Problem with RNN

19 December 2022 16:33

RNNs → sequential data → textual, time series

↓

suffer 2 major problems → LTSMs (Why)

RNN

→ Problem of long term dependency
→ ~~Unstable gradients~~ stagnated training } → Unstable gradients

Applicat next word predicts

→ RNN

timesteps

→ Marathi is spoken in Mahar

short

→ Maha is a beautiful place. I went then last year, but I could not enjoyed prop. bca I dont understa marathi

→ RNN

{ Vanishing gradient prob } → long term

→ RNN → keypad

{ Exploding gradient problem } → RNN

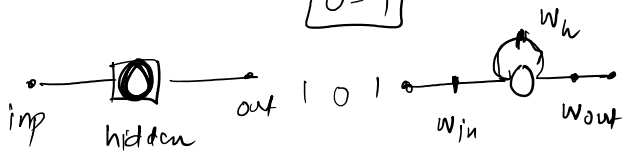
possible sol^n → remove

Problem # 1 → Problem of long term dependency → Vanishing

inp			out
1	0	1	1
0	0	1	0
0	0	0	0
1	1	1	1

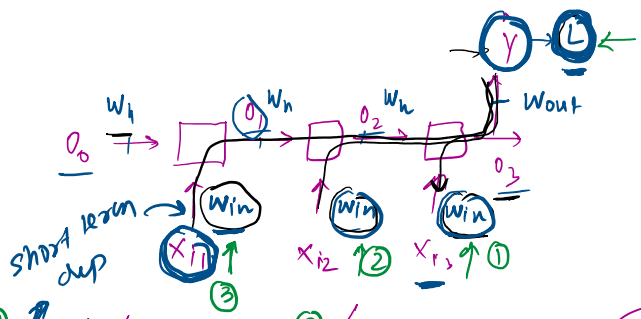
3 time stamps

0-1



L → min

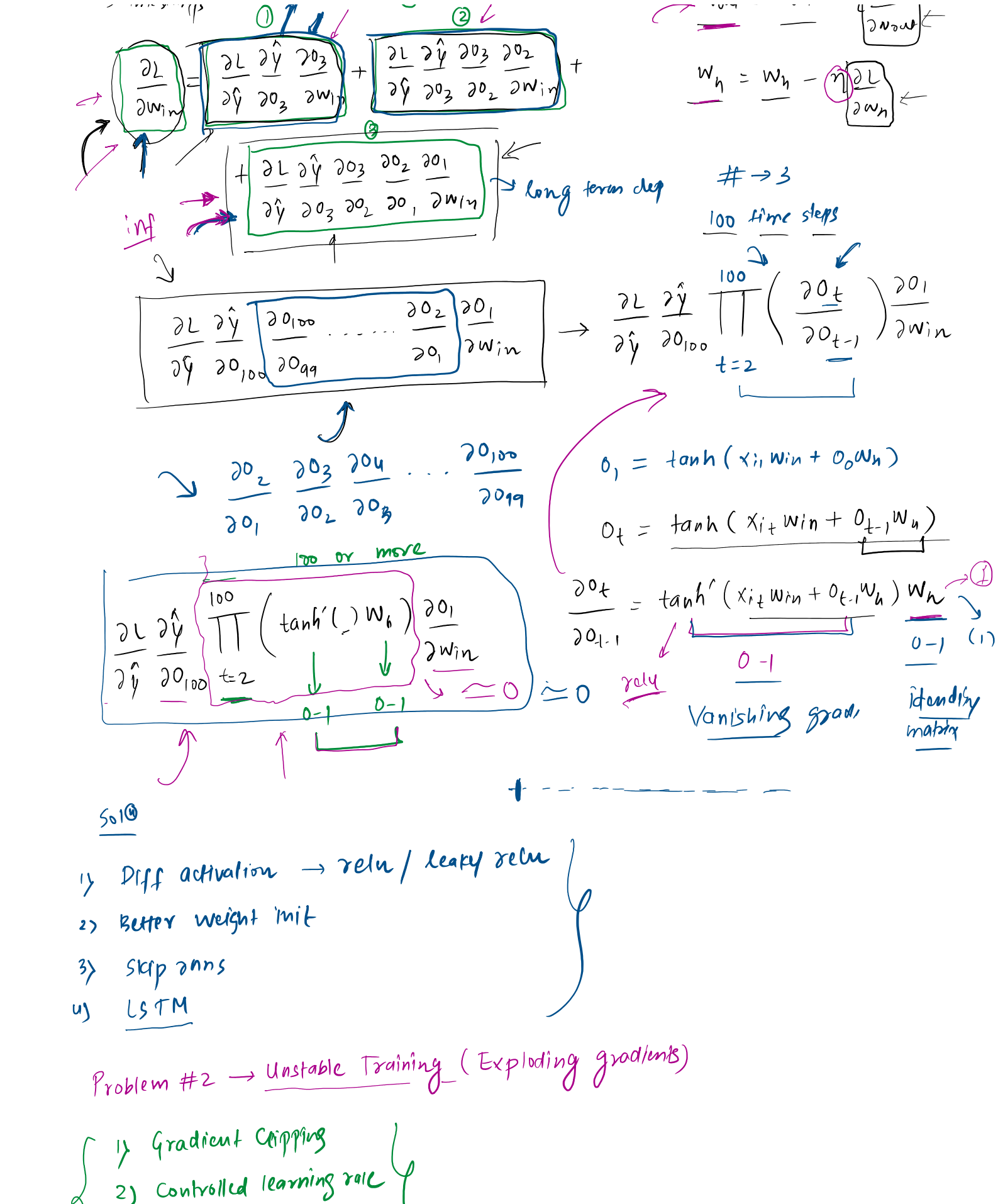
W_h, W_{in}, W_{out} → gradient des



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

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

$\partial L \quad \partial L \; \partial \hat{y} \; \partial O_3 \quad \partial L \; \partial \hat{y} \; \partial O_3 \; \partial O_2$



