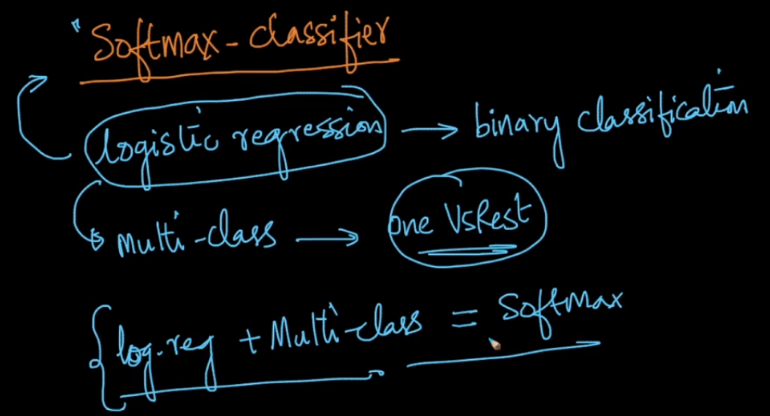
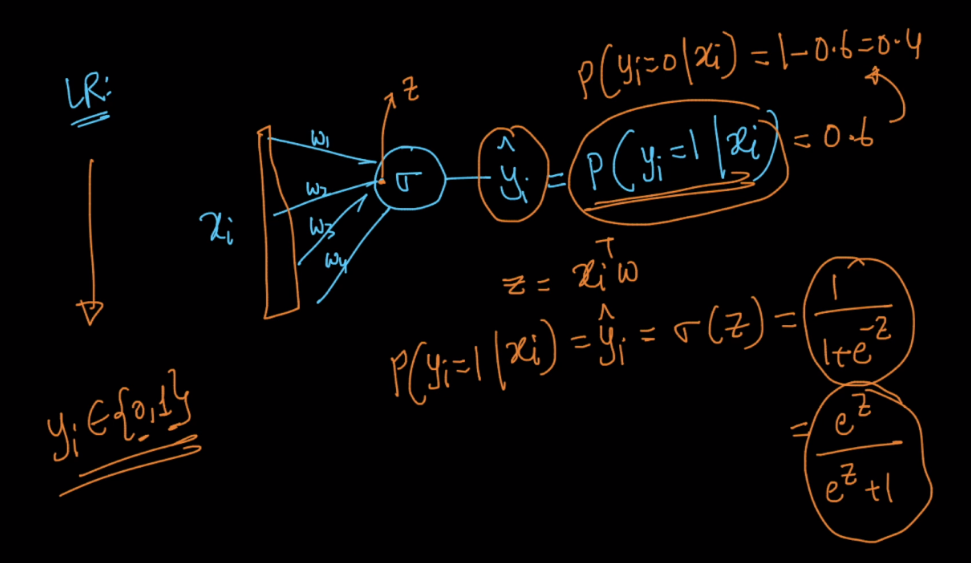
As we know that logisitic regression is a binary classification technique. For multi class classification we can perform one vs rest.

But in neural network we can extend the idea of logistic regression to perform multi class classification which is called **Softmax classifier.**

