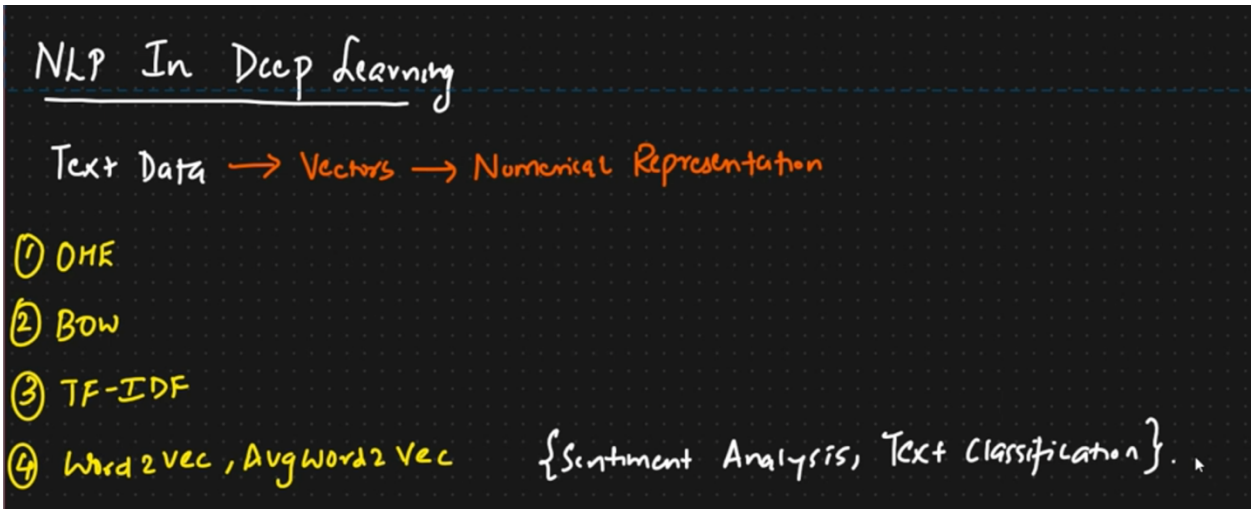
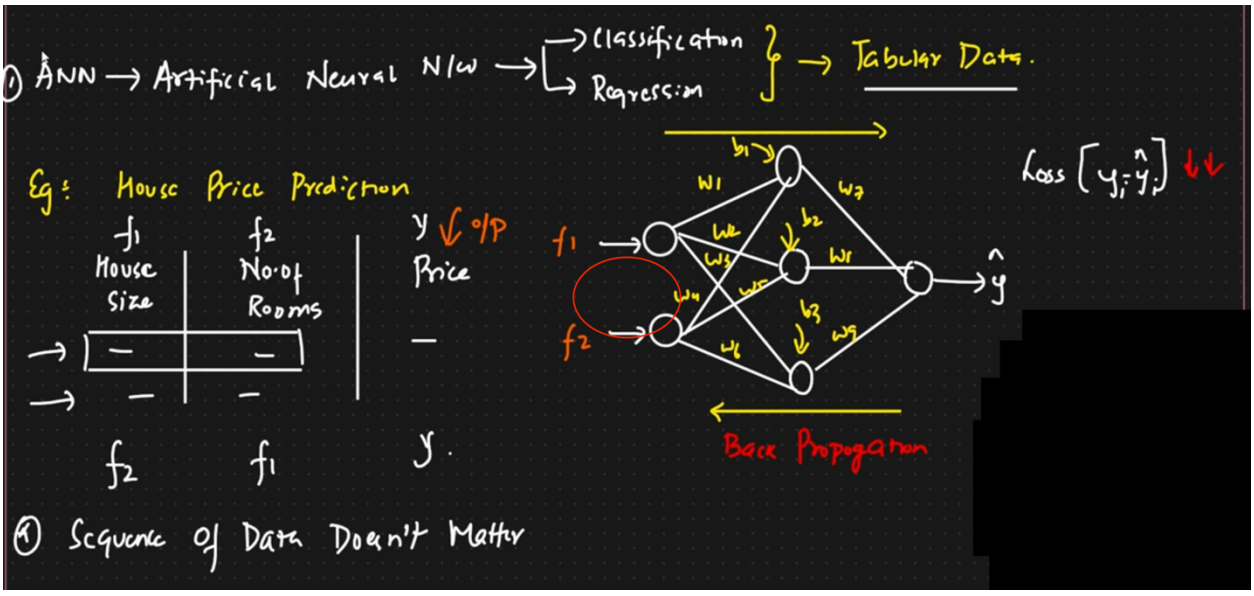
Introduction to NLP in Deep Learning





③ Data → Sequential Data.

① Text Generation → I/P: This is a Apple → juice

Sentence.

② Chatbot Conversation → Q&A — Question → Answer

Sequential Data → The food is good

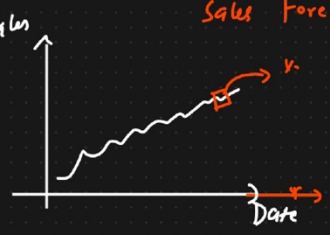
{good The food is} → Meaning may change

② Language Translation → English → ... → French → .....

