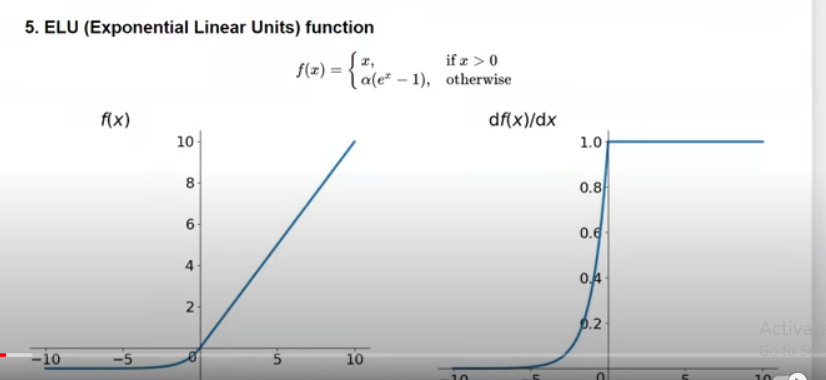
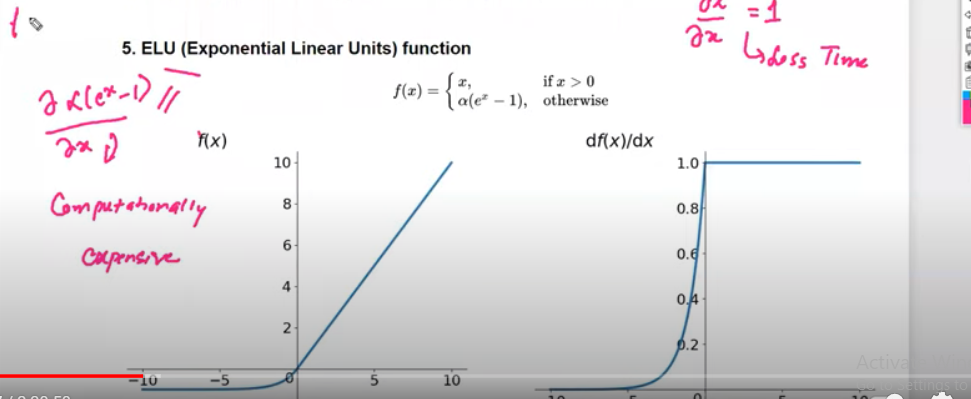


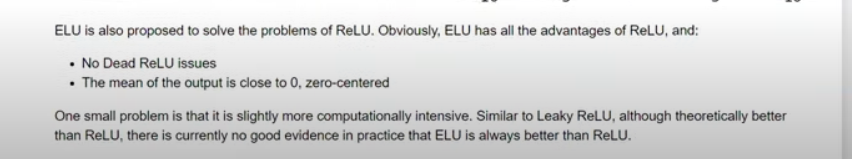
In case of sigmoid activation function data is not zero centered for this reason more computation time is taken that means time required to reach the global minima is more

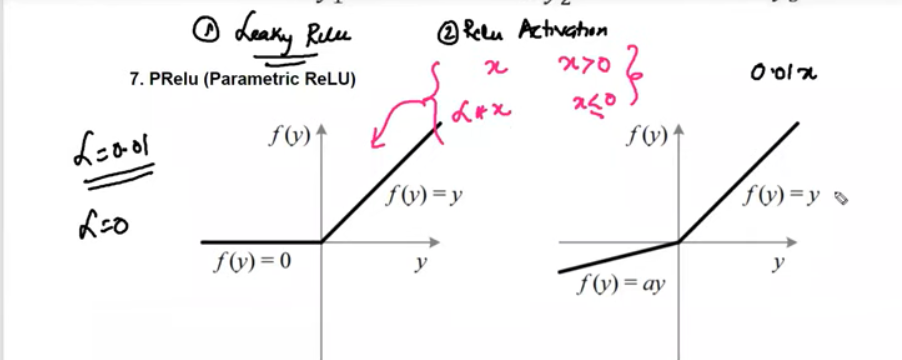
But tanh is zero centered activation function so it better than sigmoid activation function in case of time taken in reaching the global minma

