



# Vidyavardhini's College of Engineering and Technology

## Department of Artificial Intelligence & Data Science

AY: 2025-26

Class:	AI	Semester:	VII
Course Code:		Course Name:	DL

Name of Student:	BARI ANKIT VINOD
Roll No.:	61
Assignment No.:	5
Title of Assignment:	
Date of Submission:	
Date of Correction:	

### Evaluation

Performance Indicator	Max. Marks	Marks Obtained
Completeness	5	
Demonstrated Knowledge	3	
Legibility	2	
Total	10	

Performance Indicator	Exceed Expectations (EE)	Meet Expectations (ME)	Below Expectations (BE)
Completeness	5	3-4	1-2
Demonstrated Knowledge	3	2	1
Legibility	2	1	0

Checked By

Name of Faculty : Raunak Toshi

Signature :

Date :

# Assignment No. - 5

DL

Q. i) Discuss the way LSTM solve the vanishing gradient problem of RNN.

→ ① The vanishing gradient problem in RNN's:

- standard RNN's process sequences by successively applying the hidden problem:

$$h_t = \text{tanh}(W_h h_{t-1} + W_x x_t + b)$$

- during backpropagation through time (BPTT), the gradient involved states:

$$\frac{\partial L}{\partial h_t} = \frac{\partial L}{\partial h_t} \prod_{k=t}^{T-1} \frac{\partial h_{k+1}}{\partial h_k}$$

- Problem: Since  $|\tanh(x)| < 1$ , multiplying many small numbers.

ii) How LSTM solves it:-

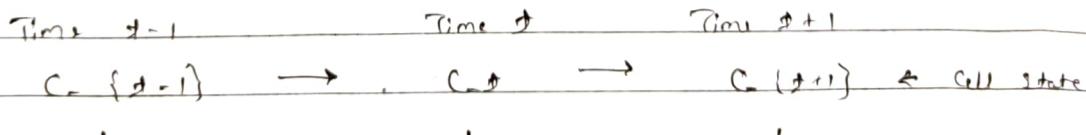
- LSTM (long - short - term - memory) introduces a special cell state  $C_t$  and gates that control information flow.

key components:-

- Cell state  $C_{t+1}$  - Acts as a highway for gradients.  
- gradient can unchanged over many time steps.

- Gates :-  
- forget gate is  
- input gate is  
- output gate is

iii) Visual Representation:-



forget & Input gates control updates in this layer



∴ The cell state acts like a "gradient highway", letting long - term information pass without decay.

Q. 2) Compare and contrast RNN and LSTM for Text Classification Task.

→

RNN

LSTM

- Simple recurrent unit with hidden state  $h_t$  | LSTM cell with hidden  $h_t$  and cell state  $c_t$ .

- Only short term memory | forget, input, output, gates.

- Prone - suffers from vanishing / exploding gradients | + emergent - gates + additive cell state preserve gradients.

- may fail when context spans long sequences. | can retain important contextual information across long sequences.

- less stable, may require gradient clipping, | more stable, better convergent in long sequences.

- dropout, or recurrent dropout required. | can use some plus intrinsic gate regularization.

- works well if sequences are short. | works well for; overfit for every short seqt.

- poor performance on long sequences | superior performance, remains constant from earlier words.

- limited context  $\rightarrow$  less accurate | + rich contextual representation  $\rightarrow$  better accuracy.

- faster, simpler

+ slower, heavier due to gates.