④ Auto Suggestion → LinkedIn, GMAIL → Autosuggestion

⑤ Sales Data → DateTime

Sales Forecasting



Can we use ANN to solve this Problem? → Sequential Data.

↓

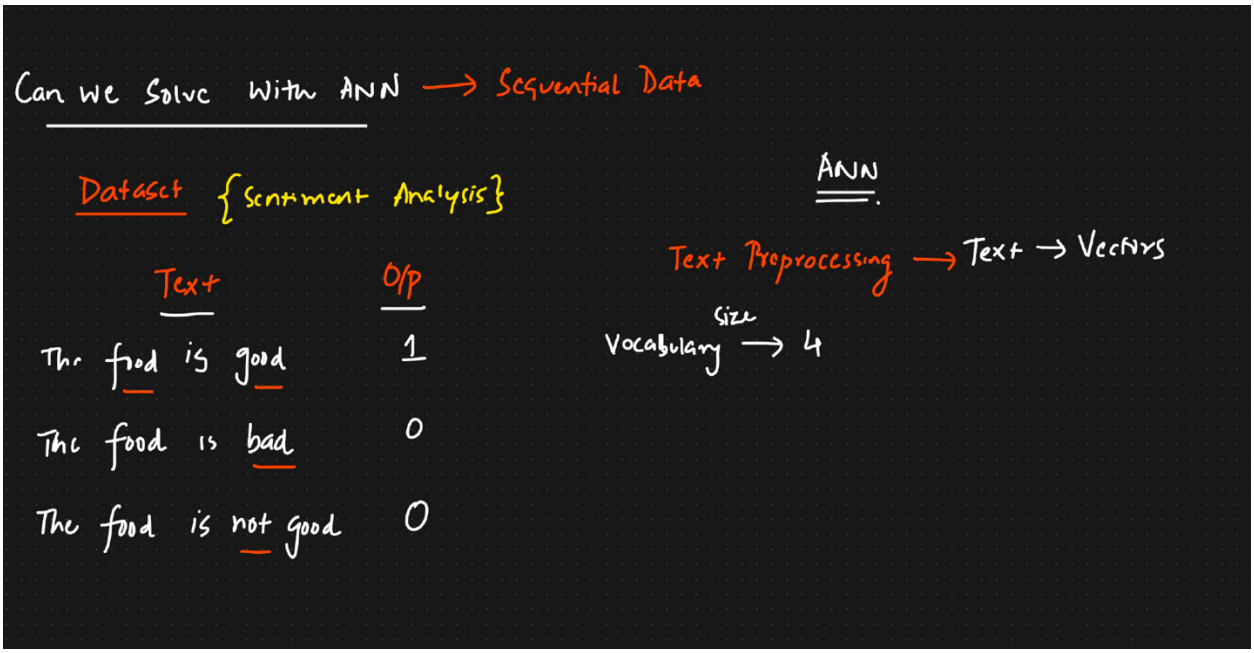
NLP In Deep Learning [Generative AI → LLM, MultiModel]

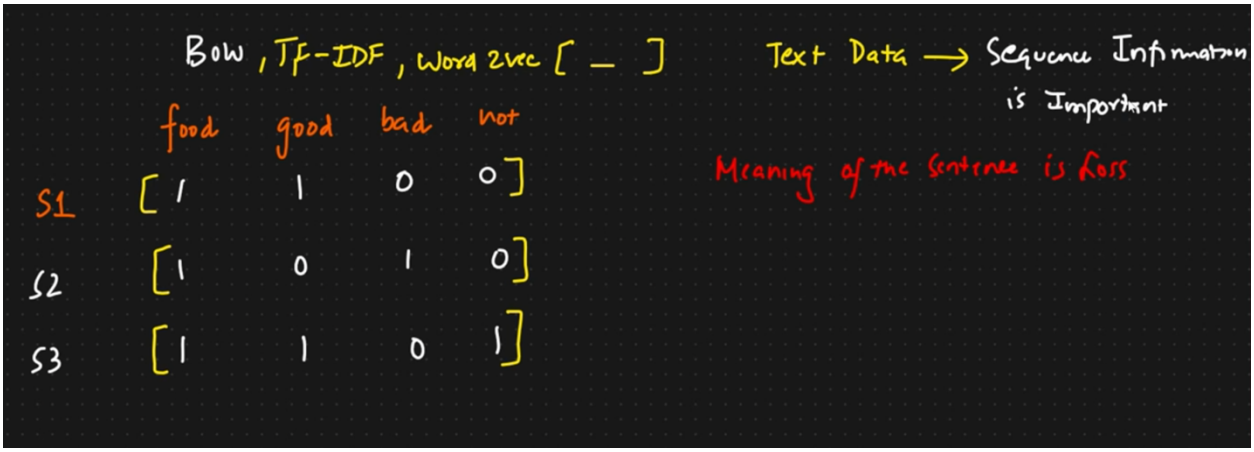
① Simple RNN → LSTM / GRU RNN → Bidirectional RNN → Encoder Decoder

