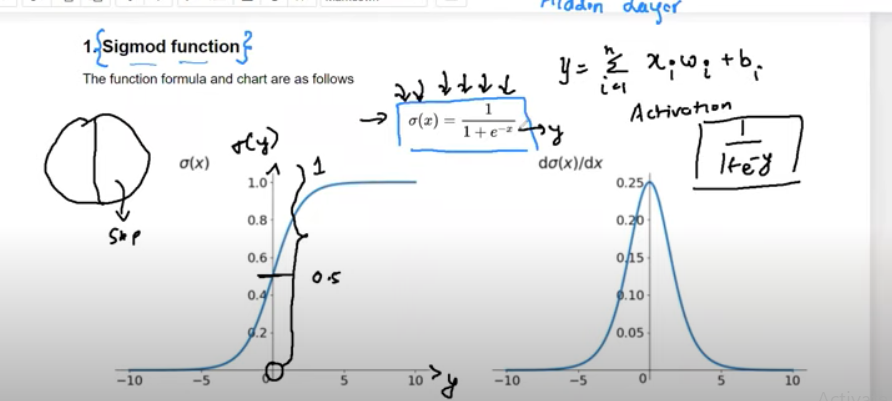
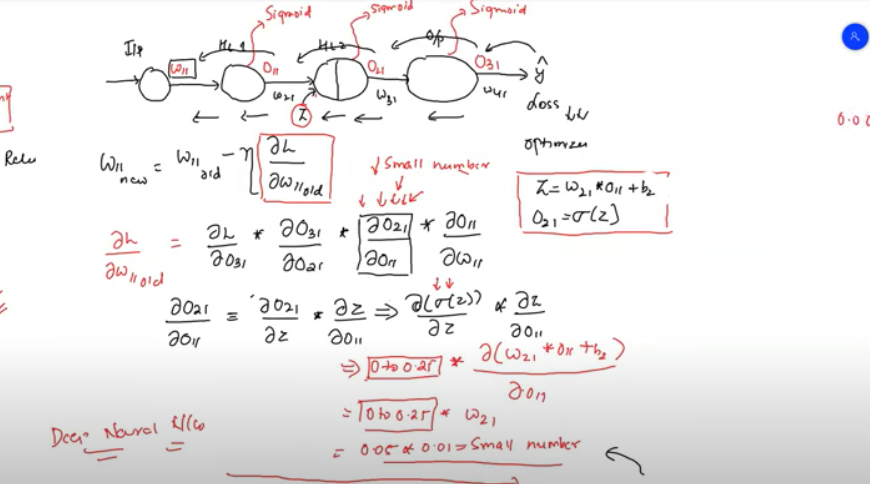


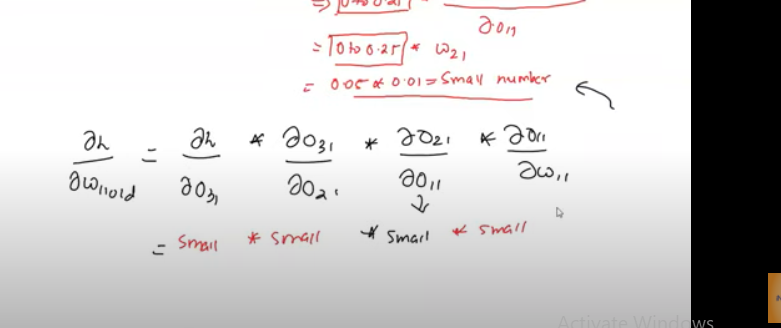
Sigmoid and tanh activation functions are undergoing with vanishing gradient problem for deep neural network

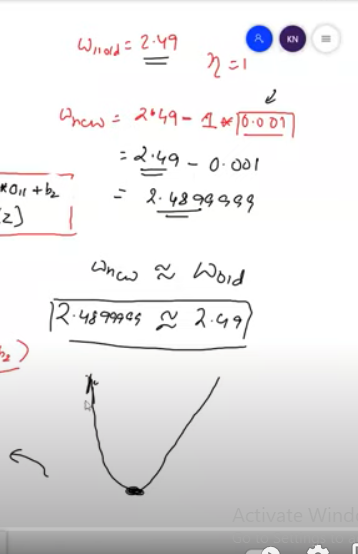


<http://ronny.rest/blog/post_2017_08_10_sigmoid/>



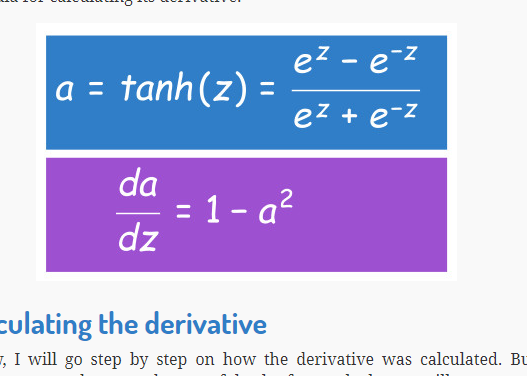
Vanishing gradient problem for increasing in deep nueral network the derivation of loss function becomes smaller with sigmoid activation function

