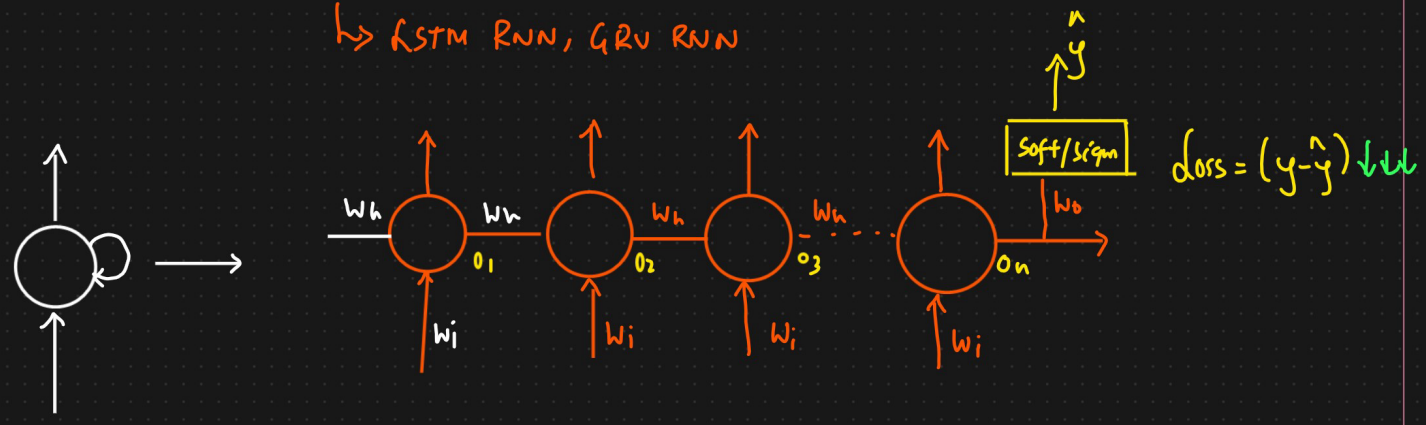
Problems With RNN

↳ LSTM RNN, GRU RNN

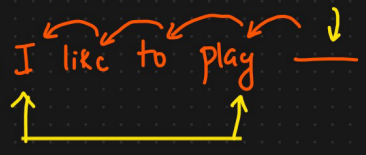


ANN → Vanishing Gradient Problem.

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

Sentences → small

Text Generation



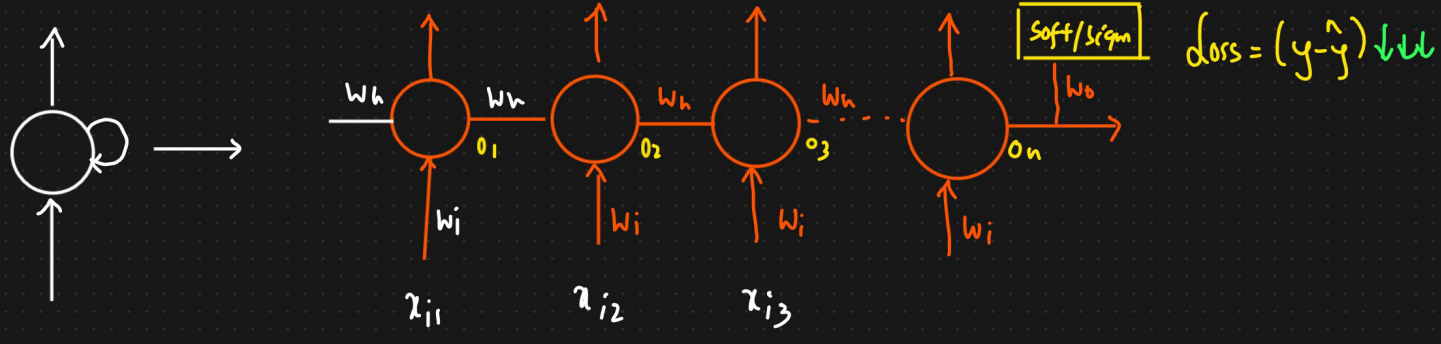
{ Sentence is . long

t:1 t=2 t=3 t=4 ... t=16

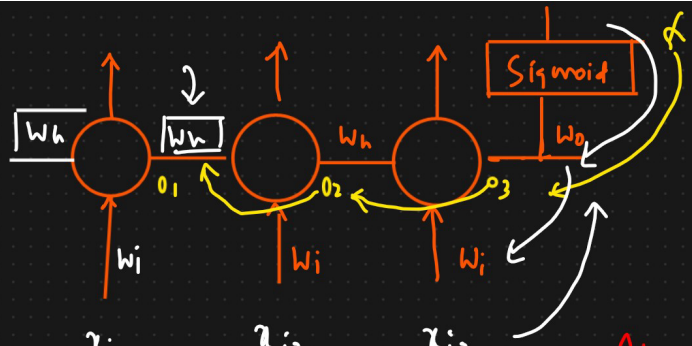
My Name is KRISH and I like sports like CRICKET, VOLLEYBALL AND ALSO LIKE TO MAKE food. Videos

