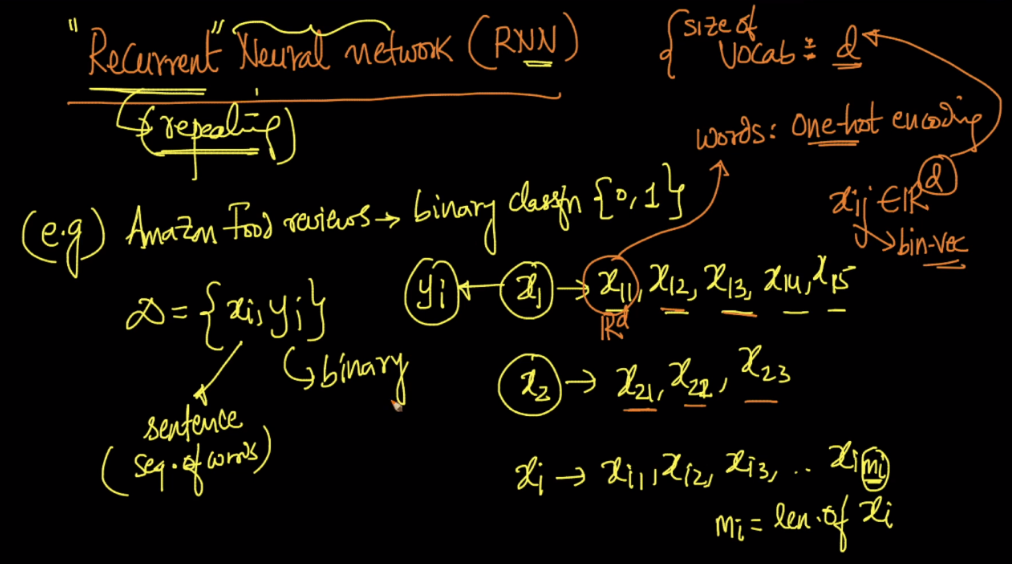
RNN: How to perform?

Lets’ take amazon fine food reviews, where input is sentence(sequence of words). Let’s say there are total ‘d’ unique words in whole training set.

Now we one hot encode each word in input sentence with ‘d’ dimensions.



We pass the first review through network which has 5 words.

Let’s say we are using network having 100 neurons.

Now at time=1 we pass first word x11 through network which outputs o1.

Here w b/w x11 and network, ‘w’ is of dim d\*100. And output o1 will be of 100 dim as there are 100 neurons.

O1 = f( w\_t \* x11 )

Now at time=2 we pass second word x12 through network(using weight w, which obtain in previous t=1) along with output of previous word ie 01, and time=2 we get o/p o2.

Here weight between o1 and network, w’ is of dim 100\*100.

O2 = f( w\_t \* x12 + w’\_t \* o1 )

Now at time=3 we pass third word x13 through network(using weight w, which obtain in previous t=2, and w’ which obtain in previous t=2) along with output of previous word ie o2 and we get o/p o3.

O3 = f( w\_t \* x13 + w’\_t \* o2 )

Similarly at time=4 we get o/p 04

Now at time=5, similarly we get o/p 05

Now this o5 will go to softmax layer which predict 0 or 1.

And now we backprpogate to update weights.

**Note:**

1. there are 3 weights:

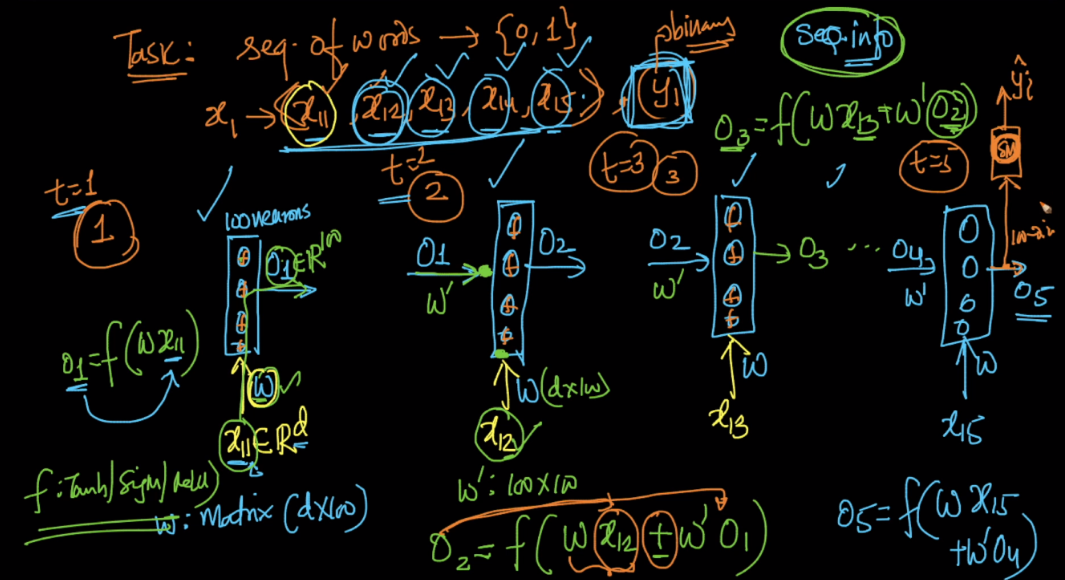
W: between word and network

W’: between previous o/p and network

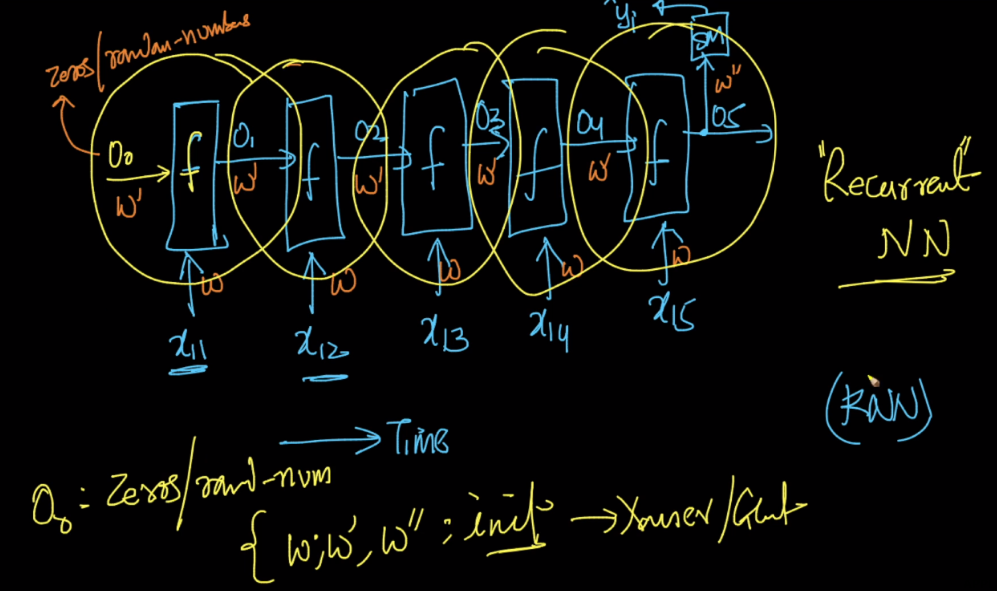
W’’: between last word o/p and softmax layer.

Those weights only updated after one iteration(completion of a sentence) and will be used for the next iteration

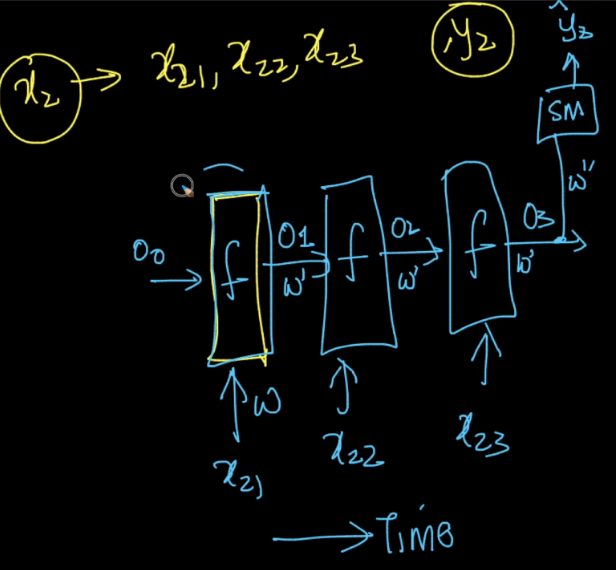
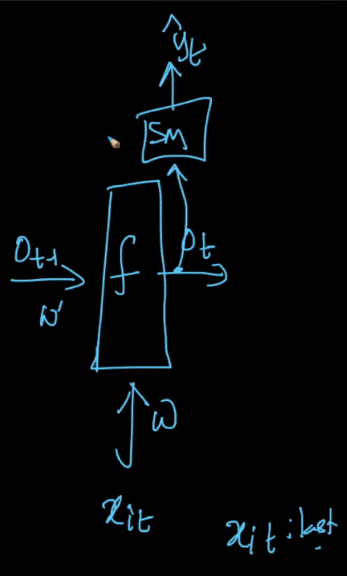
1. Here we are using one hidden layer but can use many hidden layers, and all the words have same hidden layers.
2. All sentences or inputs have same hidden layers, but different time steps based on no. of words in those sentence.
3. These all four words are passing through one network at different time step one by one, they are not connected hidden layers.



Note that at time step we are providing two inputs to network(word and o/p from previous word) except for the first word. So for first we can provide the vector of zeros or random numbers.



Now let’s say second review come and have only 3 words, so for this there will be only 3 time step and softamax will be after the output of 3rd word.

Below figure shows the box representation of RNN.