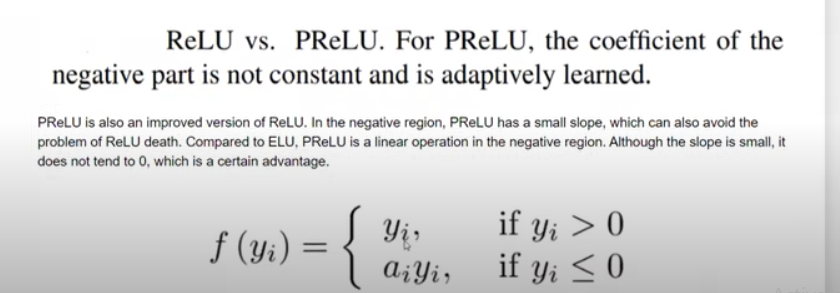
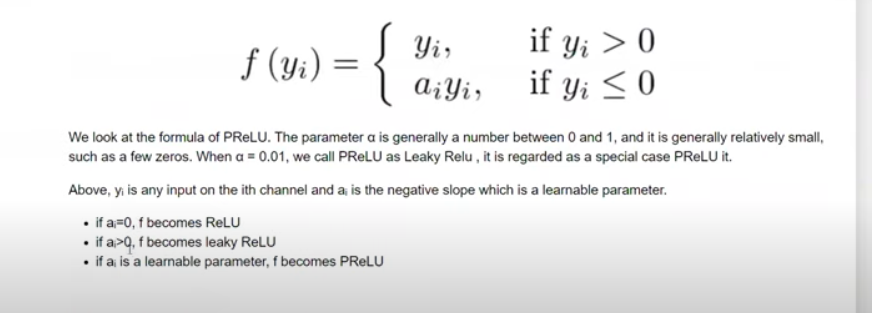
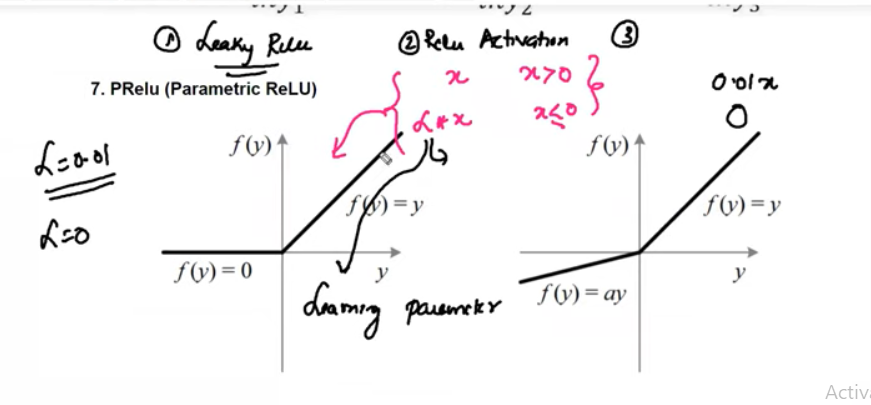
The dead neuron problem is solving by leaky relu but it will undergo vanishing gradient problem sometimes