$$\frac{\partial L}{\partial w_{in}} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial o_3}\frac{\partial o_3}{\partial w_{in}} + \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial o_3}\frac{\partial o_3}{\partial o_2}\frac{\partial o_2}{\partial w_{in}} + \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial o_3}\frac{\partial o_3}{\partial o_2}\frac{\partial o_2}{\partial o_1}\frac{\partial o_1}{\partial w_{in}}$$

① ② ③

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

inf

long term dep

→ 3

100 time steps

$$\frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial o_{100}}\frac{\partial o_{100}}{\partial o_{99}}\cdots\frac{\partial o_2}{\partial o_1}\frac{\partial o_1}{\partial w_{in}} \rightarrow \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial o_{100}}\prod_{t=2}^{100}\left(\frac{\partial o_t}{\partial o_{t-1}}\right)\frac{\partial o_1}{\partial w_{in}}$$

$$\frac{\partial o_2}{\partial o_1}\frac{\partial o_3}{\partial o_2}\frac{\partial o_4}{\partial o_3}\cdots\frac{\partial o_{100}}{\partial o_{99}}$$

$$o_1 = \tanh(x_{i1} w_{in} + o_0 w_h)$$

$$o_t = \tanh(x_{it} w_{in} + o_{t-1} w_h)$$

100 or more

$$\frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial o_{100}}\prod_{t=2}^{100}\left(\tanh'(\ )\, w_h\right)\frac{\partial o_1}{\partial w_{in}}$$

0-1 0-1 $\simeq 0$ $\simeq 0$

$$\frac{\partial o_t}{\partial o_{t-1}} = \tanh'(x_{it} w_{in} + o_{t-1} w_h)\, w_h$$

relu

0-1 0-1 (1)

Vanishing grad.

