FORWARD PROPAGATION WITH RNN

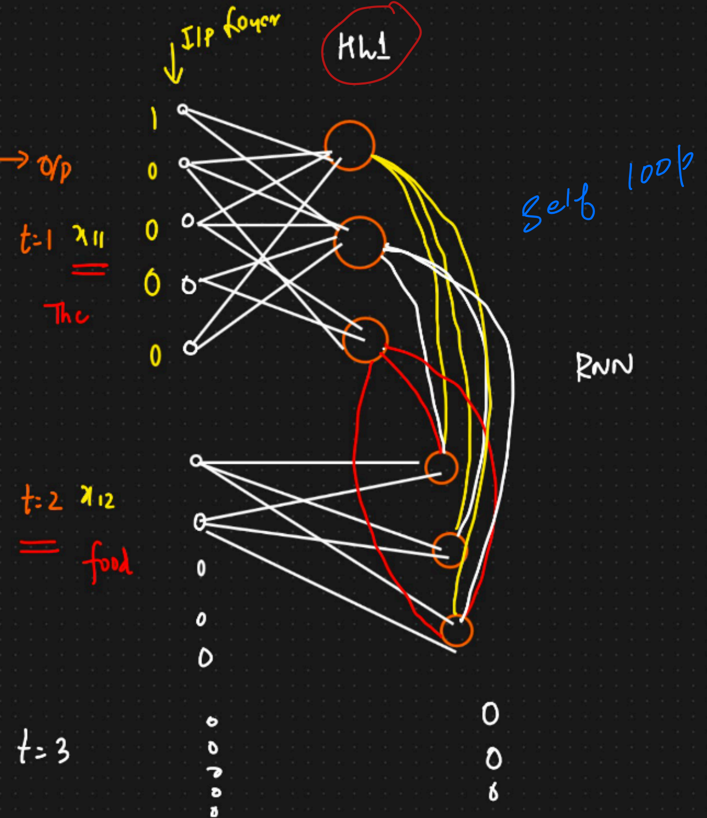
Dataset

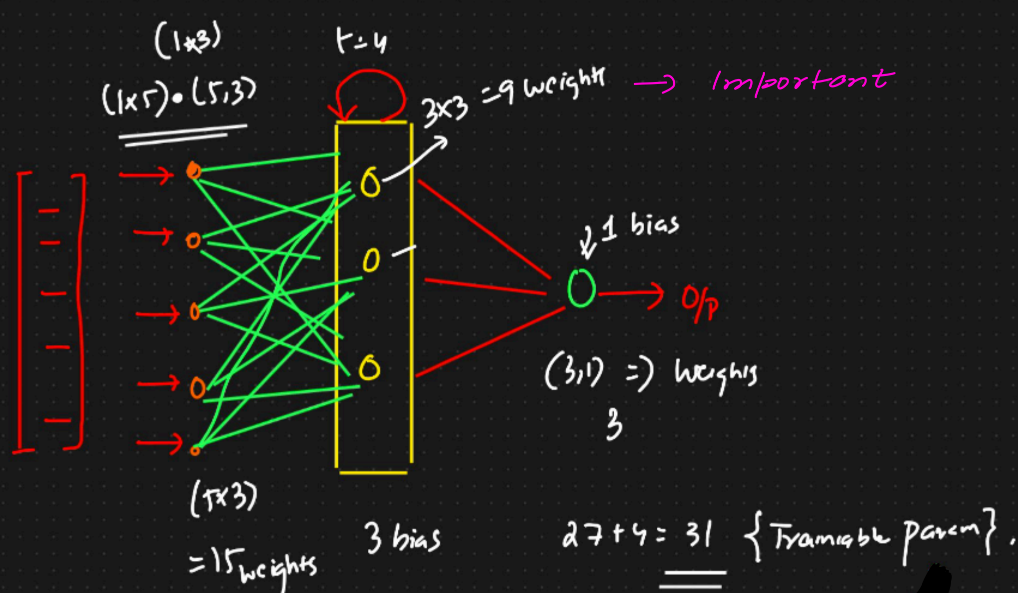
Text	O/P
The food is good (x_{11} x_{12} x_{13})	1
The food is bad	0
The food is not good	0

