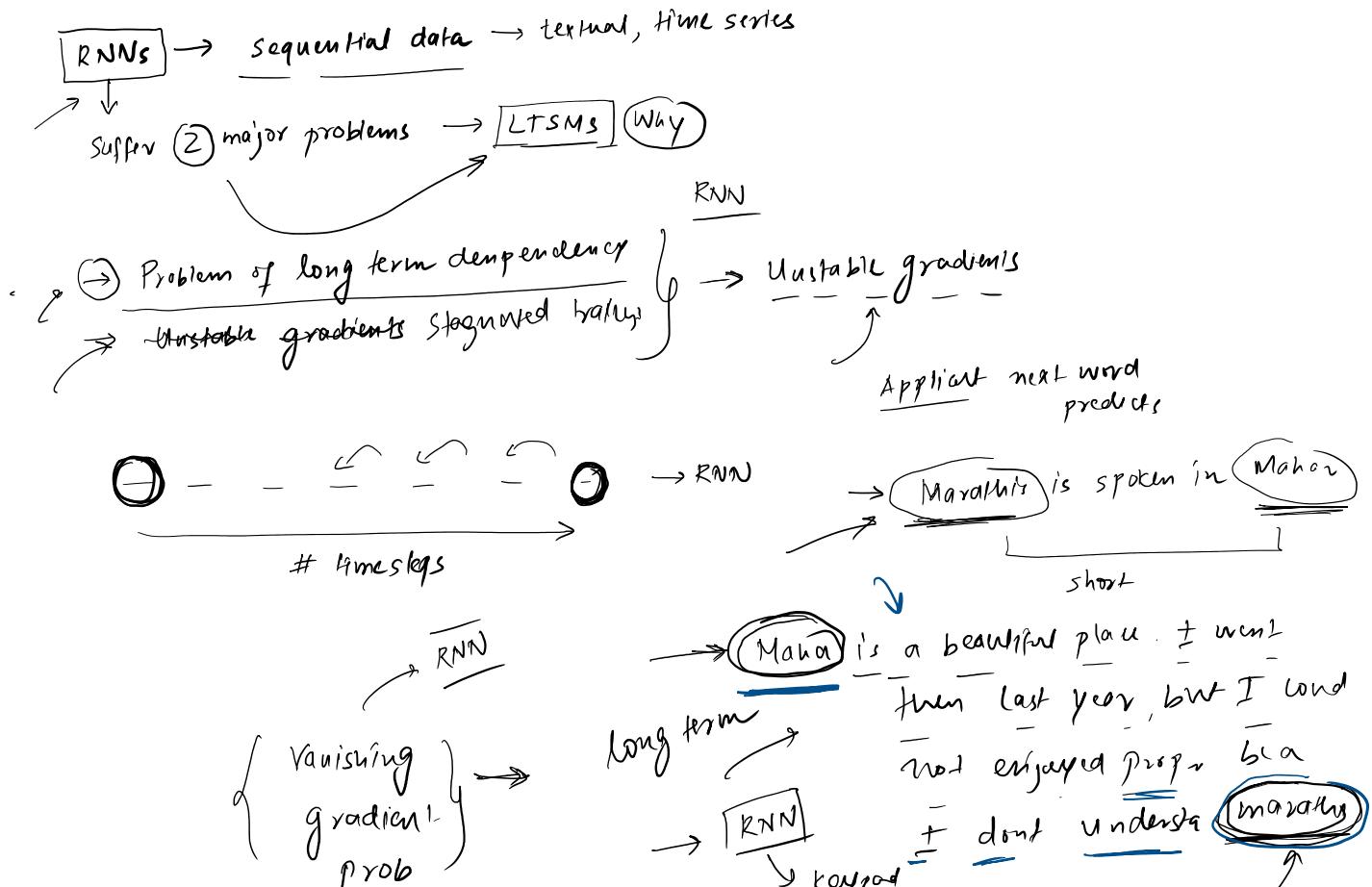
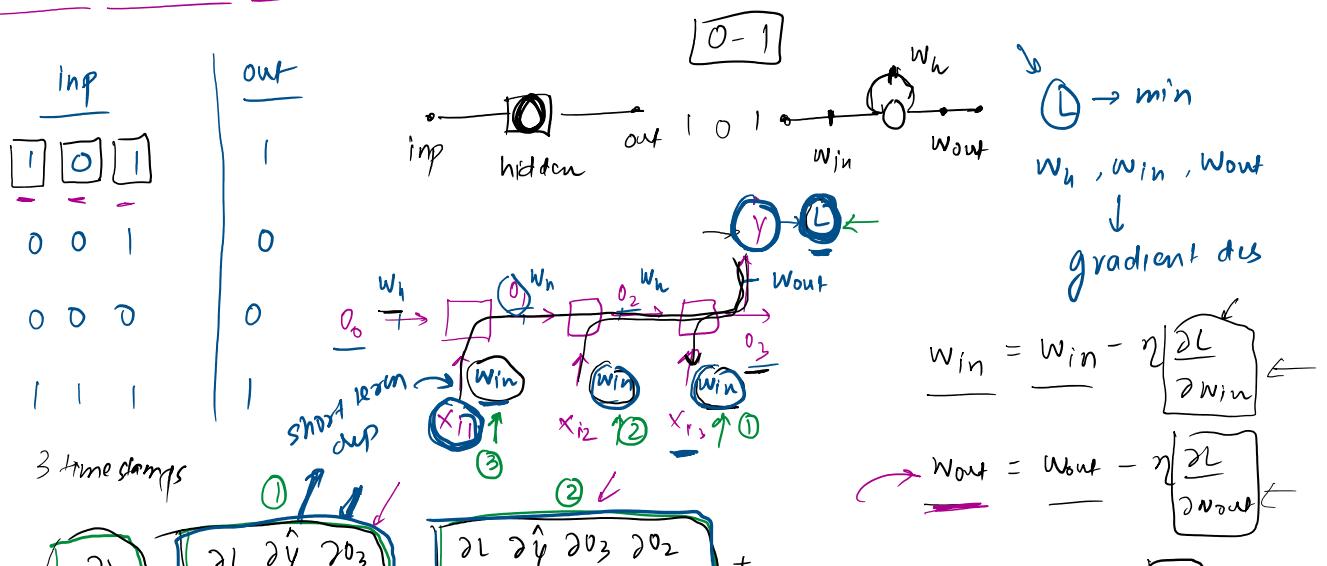