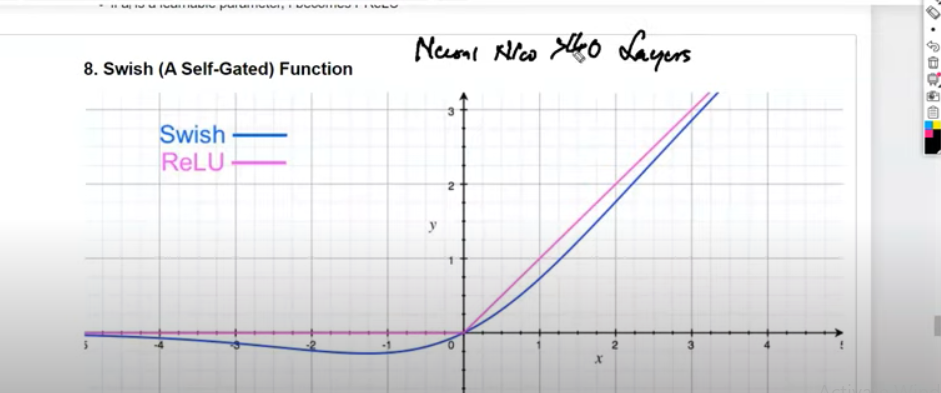


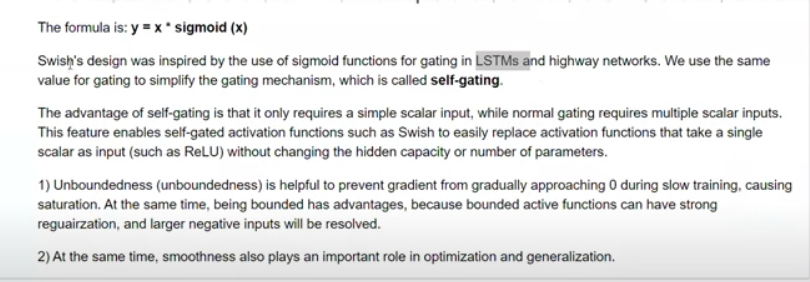


Alpha is hyper parameter since the derivative function is taking more time hence the more computational time is taken