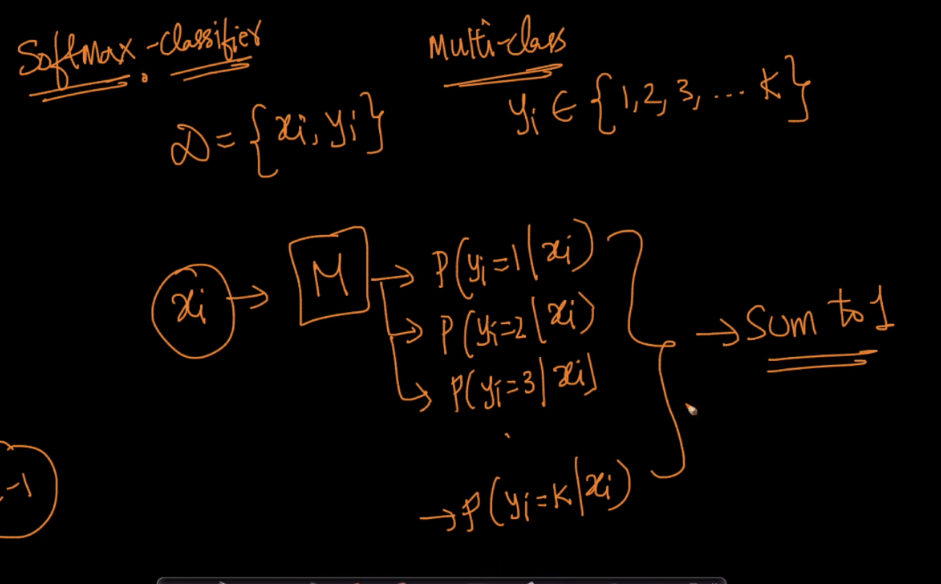


As in logistic regression the output is the probability that it’s belong to class 1, by using this probability we can find the probability of getting 0 by 1 – output\_prob, because it’s binary classification and prob\_1 + prob\_0 = 1

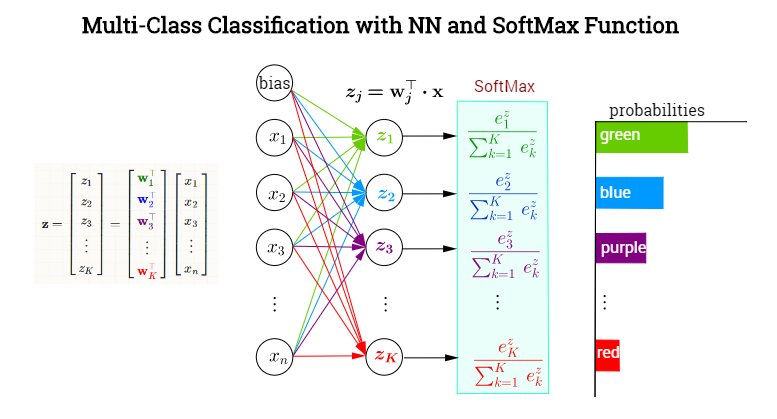
Now the formula for logistic regression is 1 / (1 + e-z ), we can generalize it by dividing by e-z in numerator and denominator and we get ez / (ez + 1 )

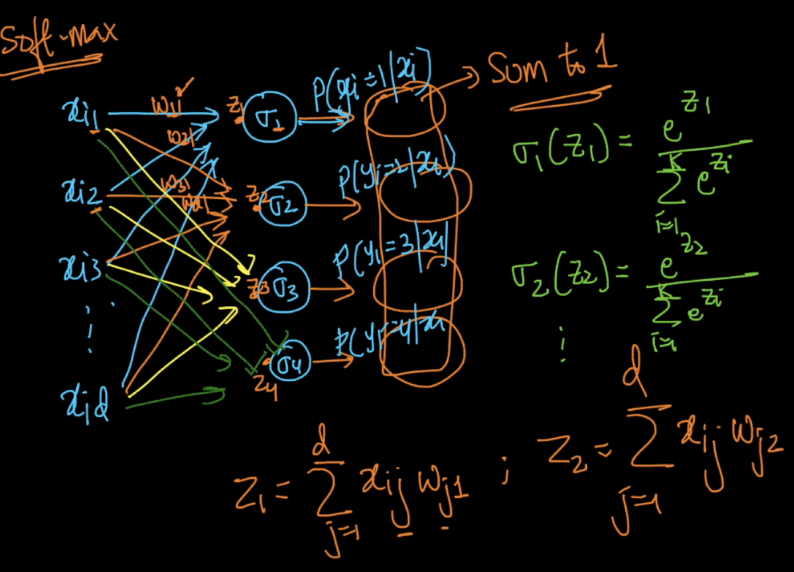


As in logistic regression we are getting 1 output which is probability of getting 1, in softmax we will get k output, one output for each class which says probability for that class, but sum of all the obtained probabilities should be 1

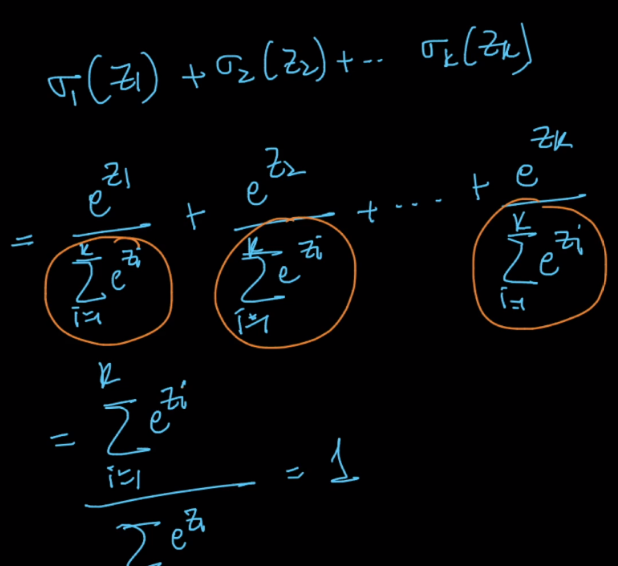


The output from each neuron is the probability of occurring that class and it is given by ez1 / sum\_i=1\_k ezi