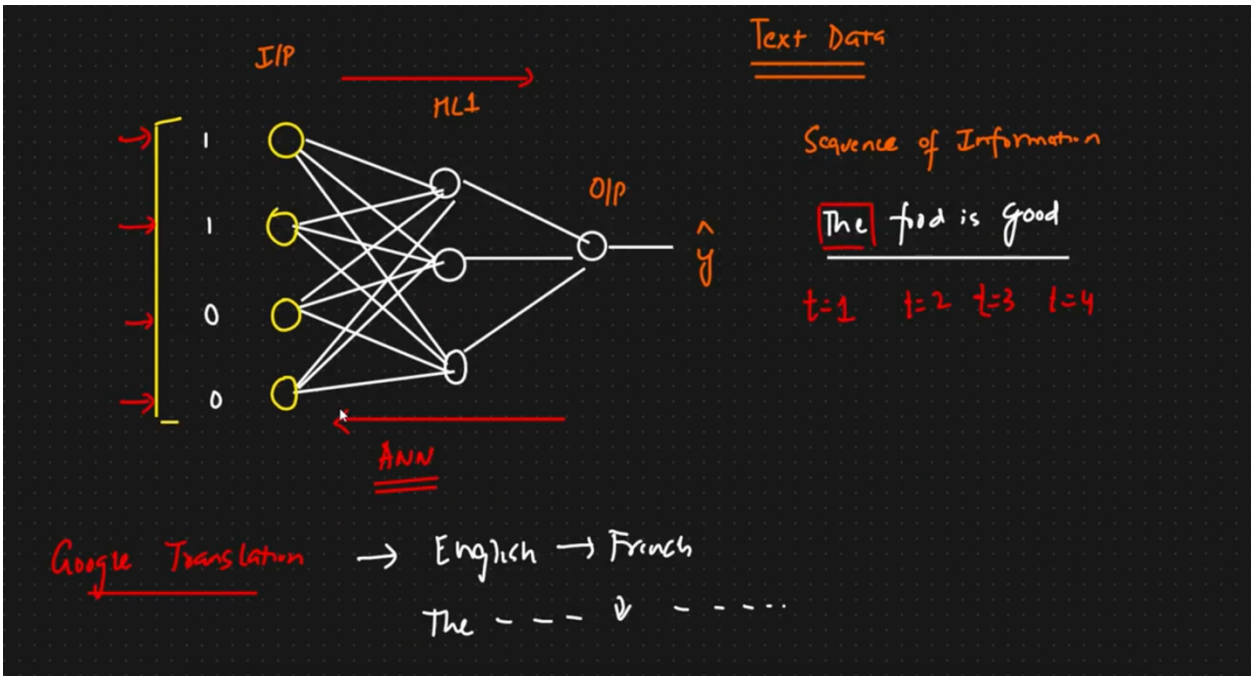
↓

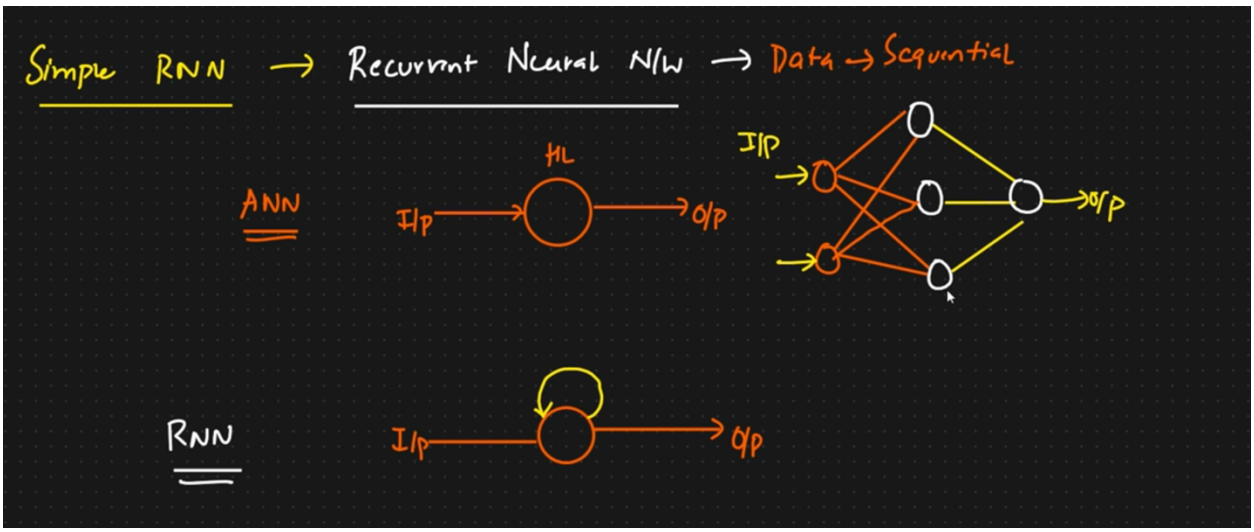
Transformers ← Self Attention

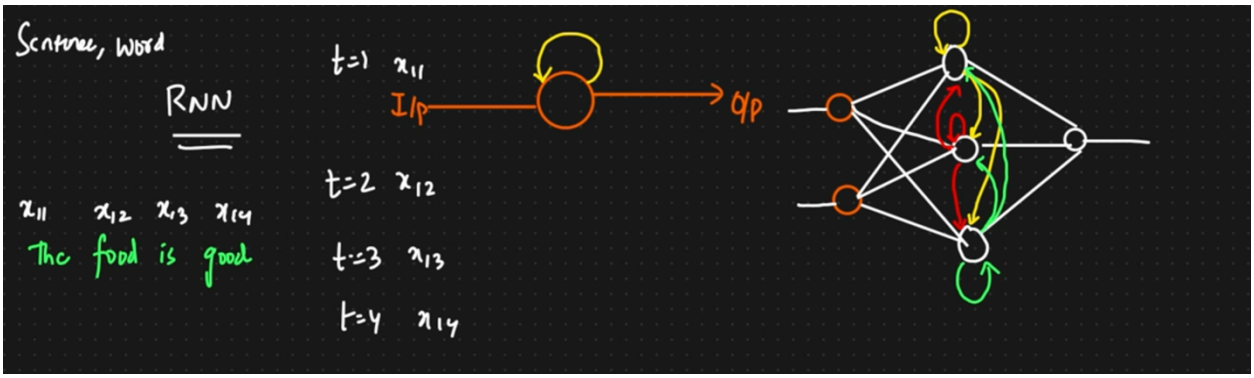
ANN VS RNN

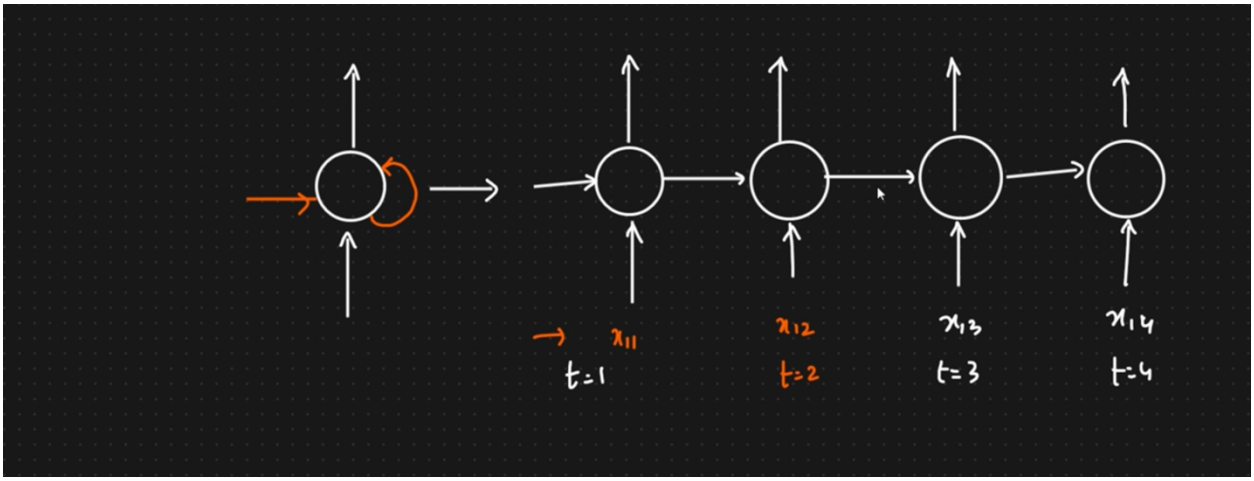

Can we Solve With ANN → Sequential Data
Dataset {Sentiment Analysis}
Text
O/p
The food is good
1
The food is bad
0
The food is not good
0
ANN
Text Preprocessing → Text → Vectors
Vocabulary Size → 4


BOW, TF-IDF, Word2vec [_]
food good bad not
S1 [1 1 0 0]
S2 [1 0 1 0]
S3 [1 1 0 1]
Text Data → Sequence Information is Important
Meaning of the Sentence is Loss