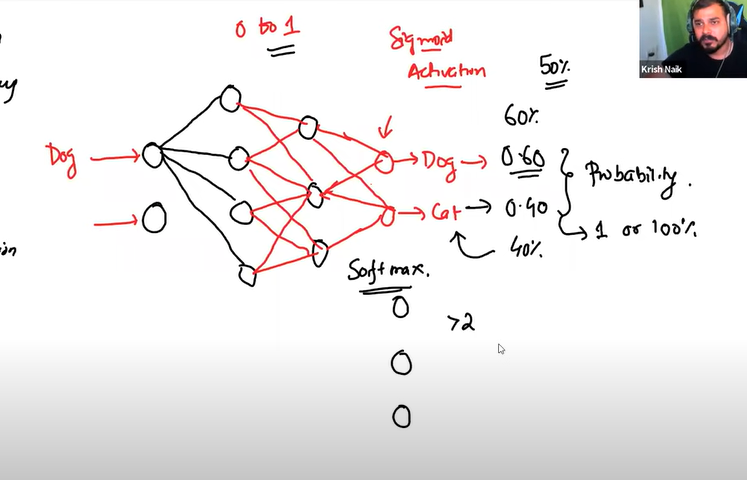


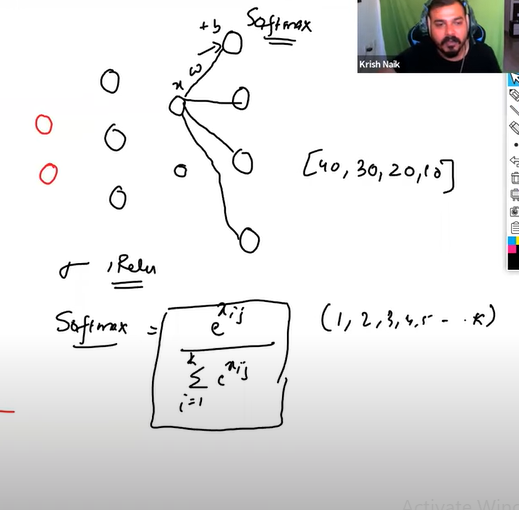
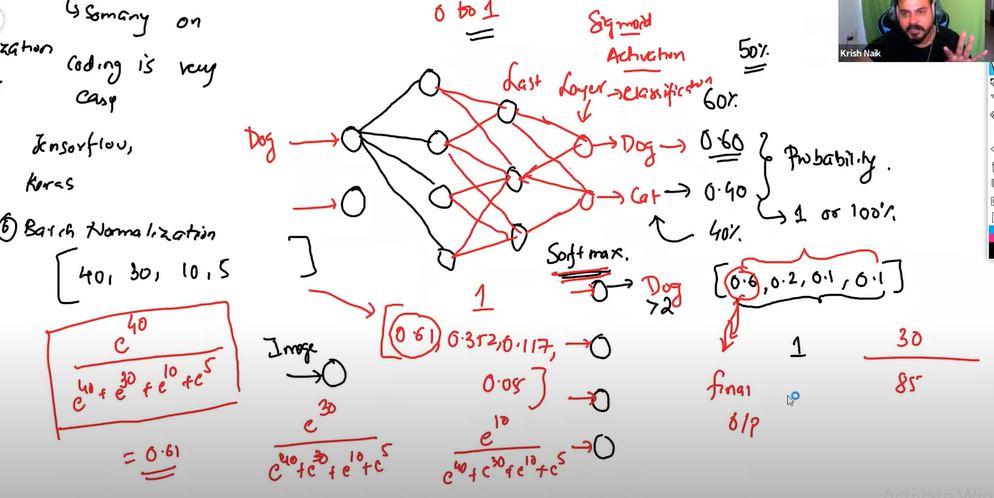
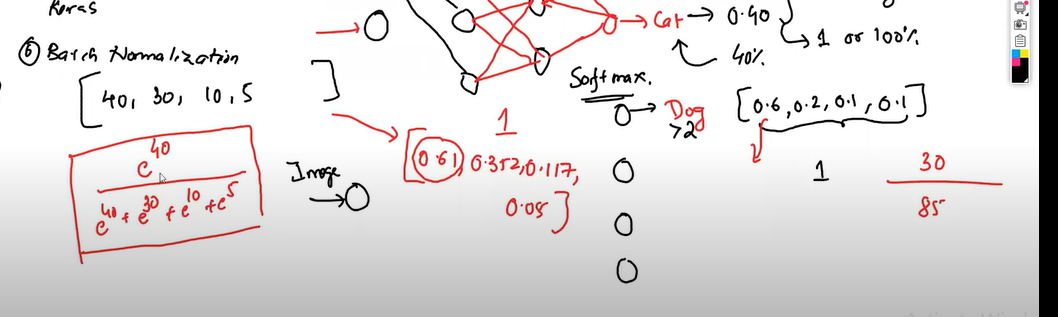
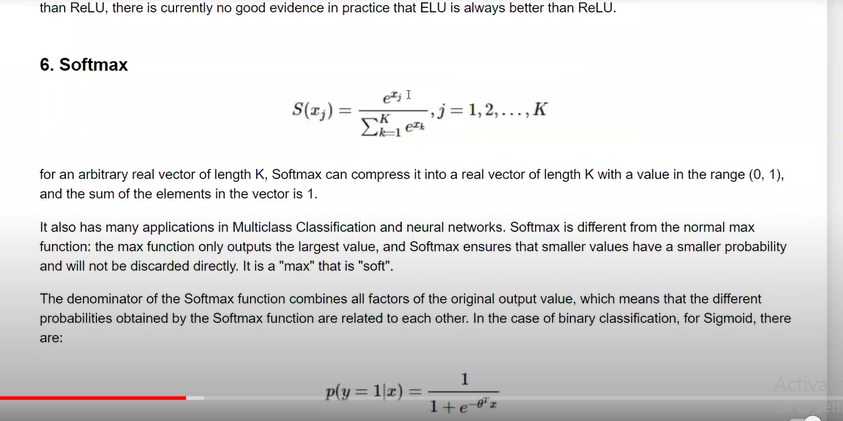


