## horizontal line

Forward Propagation in RNNs

22.04.2025

**─**

# Forward Propagation in RNN

| Review | Sentiment |
| --- | --- |
| Movie was Good | 1 |
| Movie was bad | 0 |
| Movie was not good. | 0 |

As shown there are 5 different categories (movie , was , good , bad , not).

[1,0,0,0,0], [0,1,0,0,0], [0,0,1,0,0], [0,0,0,1,0], [0,0,0,0,1]

Movie was good not bad

The input in RNN goes in the form (timesteps , input\_features)

Taking one review at a time , in the first review we have 3 words i.e for word1 (t=1), word2 (t=2), word3 (t=3).Therefore review 1 [[1,0,0,0,0],[0,1,0,0,0],[0,0,1,0,0]] has shape (3,5).

Review2 has shape (3,5) and Review3 has shape(4,5).

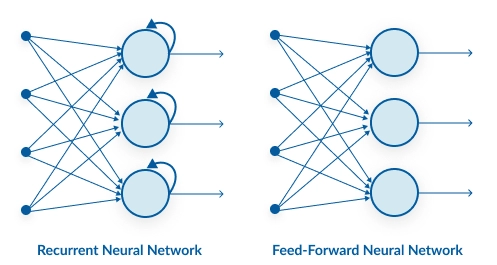
In keras , SimpleRNN has input in the form (batch\_size, timesteps, input\_features) i.e. (3,4,5).

3 reviews in a batch , 4 maximum words to be used (timesteps), 5 unique words vocab

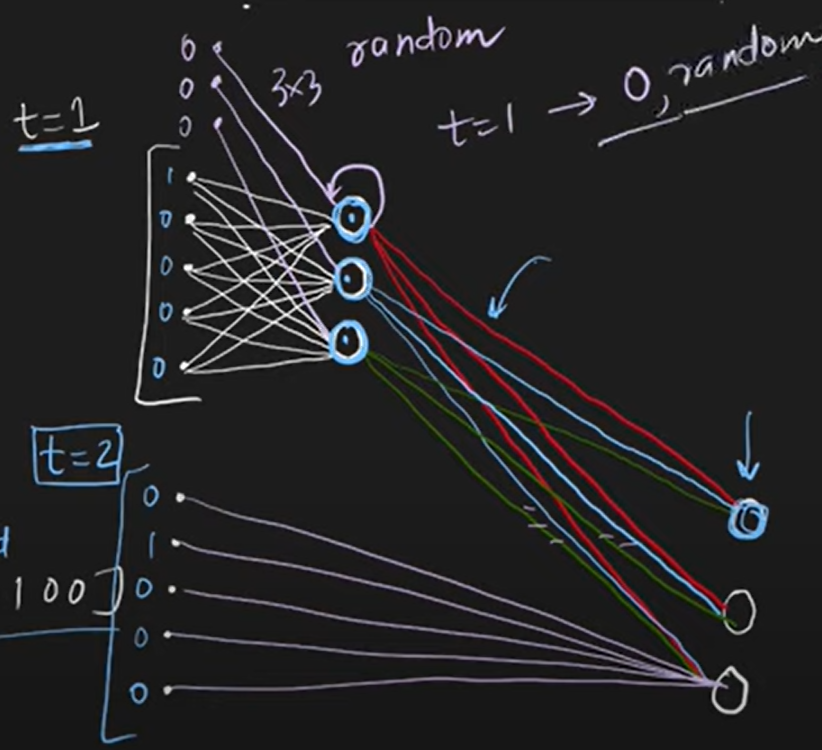
# How RNN Works ?

Simplifying the e.g.

| Review | Sentiment |
| --- | --- |
| X11 X12 X13 | 1 |
| X21 X22 X23 | 0 |
| X31 X32 X33 X34 | 0 |



A big difference between ANN and RNN is that RNN reoccurs itself in a loop in timesteps until 1 input is completed then ends it to the next layer.





Parameters :

W1 = 5x3 = 15 weights + 3 biases

W2 = 3x3 = 9 weights (reoccurring)

W3 = 3x1 = 3 weights + 1 bias In total they are 27 weights + 4 biases = 31

Every node in the hidden layer has its own activation function (tanh).

# RNN Forward Propagation

First taking review 1 : X11 , a vector of 5 dim[1,0,0,0,0]

At t=1 , X11(1,5) and Wi(5,3) so X11Wi has shape (1,3) after activation it become f(X11Wi + O0Wh) = O1



At t=2 , X12(1,5) and Wi(5,3) but also a weighted sum for t=1 will add up to this so it becomes X12Wi + O1Wh having shape ((1,3) + (1,3)(3,3)) = (1,3) . After activation it becomes f(X12Wi + O1Wh) = O2

Similarly at t=3 it would be f(X13Wi + O2Wh) = O3 having shape (1,3) .

Now review1 is completed so it reaches the next layer where O3W3 i.e (1,3)(3,1) will give you output (1,1).

For O1 there are random or zero initializations by name O0.



Here you store the information of previous time steps and at max it can store info of 10 time steps that’s why it is used for sequential data .