100 words

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



ŷ loss=(y-ŷ)



Updates W_h, W_i, W_o

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

$$\frac{\partial L}{\partial W_o} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial W_o} \Leftarrow \begin{matrix} t=2 \end{matrix}$$

$t=1$

$$\frac{\partial L}{\partial W_{h_{old}}} = \left[\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial O_3} \cdot \frac{\partial O_3}{\partial O_2} \cdot \frac{\partial O_2}{\partial W_{h_{old}}} \right] + \cdots$$

Length of sentence $\Rightarrow$ 50 words

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

$$\frac{\partial L}{\partial W_{h_{old}}} = \left[\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial O_{50}} \cdot \frac{\partial O_{50}}{\partial O_{49}} \cdot \underbrace{\frac{\partial O_{49}}{\partial O_{48}} \cdots \frac{\partial O_2}{\partial W_{h_{old}}}}_{0.25 \times 0.02 \times 0.01 \times 0.015} \right] + \cdots$$

$$O_3 = \sigma(x_{i3} * W_i + O_2 * W_h + b)$$

$$\Downarrow$$

Small Value ≈ 0

$$\frac{\partial O_3}{\partial O_2} = \frac{\partial \sigma(x_{i3} * W_i + O_2^{1} * W_h + b) \Leftarrow O_3}{\partial O_2}$$

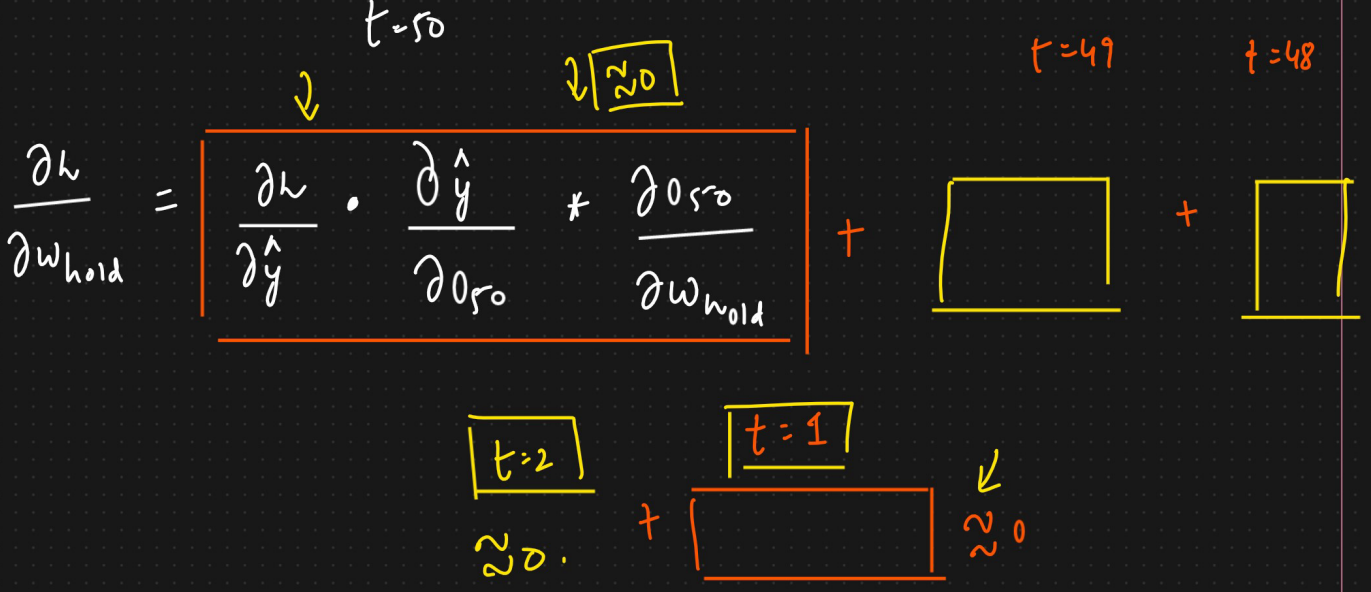
Derivative of sigmoid $0 - 0.25$

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

Vanishing Gradient Problem

$t=1 \Rightarrow$ 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 →

CONCUSION

Problems with RNN → Doesn't work well { Long Term Dependency.

- Less Accuracy for long sentences (i.e., where long term memory is required).

example $t = 50$

at $t_{50} \rightarrow t_{50} - t_0$
$t_{49} \rightarrow t_{49} - t_0$ } extremely lengthy & time consuming

Vanishing Gradient → occurs with deep neural networks (many layers) which results, overtime Gradient is 0

Act. fxns like sigmoid, tanh are more susceptible

To Solve

- we can use Relu, Leaky Relu
- LSTM RNN
- GRU RNN } Better options { Long Sentences