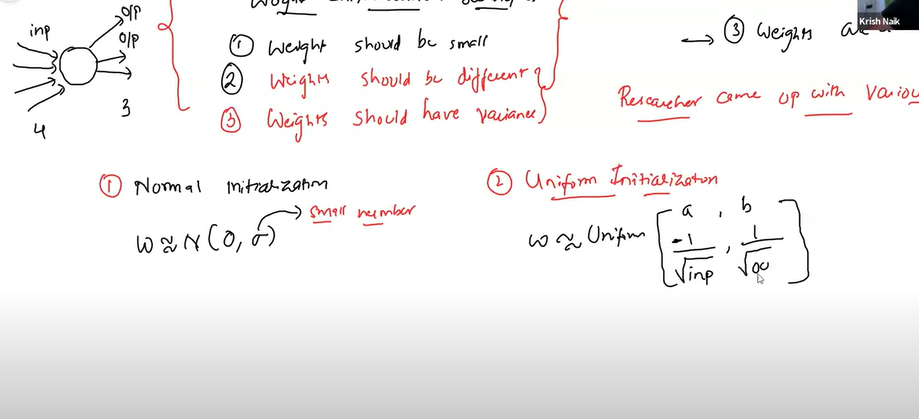
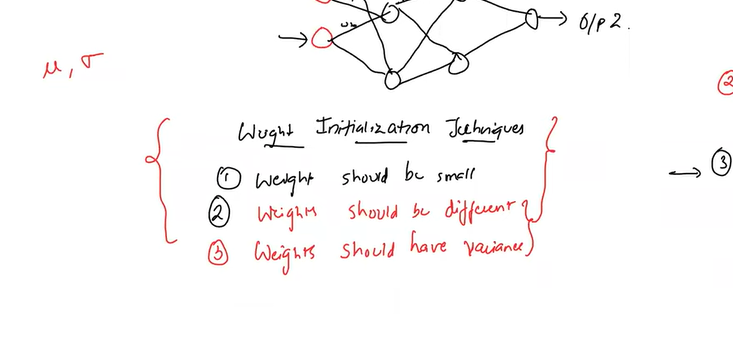
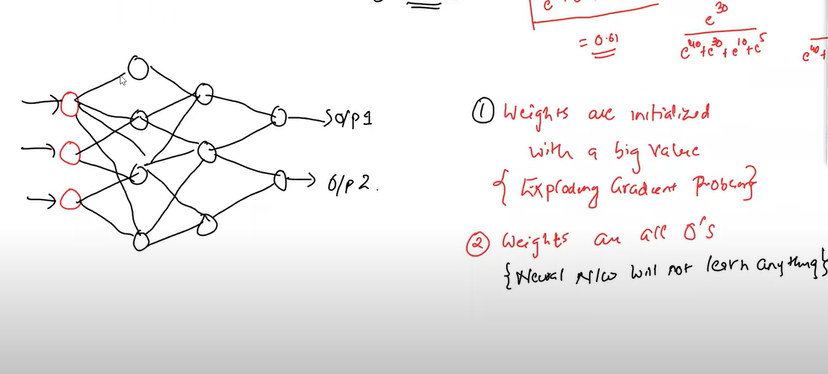


