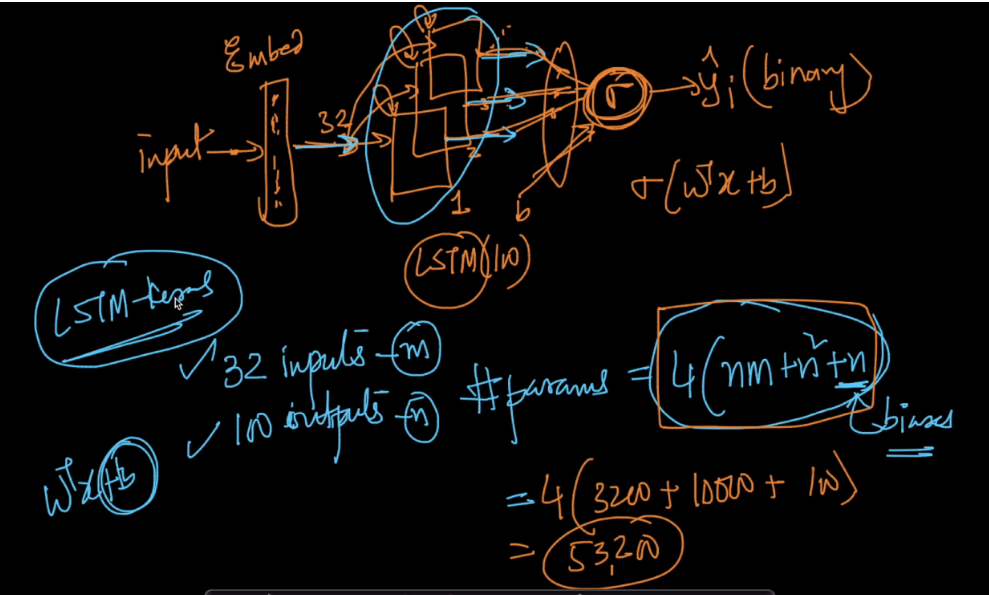
If there are m LSTM cells(or we can think of them as ‘m’ LSTMs are running parallely) and ‘n’ size of inputs then

No. of parameters = 4 ( n\*m + n2 + n ), here the last ‘n’ is bias(it can be removed if bias is not taken).



How this parameters are generated.

Remember we have four gates in LSTM cell, all four have same counts, so let’s see counts for one gate

Here h\_t-1 is the hidden state and if we use 100 LSTM cells then we take hidden state of all cells, so let’s assume there are ‘m’ such cells so it’s size is ‘m’.

X\_t is of size n, and bias is of size 1. So in total they become m+n+1.

Now let’s see number of weights, since concatenation of h\_t-1 and xt is m+n, so ofcourse weights will be of size m+n.

So eventually for parameter we have only weights, so total parameters for one cell will be m+n and +1 for bias, that means 4\*(m+n+1)

Now if we have ‘m’ cells, then total parameters for ‘m’ cells will be m\*4\*(m+n+1) ie 4\*(m2 + m\*n +1)