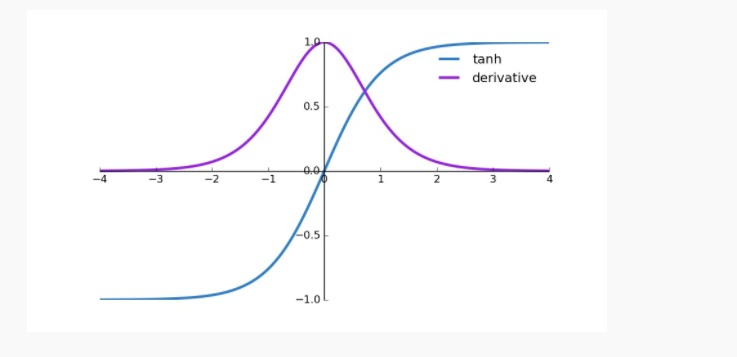


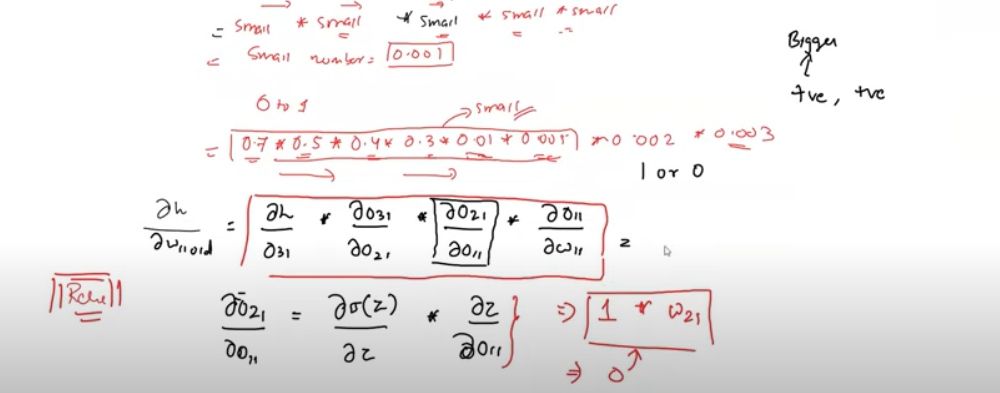
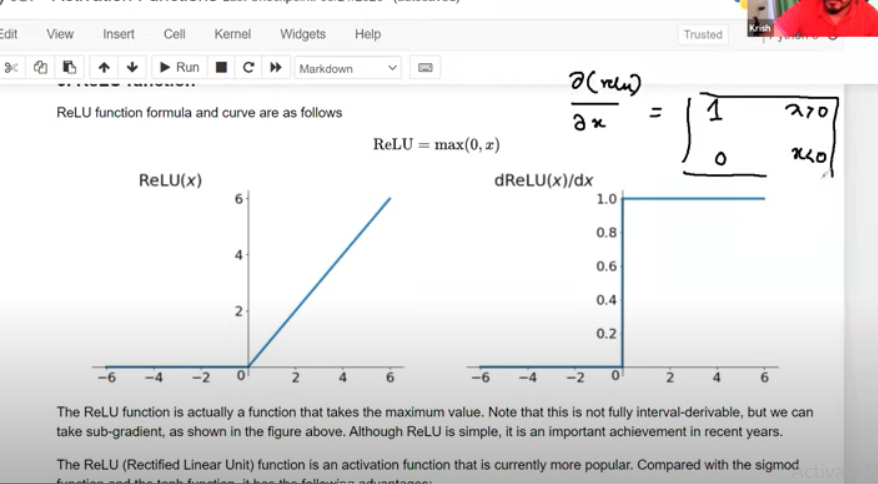
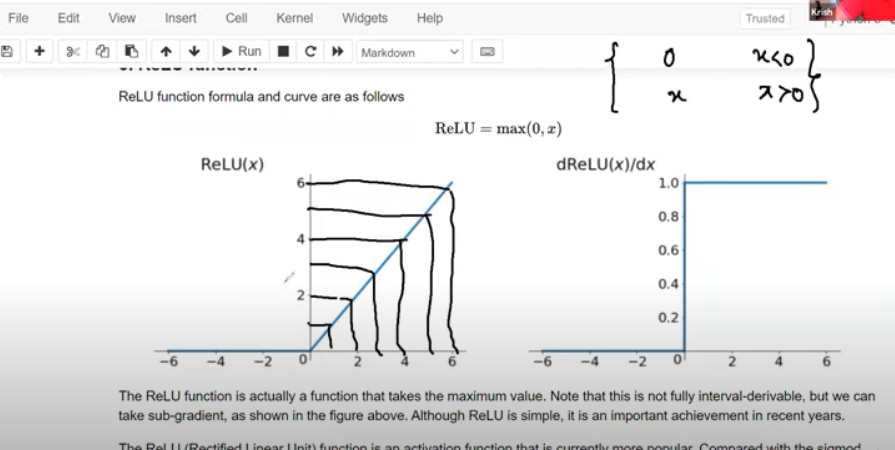
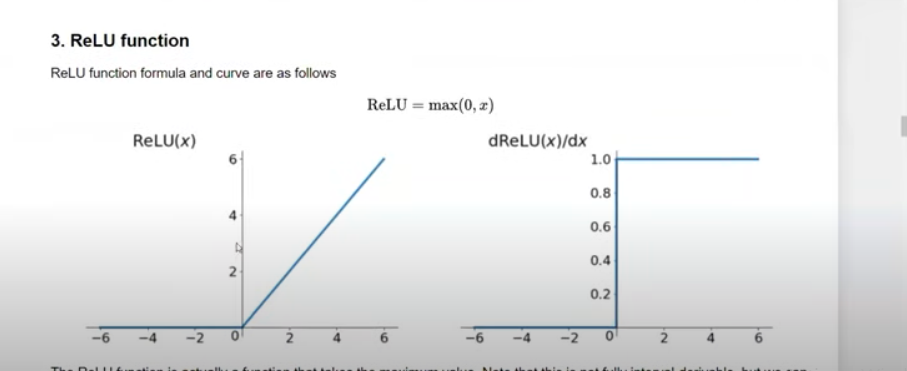


