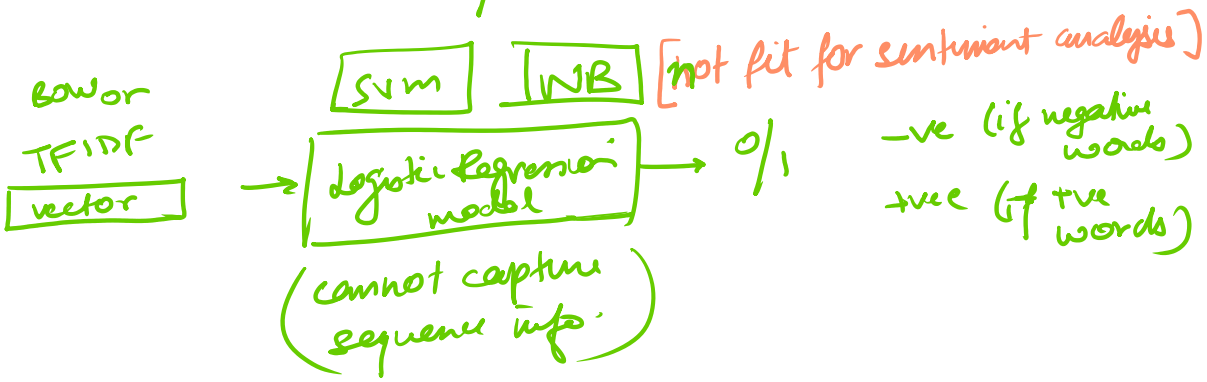
Transformers Vs LSTM

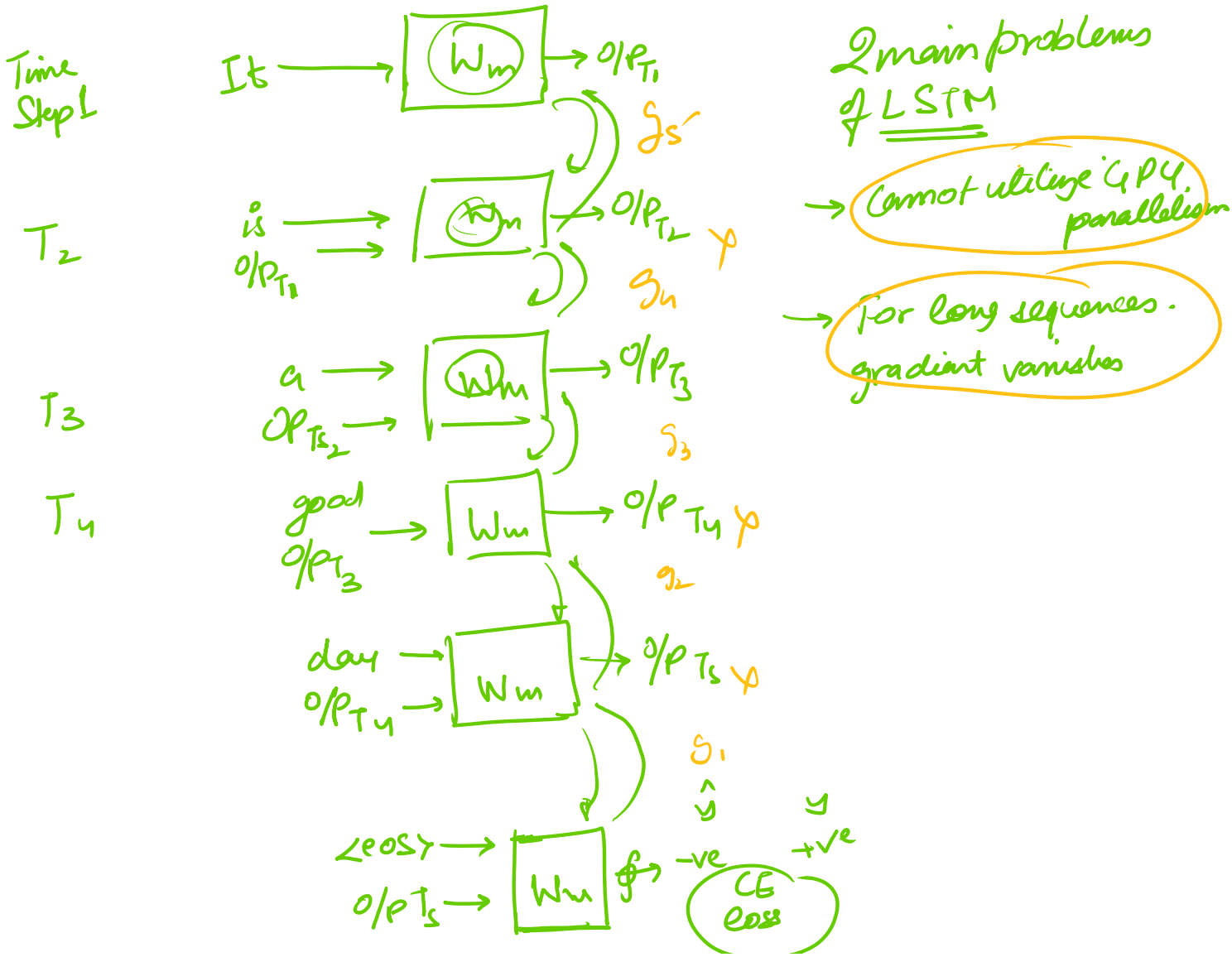
28 May 2022 23:16

Txt is a sequential data → T_1 T_2 T_3 T_4 Lets go there! → sequence matters

wrong order: go there lets !



LSTM can capture. Input → It is a good day [Sentiment analysis]

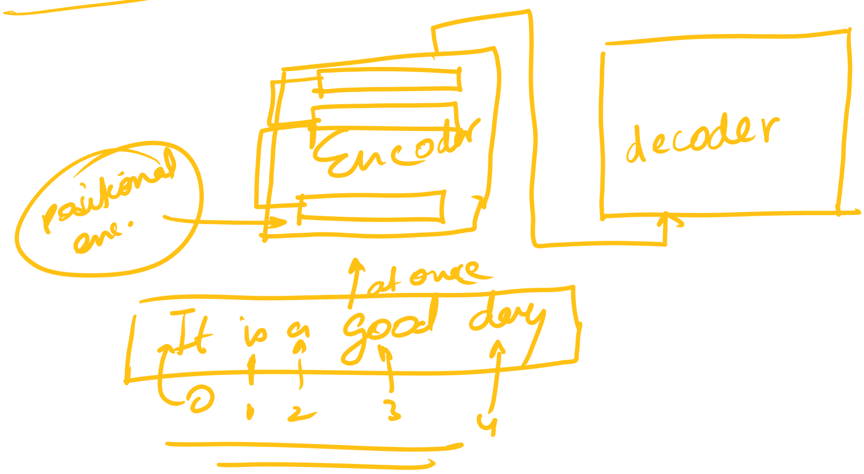


2 main problems of LSTM

→ Cannot utilize GPU parallelism

→ For long sequences. gradient vanishes

Transformers.



① Since i/p goes as one it can be parallelized.

② Residual connections helps with vanishing gradients

③ Attention improves the performance.

LSTM

→ Due to the recurrence it cannot be parallelized hence slow in training.

→ Suffers from vanishing gradient problem for long sequences.

→ No attention implementation

Transformers

→ Consumes the input at once hence can be parallelized
+ Position info is maintained by positional encoding layer.

→ Has residual connections to overcome vanishing gradient problem.

→ Attention layer improves the performance.