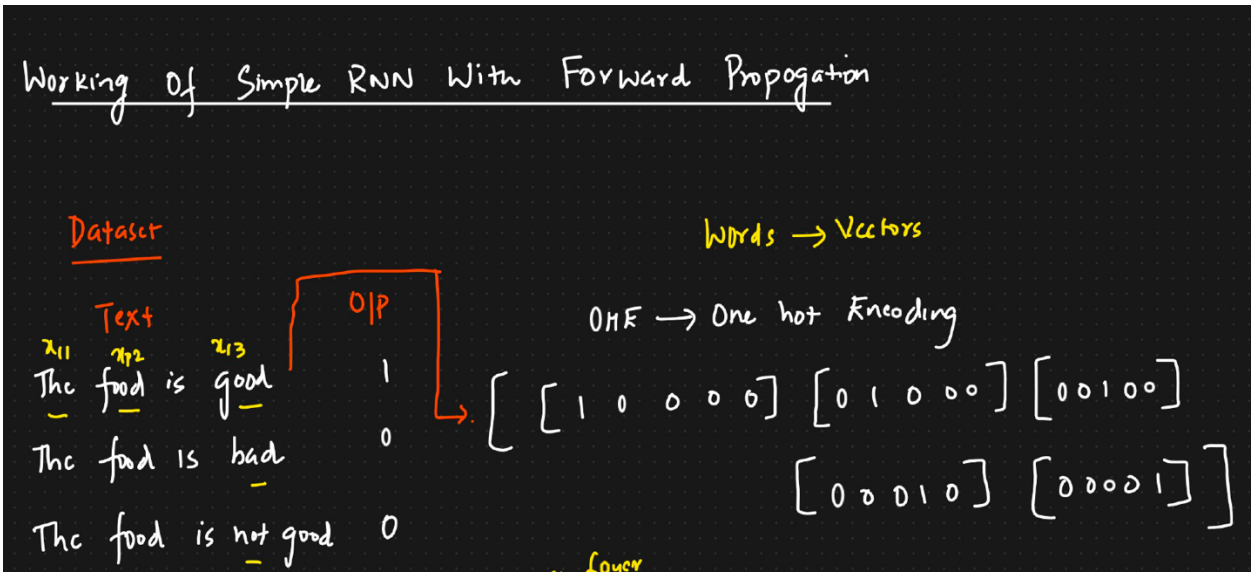

I/P
HL1
O/P
ŷ
1
1
0
0
ANN
Text Data
Sequence of Information
The food is good
t=1 t=2 t=3 t=4
Google Translation → English → French
The - - - - - - - - -

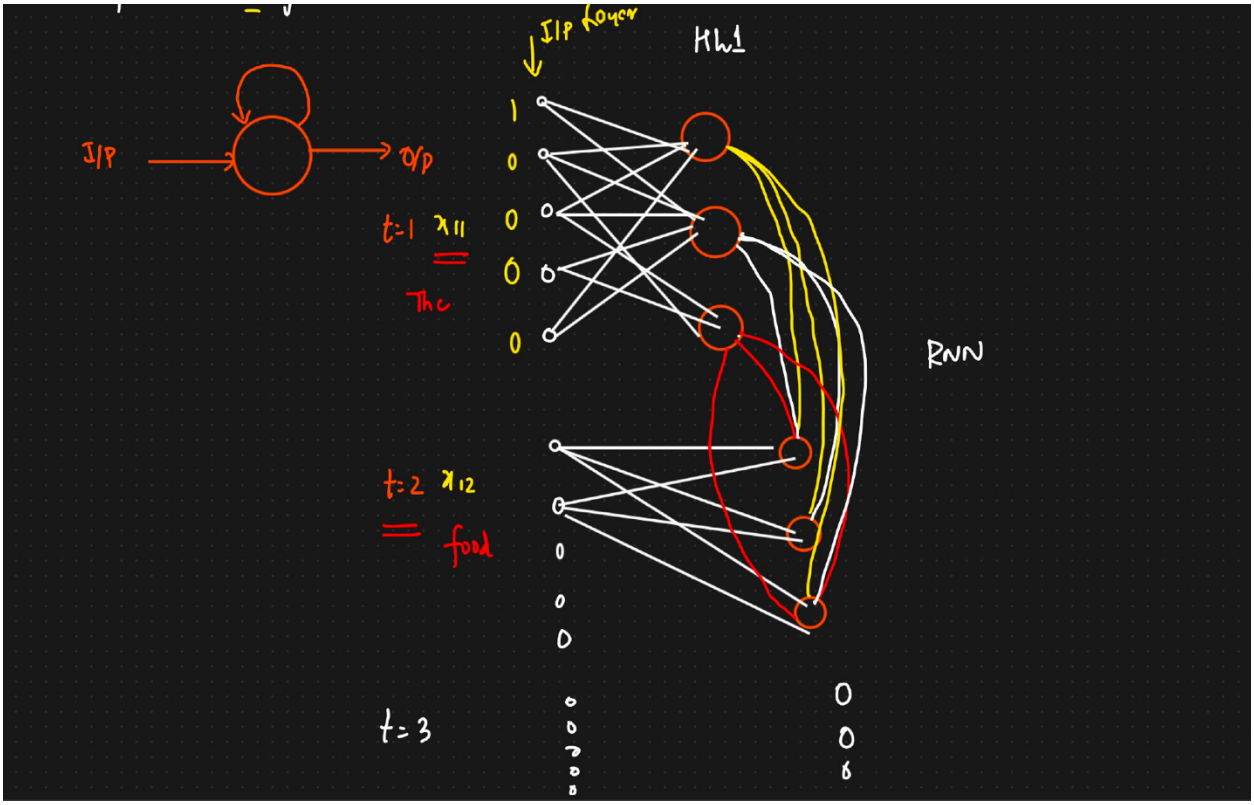

Simple RNN → Recurrent Neural N/W → Data → Sequential
ANN
HL
I/P
O/P
I/P
O/P
RNN
I/P
O/P

