problem with RNN :-

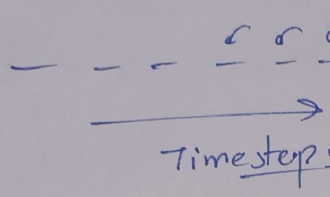
RNNs → Sequential data → Textual, time Series

↓

suffer (2) major problems → LSTMs why??

① → problem of long term dependencies
② → unstable gradients / stagnated Training. } → unstable gradients.

ent:



① Applications
↳ next word predicts -

→ Marathi is spoken in Mahara
short distance -

means :-
↳ RNN is Good working in short term distance like To make Kypad using RNN it is good in short term. but in long term they work poor.

→ Maratha is a beautiful place I went there last year but I could not enjoyed proper. because I don't understand marathi

long term } → if arises due to [Vanishing gradient problem]

② Stagnated Training / unstable Gradient :-

→ possible Solution

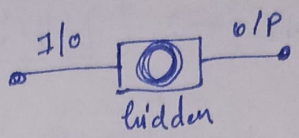
#

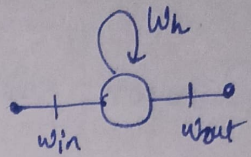
{ exploding gradient problem } → RNN Struggles -

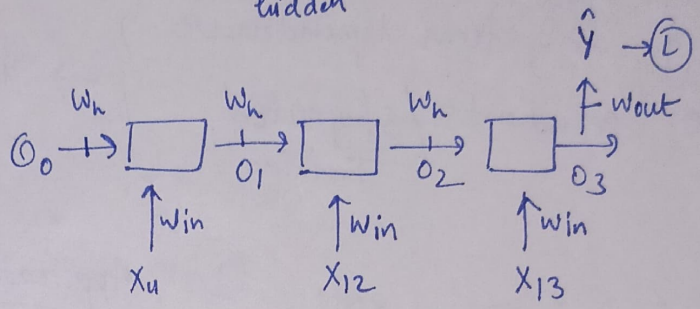
- problem #1 ⟹ problem of long term dependency
  - ↳ Vanishing Gradient problem

Input	output
1 0 1	1
0 0 1	0
0 0 0	0
1 1 1	1

3 time stamps







back propagation → L → min

W_h, W_{in}, W_{out}

↓

Gradient Descent

$$W_{in} = W_{in} - \eta \frac{\partial L}{\partial W_{in}}$$

$$W_{out} = W_{out} - \eta \frac{\partial L}{\partial W_{out}}$$

$$W_h = W_h - \eta \frac{\partial L}{\partial W_h}$$

Isko nikalne ka sbse bdaa contribution short term waalo ka hota hai

$$\frac{\partial L}{\partial min} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_3} \frac{\partial O_3}{\partial W_{in}} + \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_3} \frac{\partial O_3}{\partial O_2} \frac{\partial O_2}{\partial W_{in}} + \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_3} \frac{\partial O_3}{\partial O_2} \frac{\partial O_2}{\partial O_1} \frac{\partial O_1}{\partial W_{in}}$$

Short term dependencies.

long term dependencies

for 100 time stamps ÷

↳ $$\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_{100}} \frac{\partial O_{100}}{\partial O_{99}} \cdots \frac{\partial O_2}{\partial O_1} \frac{\partial O_1}{\partial W_{in}}$$

means ÷

Jitna jyada time stamps huga utna hi chote humara long term dependencies hugi !!

$$\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_{100}} \frac{\partial O_{100}}{\partial O_{99}} \cdots \frac{\partial O_2}{\partial O_1} \frac{\partial O_1}{\partial w_{in}} \rightarrow \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_{100}} \prod_{t=2}^{100} \left(\frac{\partial O_t}{\partial O_{t-1}} \right) \frac{\partial O_1}{\partial w_{in}}$$

Generalize

$$\frac{\partial O_2}{\partial O_1} \frac{\partial O_3}{\partial O_2} \frac{\partial O_4}{\partial O_3} \cdots \cdots \frac{\partial O_{100}}{\partial O_{99}}$$

Solution to Reduce this problem :-

1) Diff Activation → Relu / leaky relu.
2) Better weight initialization.
3) Skip rnns.
4) LSTM

⇒ #

⇒ problem # 2 → unstable Training (Exploding Gradients) :-

↳ Training thoda hi nhi hai.

1) Gradient Gripping
2) Control learning rate
3) LSTM