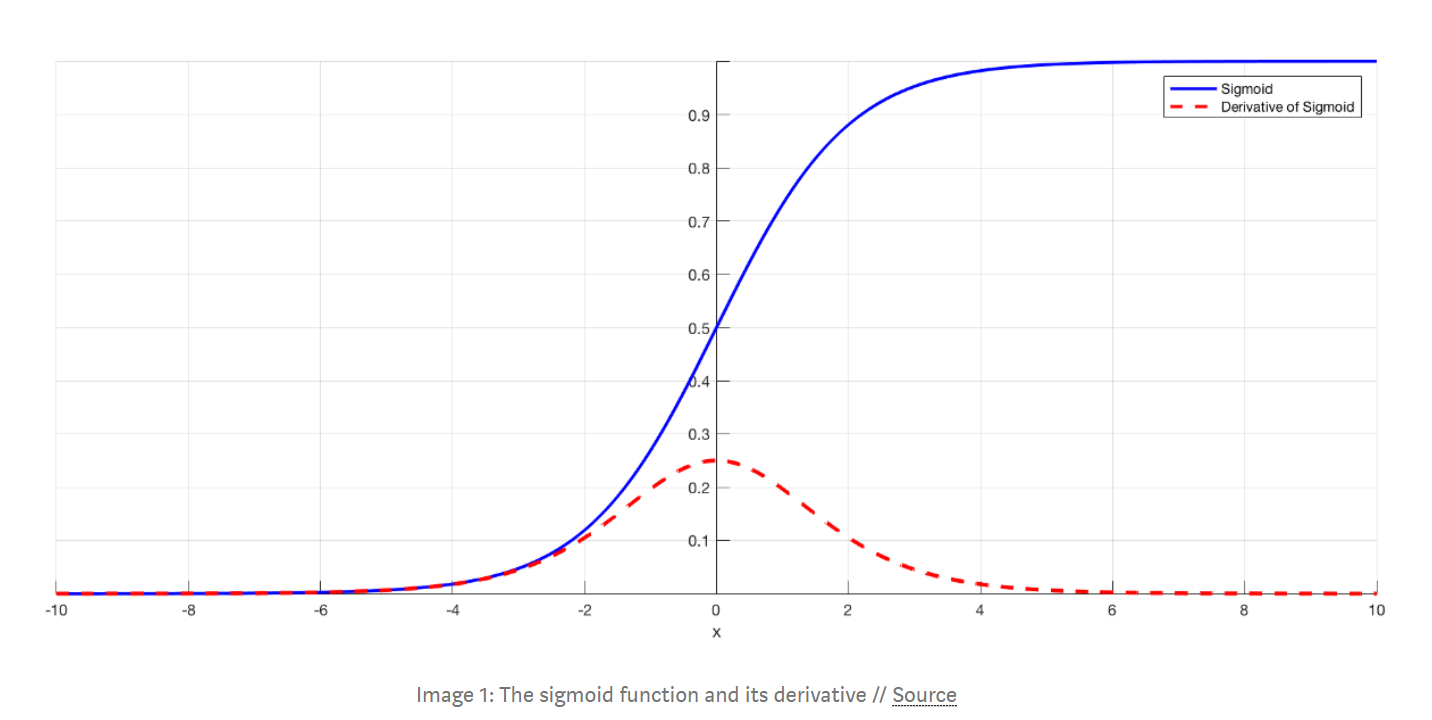
When you have a huge and deep neural network and see that in gradient based methods during back propagation. Backpropagation’s goal is to find the gradients for the network to update the weights and biases so the predicted output could be close to the actual output. When the gradients approach zero, it is called as “vanishing gradient problem”.

**Significance:**

If the gradients get close to zero, the network becomes harder to train because the weights and biases remain almost unchanged and model will not produce the expected output.

**Causes:**

This problem is mostly caused by “sigmoid” and “tanh” activation functions. Let us consider the sigmoid function. We know that the sigmoid function g(x) = 1/(1+e^-x). The derivative of it is g(x)\*(1-g(x)). It is shown that as the value ‘x’ increases the derivative approaches zero. Imagine a network with 10 layers with each hidden layer having sigmoid activation function. When you back propagate to find the gradients, the gradients will seem to decrease exponentially leaving the weights and biases unaffected.



**Solutions:**

1. Use ReLU activation function
2. Perform batch normalization