Problem with RNN

19 December 2022 16:33



Problem #1 → Problem of long term dependency → vanishing



$$\begin{aligned}
 & \text{①} \quad \frac{\partial L}{\partial w_{in}} = \frac{\partial L}{\partial y} \frac{\partial y}{\partial w_3} \frac{\partial w_3}{\partial w_{in}} + \frac{\partial L}{\partial y} \frac{\partial y}{\partial w_3} \frac{\partial w_3}{\partial w_2} \frac{\partial w_2}{\partial w_{in}} + \\
 & \quad \left(\frac{\partial L}{\partial y} \frac{\partial y}{\partial w_3} \frac{\partial w_3}{\partial w_2} \frac{\partial w_2}{\partial w_1} \frac{\partial w_1}{\partial w_{in}} \right) \\
 & \text{②} \quad \text{long term dep}
 \end{aligned}$$

$$\frac{\partial L}{\partial y} \quad \frac{\partial \hat{y}}{\partial y_{100}} \quad \boxed{\frac{\partial y_{100}}{\partial w_{99}}} \quad \dots \quad \frac{\partial y_2}{\partial y_1} \quad \frac{\partial y_1}{\partial w_{\text{min}}}$$

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

$$\rightarrow \frac{20_2}{20_1} \quad \frac{20_3}{20_2} \quad \frac{20_4}{20_3} \quad \dots \quad \frac{20_{100}}{20_{99}}$$

$$o_i = \tanh(x_i w_{in} + o_o w_h)$$

$$o_t = \tanh(x_i w_{in} + o_{t-1} w_u)$$

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

0 - 1

Vanishing grad.

Identify
matrix

5010

- 1) Diff activation \rightarrow relu / leaky relu
 - 2) Better weight init
 - 3) Skip conn
 - 4) LSTM

1. Long term dependency , as we seen in backpropagation while calculating gradients , there are so many terms .
If many as b/w 0-1 then it leads to problem of vanishing gradient descent.

2. To avoid vanishing gradient descent , we can activation functions that as non-saturating like relu , can do better weight initialization , can have skip rnns(we consider only few times , read about this)

Problem #2 → Unstable Training (Exploding gradients)

- { 1) Gradient Clipping
 - 2) Controlled learning rate
 - 3) LSTM }

1. There can be possibility of exploding gradients which lead to unstable training .

Recap

21 August 2023

11:55