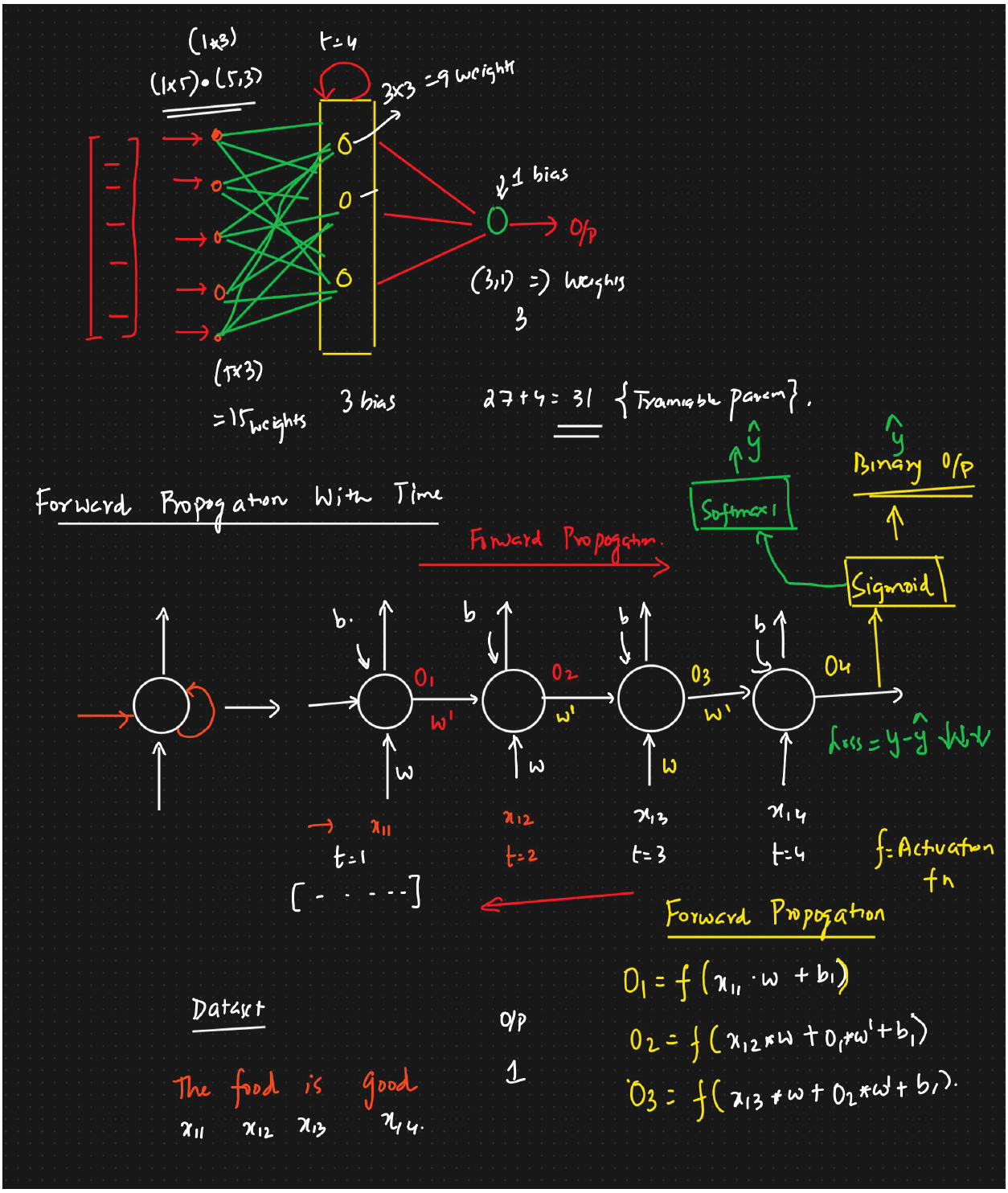

Sentence, Word
RNN
x_{11} x_{12} x_{13} x_{14}
The food is good
$t=1$ x_{11}
I/p
O/p
$t=2$ x_{12}
$t=3$ x_{13}
$t=4$ x_{14}


x_{11}
$t=1$
x_{12}
$t=2$
x_{13}
$t=3$
x_{14}
$t=4$

RNN Forward Propogation with Time

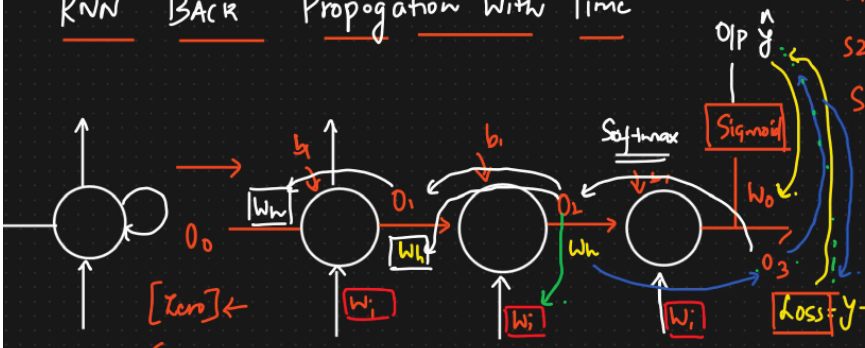






Simple RNN Backward Propogation

RNN BACK Propogation With Time



O/P

S1 $\langle x_{11} \quad x_{12} \quad x_{13} \rangle$ 1

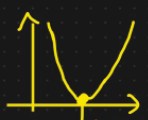
S2 $\langle x_{21} \quad x_{22} \quad x_{23} \rangle$ 0

S3 $\langle x_{31} \quad x_{32} \quad x_{33} \rangle$.

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

i = Sentence No.

