ourdiu.com | Shared by Abdullah Mahmud Tawsif (CSE-232)

Daffodil International University

Faculty of Science & Information Technology
Department of Computer Science and Engineering
Final Examination, Summer-2026
Course Code: CSE432, Course Title: Natural Language Processing
Level: 4 Term: 2 Batch: 64
Time: 2 Hours **Marks: 40**

Answer ALL Questions

[The figures in the right margin indicate the full marks and corresponding course outcomes. All portions of each question must be answered sequentially.]

1. Consider the fully-connected 2-2-2 network below: two inputs (i1, i2), one hidden layer (h1, h2) with a shared bias b1, and one output layer (o1, o2) with a shared bias b2. Hidden uses swish activation function and the output layer uses the sigmoid function and the network is trained with the sum-of-squared-errors loss. One forward/backward cycle is performed with learning rate η. The initial parameters are:

Parameter	Value	Parameter	Value	Parameter	Value
i1	0.05	w1	0.15	w6	0.45
i2	0.10	w2	0.20	w7	0.50
b1	0.35	w3	0.25	w8	0.55
b2	0.60	w4	0.30	target o1	0.01
lr (η)	0.5	w5	0.40	target o2	0.99

0.622 0.6229 Khalidur Rahman

CO2

a) Compute the net output of h1 and h2, then calculate o1 and o2. Hence compute E1, E2 and total error. [5]

b) Update w1 using chain rules showing all necessary steps. [10]

2. a) A feed-forward network can not model a sentence well. State two specific limitations of feed-forward network for sequential text that the RNN was designed to overcome and explain how the recurrent state addresses them. [5]

b) Explain types of RNN models with their applications. [5]

3. a) Analyze how the LSTM model overcome the memory limitation of RNN with figure. [5]

b) A Neural Machine Translation model produces the candidate translation below, evaluate it against a single human reference. [5] CO3

Candidate: the cat sat on mat
Reference: the cat sat on the mat
Compute the BLEU-2 score of the candidate (i.e. using modified n-gram precision for n=1 and n=2 with equal weights). Show all steps:

ourdiu.com | Shared by Abdullah Mahmud Tawsif (CSE-232)

	i)	The modified (clipped) unigram precision P_1 and bigram precision P_2.		
	ii)	The Brevity Penalty BP stating the formula you use and why it is needed.		
	iii)	The Final BLEU-2 score.		
	c)	Draw the encoder decoder architecture of a transformer. State its core formula for Scaled Dot-Product Attention and explain why it is scaled by $\sqrt{d_k}$.	[5]	