Words → Vectors

OHE → One hot Encoding

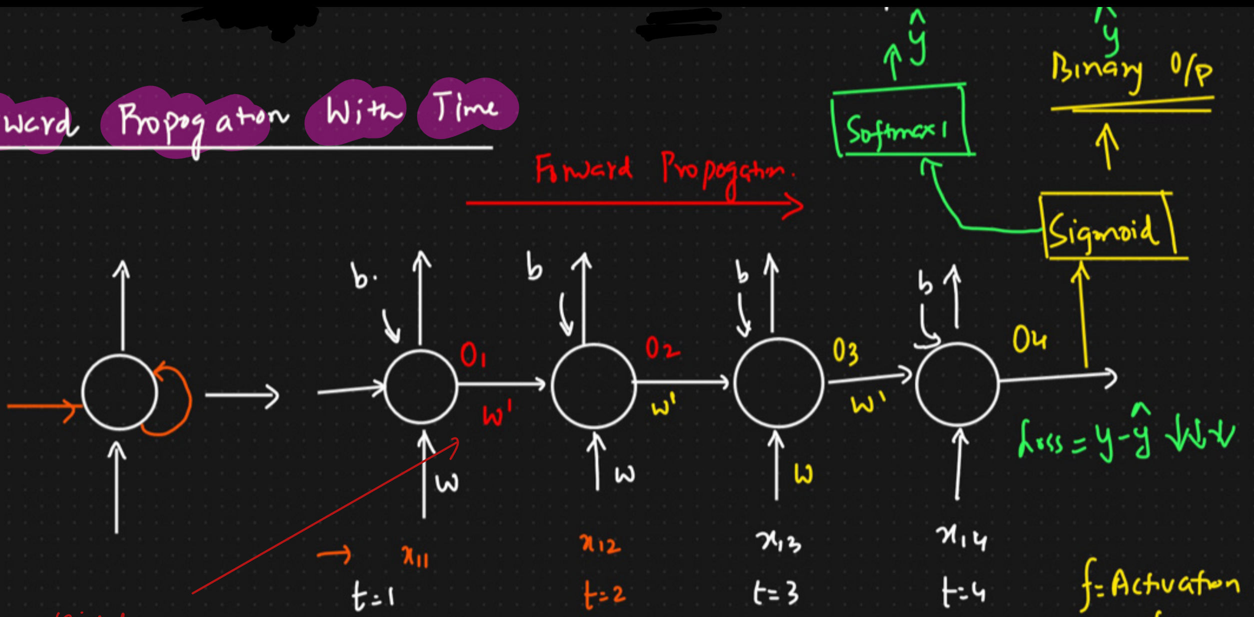
[[1 0 0 0 0] [0 1 0 0 0] [0 0 1 0 0]
[0 0 0 1 0] [0 0 0 0 1]]





27 + 4 = 31 {Trainable param}.

Forward Propogation With Time



hidden Neuron weight.

[- . . . - -]

Forward Propogation

$O_1 = f(x_{11} \cdot w + b_1)$

$O_2 = f(x_{12} * w + O_1 * w' + b_1)$

$O_3 = f(x_{13} * w + O_2 * w + b_1)$.

Dataset	O/P
The food is good	1
x_{11} x_{12} x_{13} x_{14}.	

Conclusion

1. **Input at time t:**
   You feed the input x_t (e.g., a word embedding) into the RNN.
2. **Hidden state update:**
   The RNN combines the current input and the **previous hidden state** h_{t-1} to make the new hidden state:

$$h_t = \tanh(\underbrace{W_x x_t}_{\text{current input}} + W_h h_{t-1} + b)$$ → previous input

   - W_x → weight for current input
   - W_h → weight for previous hidden state
   - b → bias
   - **tanh** adds non-linearity so it can model complex patterns.
3. **Output at time t:**
   From the new hidden state, the RNN predicts the output:

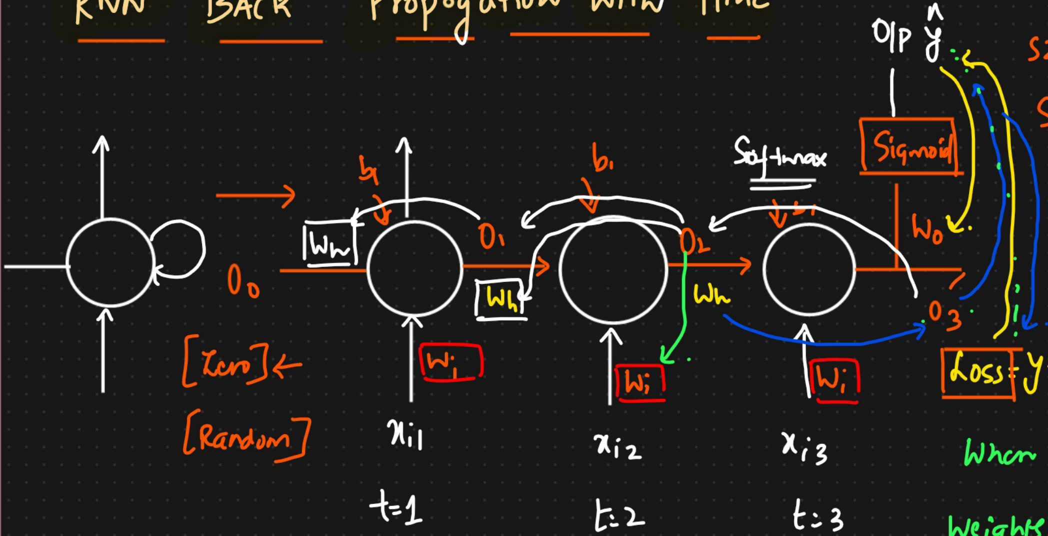
$$y_t = W_y h_t + b_y$$

4. **Repeat for next time step:**
   The new hidden state h_t becomes h_{t-1} for the next step.