When we update all the Weights $[W_i, W_h, W_o]$ ← Back Propogation



→ global Minima → loss value

Forward Propogation

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

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

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

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

Backward Propagation with Time

Update $[W_i, W_h, W_o]$

Weight updation formula

$$W_{new} = W_{old} - \eta \frac{\partial h}{\partial w_{old}}$$

$\frac{\partial h}{\partial w_{old}}$ → Derivative, slope of GRADIENT DESCENT

$\eta = 0.001$

$$W_{o_{new}} = W_{o_{old}} - \eta \frac{\partial h}{\partial w_{old}}$$

① Update W_o

Chain Rule of Derivative

$$\frac{\partial h}{\partial W_{oold}} = \frac{\partial h}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial W_{old}}$$

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

$$W_{h_{new}} = W_{h_{old}} - \eta \frac{\partial h}{\partial W_{hold}}$$

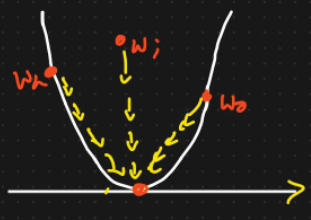
t=1, 2, 3

$$\frac{\partial h}{\partial w_{hold}} = \left[\frac{\partial h}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial o_3} * \frac{\partial o_3}{\partial w_h}\right]_{t=3} + \left[\frac{\partial h}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial o_3} * \frac{\partial o_3}{\partial o_2} * \frac{\partial o_2}{\partial w_h}\right]_{t=2} + \left[\frac{\partial h}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial o_3} * \frac{\partial o_3}{\partial o_2} * \frac{\partial o_2}{\partial o_1} * \frac{\partial o_1}{\partial w_h}\right]_{t=1}$$

③ Updating Weights w_i → Timestamp

$$w_{inew} = w_{iold} - \eta \boxed{\frac{\partial h}{\partial w_{iold}}} \quad \{\text{update } w_i \text{ weights}\}$$

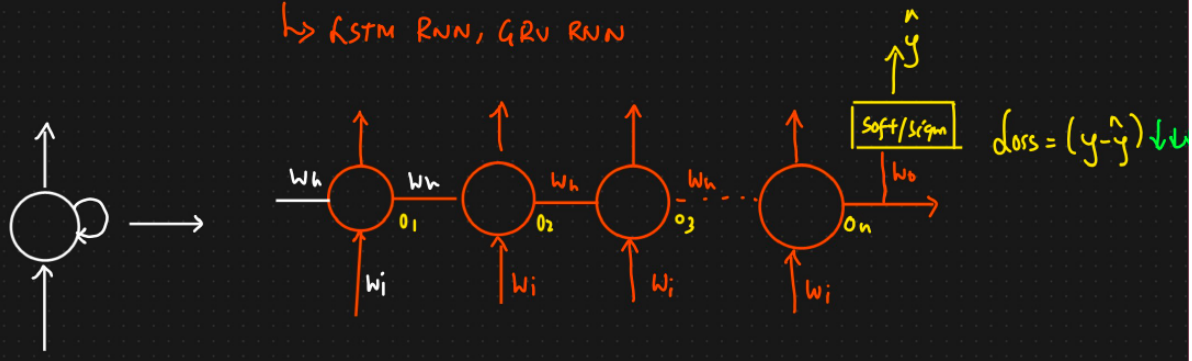
$$\boxed{\frac{\partial h}{\partial w_{iold}}} = \left[\frac{\partial h}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial o_3} * \frac{\partial o_3}{\partial w_{iold}}\right]_{t=3} + \left[\frac{\partial h}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial o_3} * \frac{\partial o_3}{\partial o_2} * \frac{\partial o_2}{\partial w_{iold}}\right]_{t=2} + \left[\frac{\partial h}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial o_3} * \frac{\partial o_3}{\partial o_2} * \frac{\partial o_2}{\partial o_1} * \frac{\partial o_1}{\partial w_{iold}}\right]_{t=1}$$



Problems with RNN

Problems With RNN

↳ LSTM RNN, GRU RNN



ANN → Vanishing Gradient Problem.