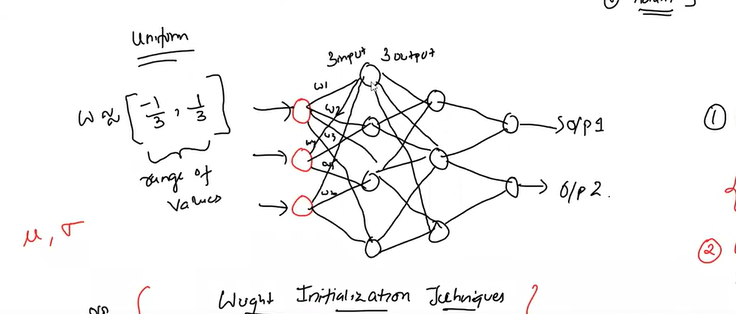
Softmax

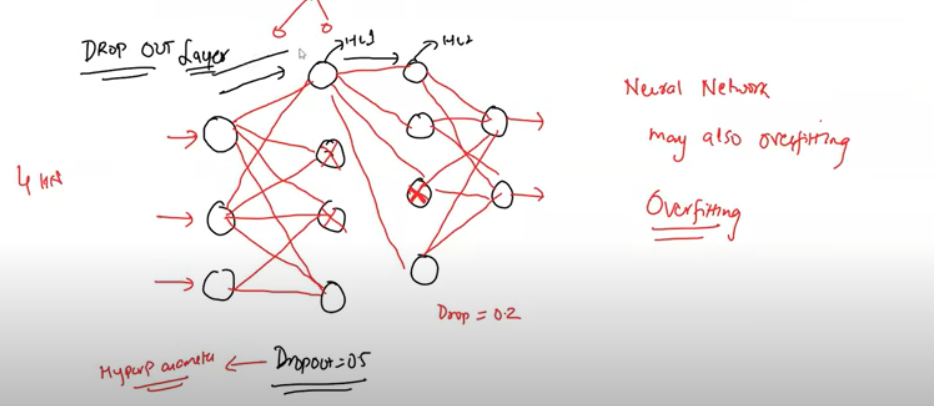
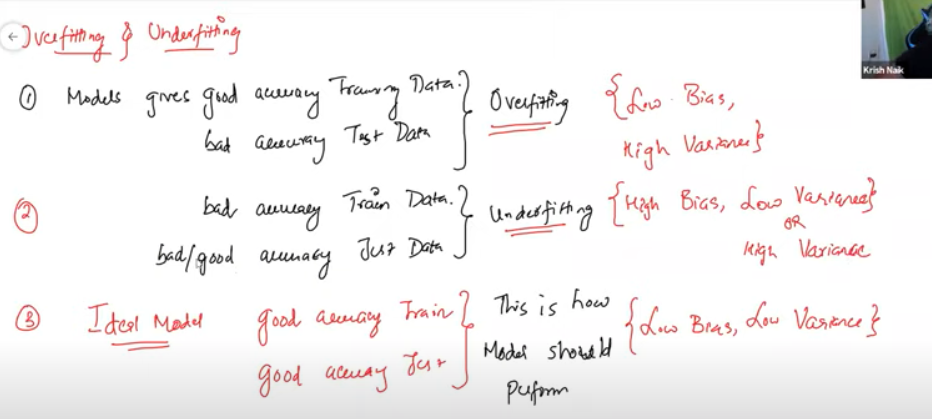
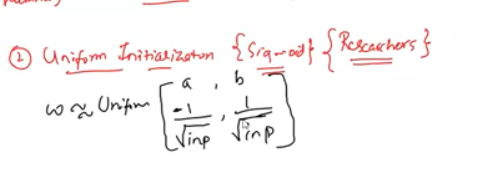
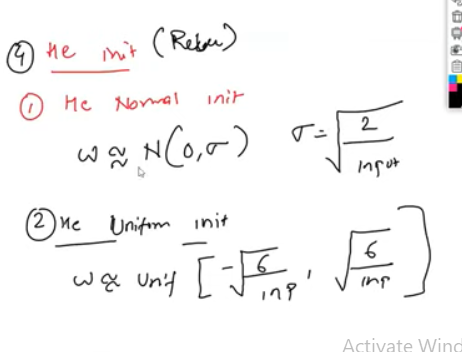
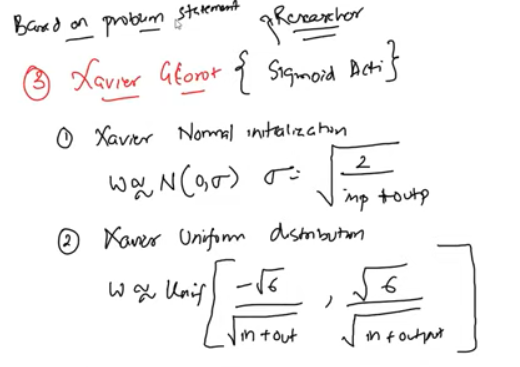
When ouputs >2 we use softmax activation function









Uniform intiialization works well with sigmoid activation function

Neural network may also overfitting to make good accuracy model drop out layer is used in which if dropout=0.5 for hidden layer then 50% of nuerons randomly gets disabled and weight intialization and activation does not happen for those nuerons in the both forward and backward propogation

In next iteration different neurons are disabled not the same neurons

For the test data the weights are multiplied by the dropout ratio by this we fix the overfitting problem and make low bias and low variance

Choosing the activation function

For hidden layers we use the different varaints of RELU ACTIVATION FUNCTION

For output layers we use the SIGMOID OR SOFTMAX (SOFTMAX IF OUTPUTS ARE >2)

