Problems with RNN

Dr. Muhammad Safyan

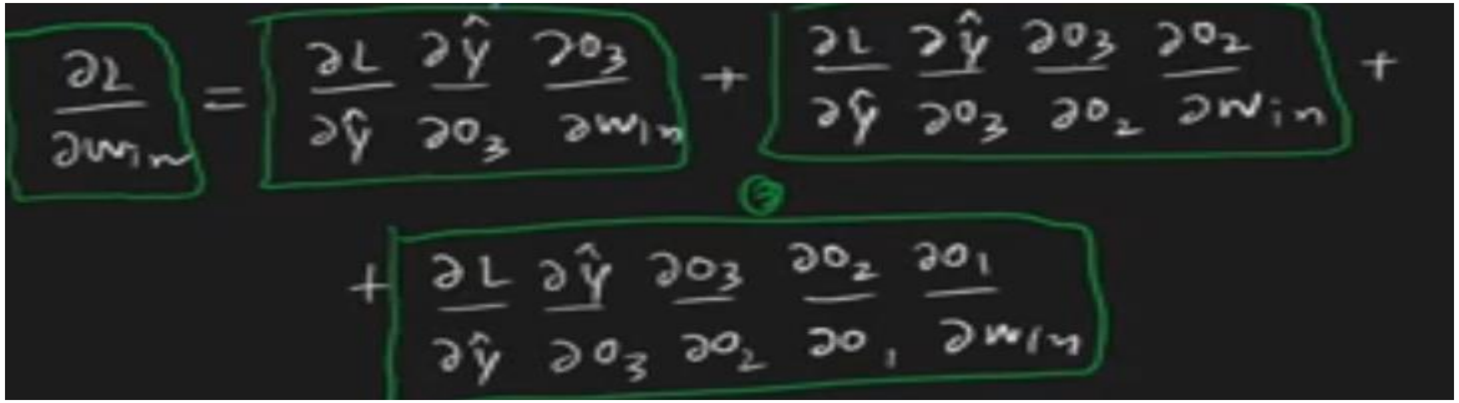
Problem with RNN

- Suitable for sequential data
  - Text , Time series
- Not used too much
  - Suffer with 2 major problems
    - Problem of long term dependency
    - Unstable gradient.
- Start to forget with the time step.
  - Next word prediction.
    - Punjabi is spoken in Punjab. Lahore is beautiful city. But I could not enjoy because I don't understand Punjabi.
      - Vanishing Gradient Descent

Unstable Training

- **Stagnant Training:**
- Exploding Gradient problem
- Longer term having so much large number , dominate the short term and become finite.
- e.g. relu +ve term derivative.
- Learning rate is not proper
    - Gradient Cliping
    - Control learning rate
    - LSTM

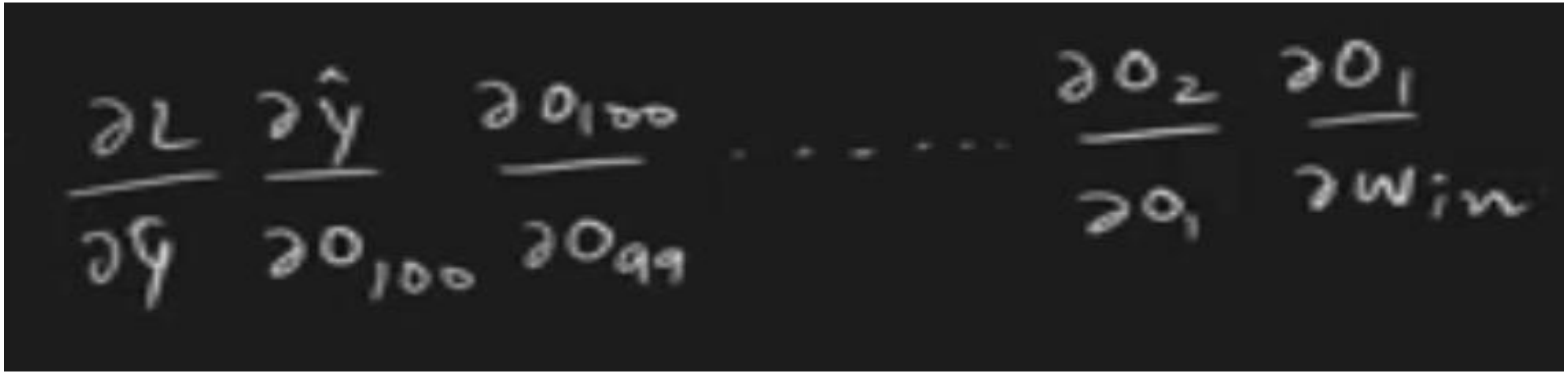
- RNN unfold input times
- Its length depends upon values in time steps(100 time steps)
- In Back propagation your tried to minimize the loss.
- Its done with Gradient Descent formula
- Wi, wh,wo



$$\boxed{\frac{\partial L}{\partial w_{in}}} = \boxed{\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_3} \frac{\partial O_3}{\partial w_{in}}} + \boxed{\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_3} \frac{\partial O_3}{\partial O_2} \frac{\partial O_2}{\partial w_{in}}} + \boxed{\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_3} \frac{\partial O_3}{\partial O_2} \frac{\partial O_2}{\partial O_1} \frac{\partial O_1}{\partial w_{in}}}$$

Long Term Dependency Problem

- In long sequence, Gradient Descent of short term Dependency contribute more then long term dependency.



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

How to reduce this problem

- Proper Activation function
  - Relu, Leaky Relu
- Better weight Initiation
- Skip gram
- LSTM