Text	O/P
The food is good	1
The food is bad	0

Sentences → small

Text Generation

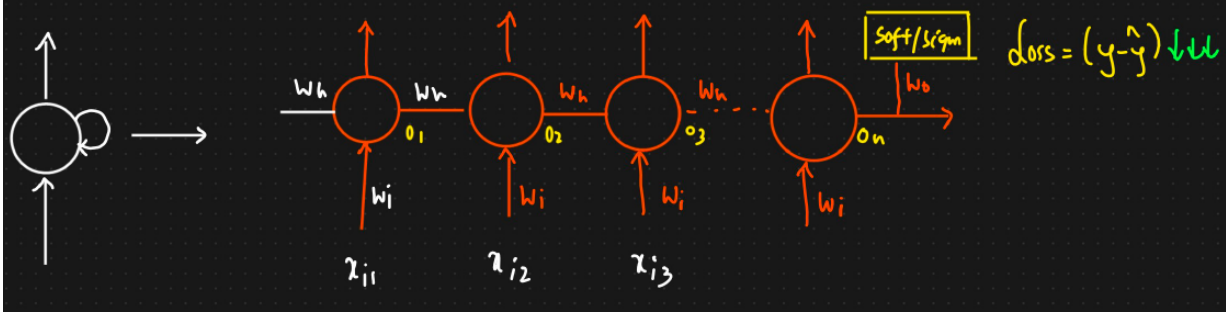
{ Sentence is . Long

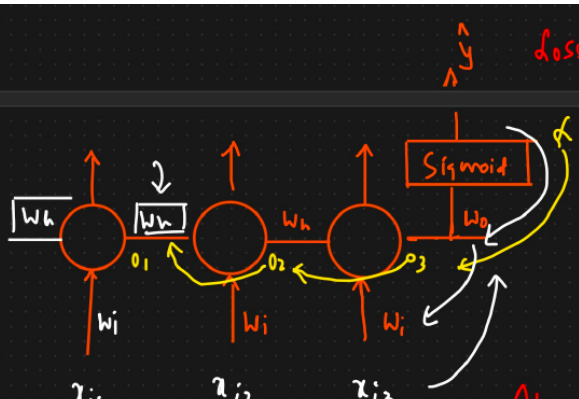
{ My Name is Malik and I like sports like CRICKET, VOLLEYBALL AND ALSO LIKE TO MAKE food. Videos

t=1 t=2 t=3 t=4 t=16

100 words

① long term Dependency cannot be captured by RNN → Provide Accuracy ↓ Dependency.





Updates W_h, W_i, W_o

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

$$\frac{\partial L}{\partial W_o} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial W_o} \Leftarrow \quad t=2$$

$t=1$

$$\frac{\partial L}{\partial W_{h_{old}}} = \left[\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial o_3} \cdot \frac{\partial o_3}{\partial o_2} \cdot \frac{\partial o_2}{\partial W_{h_{old}}}\right] + \ldots$$

Length of sentence ⇒ 50 words

$t=1$ $\qquad$ $t=2$

$$\frac{\partial L}{\partial W_{h_{old}}} = \left[\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial o_{50}} \cdot \frac{\partial o_{50}}{\partial o_{49}} \left[\frac{\partial o_{49}}{\partial o_{48}}\right] \cdots \left[\begin{matrix}0\\ \vdots \\ 0\end{matrix}\right] * \frac{\partial o_2}{\partial W_{h_{old}}}\right] + \ldots$$

$0.25 \times 0.02 \times 0.01 \times 0.015$

$o_3 = \sigma(x_{i3} * W_i + o_2 * W_h + b)$

$$\frac{\partial o_3}{\partial o_2} = \frac{\partial \sigma(x_{i3} * W_i + o_2 * W_h + b)}{\partial o_2} \Leftarrow o_3$$

Small Value $\simeq 0$

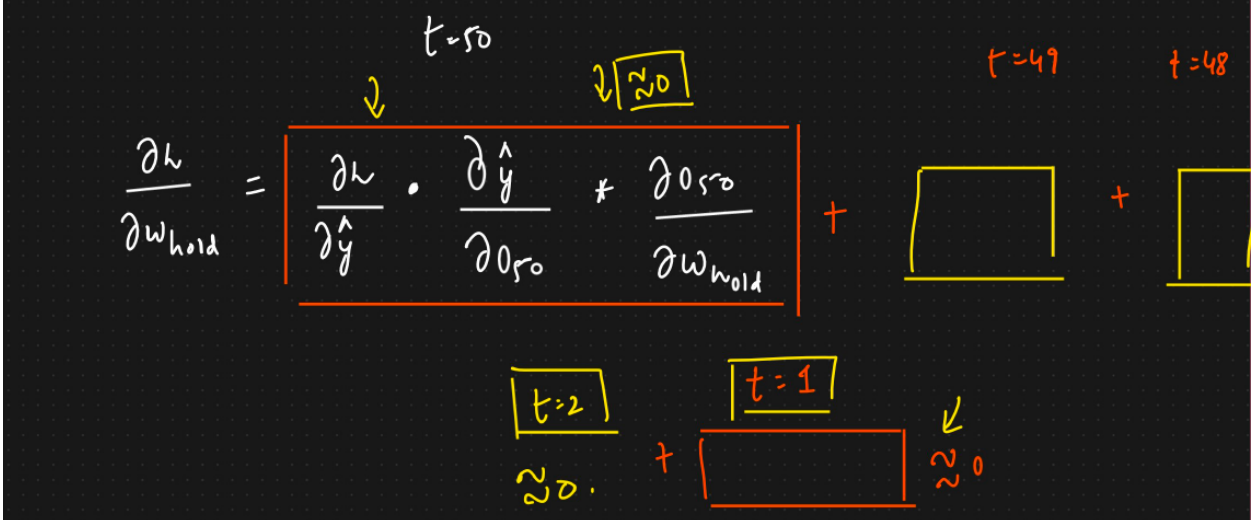
$= \sigma'(1 * W_h) \Rightarrow [0 - 0.25] * (W_h)$

Derivative of sigmoid $0 - 0.25$

Vanishing Gradient Problem

$t=1$ ⇒ The word is not participating to update the weights value.

Chain Rule is Big



① Relu, Leaky Relu →

② LSTM RNN → long short term Memory RNN } ⇒ Simple RNN.

③ GRU RNN →

Things we will be covering next class (ANN Project Implementation)