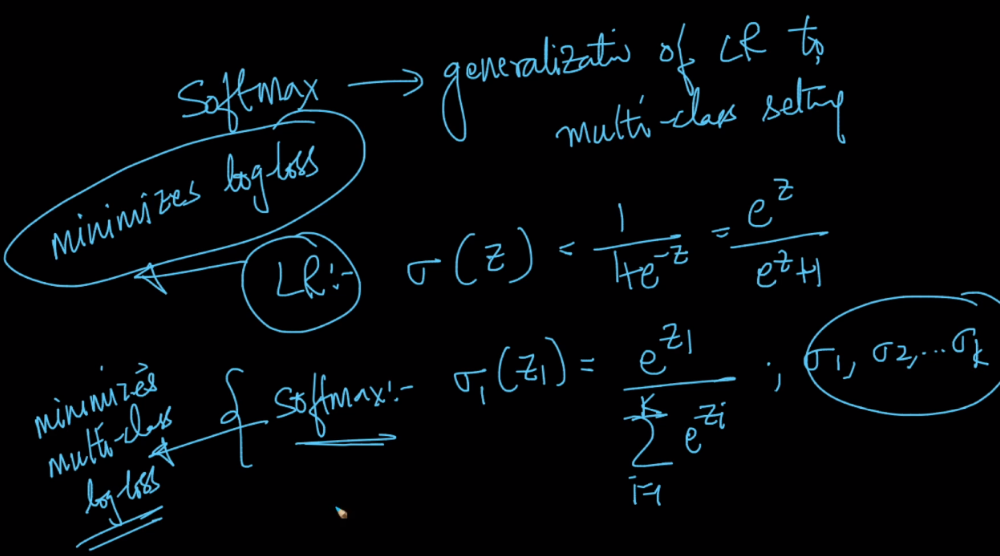




Let’s check whether the sum of all probabilities is 1, proof is given in below image

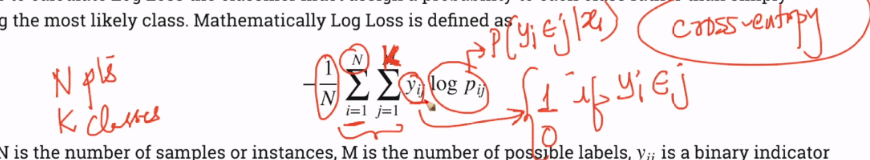


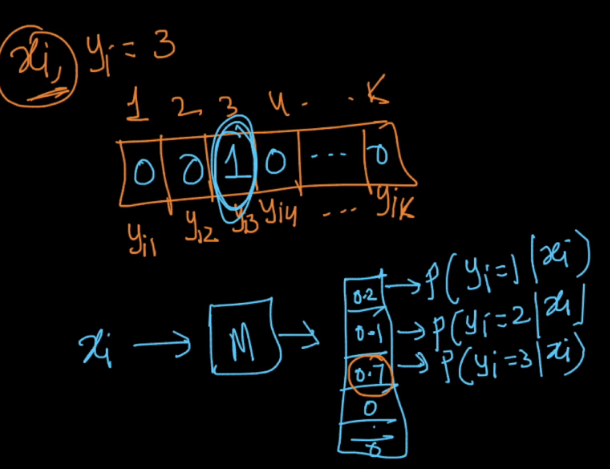
As logistic regression minimizes log loss, softmax classifier minimizes multi-class log loss



Multi class log loss is given by as given in below image which is also called cross entropy.

Here y will be 1 if the output is 1 for a particular neuron/class otherwise it will be 0, so eventually we only get the log of probability of the class for which we are finding in training time.





In Deep neural network the output layer will be softmax layer and it will have neurons equal to no of classes