💡 Think of it like:
"Current Output = Function(Current Input + Memory from the Past)"

→ This is really important

RNN BACK Propogation With Time



				O/P
S_1	$\langle x_{11}$	x_{12}	$x_{13}\rangle$	1
S_2	$\langle x_{21}$	x_{22}	$x_{23}\rangle$	0
S_3	$\langle x_{31}$	x_{32}	$x_{33}\rangle$.	

$\langle x_{i1} \quad x_{i2} \quad x_{i3}\rangle$.

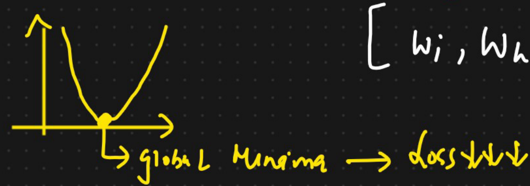
i = Sentence No.

When we update all the weights

$[W_i, W_h, W_o] \leftarrow$ Back Propogation

Forward Propogation

f = tanh



$$O_1 = f(x_{i1} * W_i + O_0 W_h + b_1)$$

$$O_2 = f(x_{i2} * W_i + O_1 * W_h + b)$$

$$O_3 = f(x_{i3} * W_i + O_2 * W_h + b).$$

$$\hat{y} = \sigma(O_3 * W_o).$$

Backward Propogation with Time

Update $[W_i, W_h, W_o]$

Weight updation formula

$$W_{new} = W_{old} - \eta \boxed{\frac{\partial L}{\partial W_{old}}} \leftarrow$$

Derivative, slope of GRADIENT DESCENT

$\eta = 0.001$

$$\boxed{W_{o_{new}} = W_{o_{old}} - \eta \frac{\partial L}{\partial W_{o_{old}}}}$$

Imp update weights (Back prop.)

- Initial weights
- hidden weights
- output weights

① Update W_o

Chain Rule of Derivative

$$\frac{\partial L}{\partial W_{o_{old}}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial W_{old}}$$

② Update W_h. [Hidden layer Weights] → Time Stamps

$$W_{h_{new}} = W_{h_{old}} - \eta \boxed{\frac{\partial L}{\partial W_{h_{old}}}}$$

t=1,2,3

$$\Rightarrow \frac{\partial L}{\partial W_{h\,old}} = \overbrace{\left[\frac{\partial L}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial O_3} * \frac{\partial O_3}{\partial W_h}\right]}^{t=3} + \overbrace{\left[\frac{\partial L}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial O_3} * \frac{\partial O_3}{\partial O_2} * \frac{\partial O_2}{\partial W_h}\right]}^{t=2}$$

$$+ \overbrace{\left[\frac{\partial L}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial O_3} * \frac{\partial O_3}{\partial O_2} * \frac{\partial O_2}{\partial O_1} * \frac{\partial O_1}{\partial W_h}\right]}^{t=1}$$

$\longleftarrow$

update weights w.r.t w.

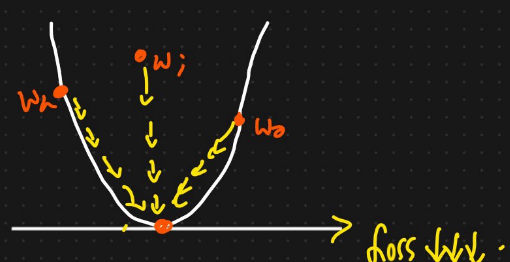
③ Updating Weights W_i $\longrightarrow$ Timestamp

{update W_i weights}

$$W_{i\,new} = W_{i\,old} - \eta \boxed{\frac{\partial L}{\partial W_{i\,old}}}$$

$$\boxed{\frac{\partial L}{\partial W_{i\,old}}} = \overbrace{\left[\frac{\partial L}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial O_3} * \frac{\partial O_3}{\partial W_{i\,old}}\right]}^{t=3 \quad h_3} + \overbrace{\left[\frac{\partial L}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial O_3} * \frac{\partial O_3}{\partial O_2} * \frac{\partial O_2}{\partial W_{i\,old}}\right]}^{t=2 \quad h_2}$$

$$+ \overbrace{\left[\frac{\partial L}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial O_3} * \frac{\partial O_3}{\partial O_2} * \frac{\partial O_2}{\partial O_1} * \frac{\partial O_1}{\partial W_{i\,old}}\right]}^{t=1 \quad h_1}$$



$\longrightarrow$ close to Global minima.