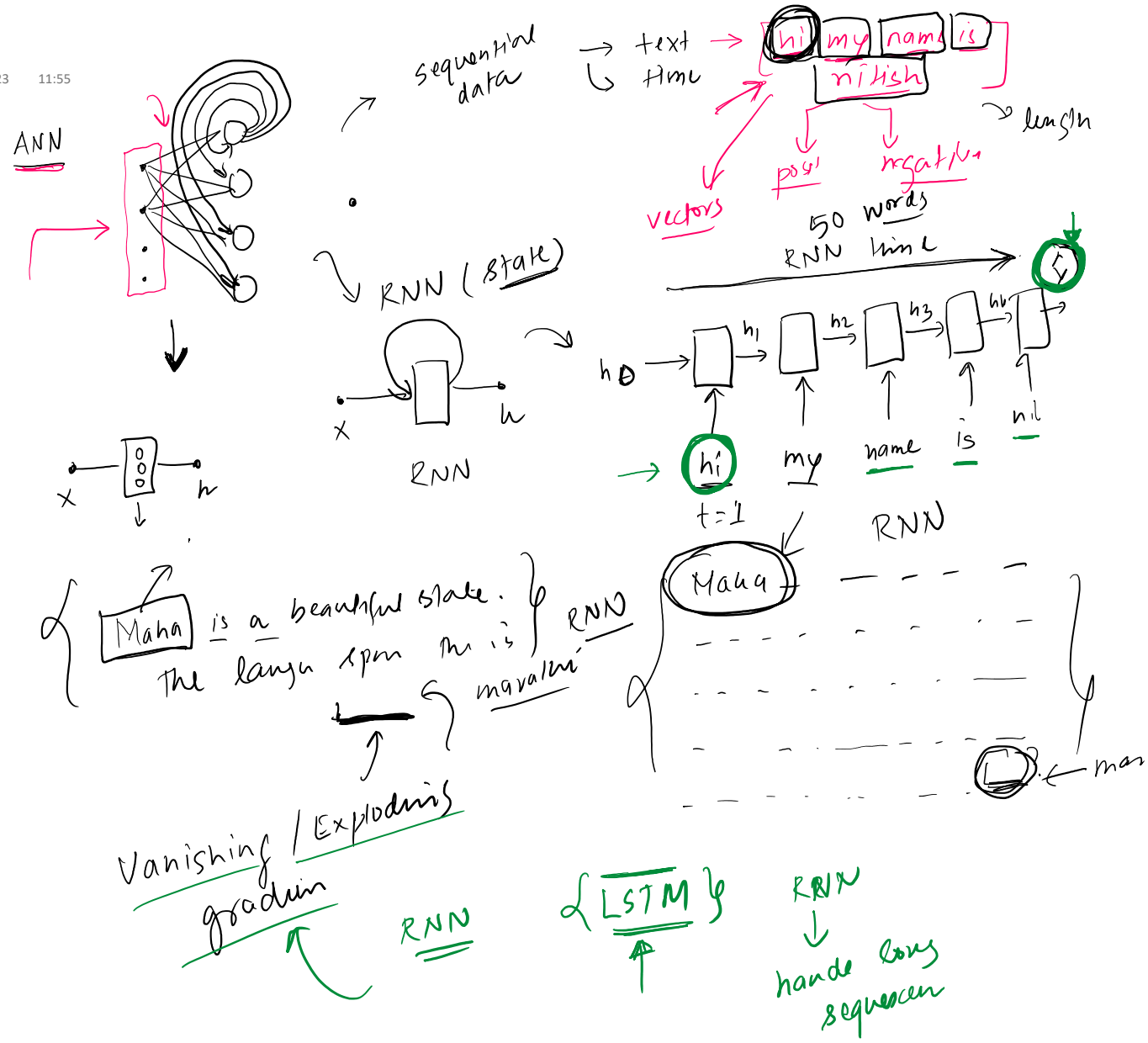
identity matrix

Sol^n

1) Diff activation → relu / leaky relu
2) Better weight init
3) Skip rnns
4) LSTM

Problem #2 → Unstable Training (Exploding gradients)

1) Gradient Clipping
2) Controlled learning rate
3) LSTM

Recap

21 August 2023 11:55