Weight updattion is not happening with the sigmoid activation function for deep neural network so it cant reach the global minima

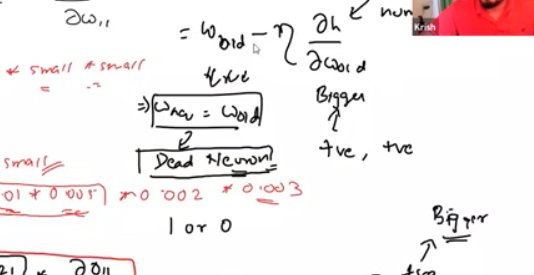
<http://ronny.rest/blog/post_2017_08_16_tanh/>





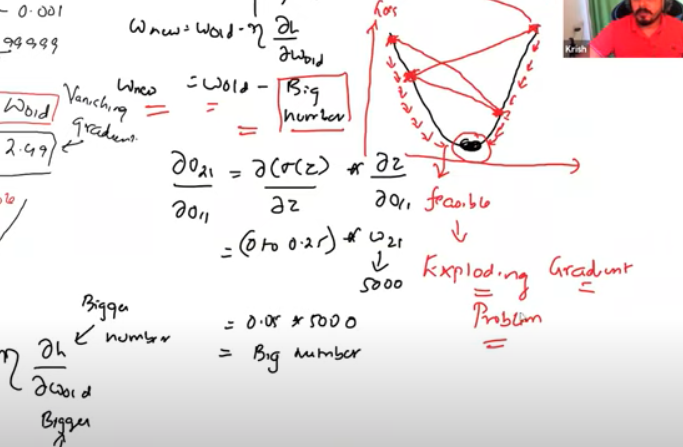
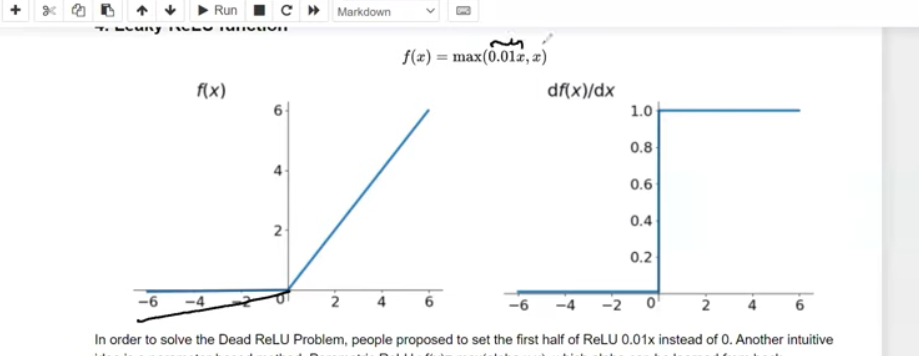
With tanh activation function also we undergo with vanishing gradient problem

Relu activation function eliminates the vanishing gradient problem because the derivative of activation function will be 0 or 1 always



If the derivative of loss function is equal to zero then the wn = wold then it is called as DEAD NEURON

IN ORDER TO FIX the dead neuron Leaky Relu activation function is used



Exploring gradient problem due to high weights