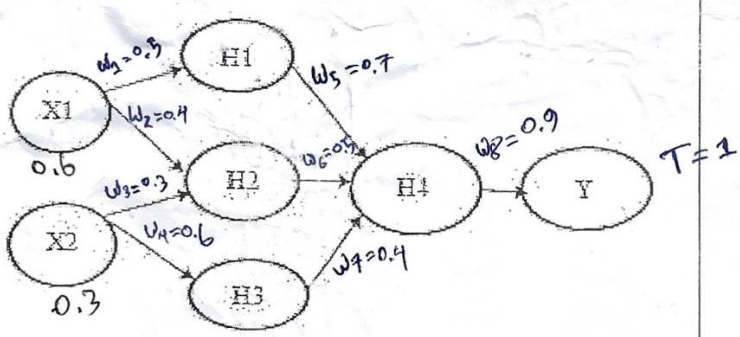
Daffodil International University

Faculty of Science & Information Technology
Department of Computer Science & Engineering
Final Examination, Fall 2025
Course Code: CSE 445, Course Title: Natural Language Processing
Level: 3 Term: 2 Batch: 63

Time: 02:00 Hrs **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.]

			[Marks]	
1.	*a)*	A company wants to build an AI model that analyzes **a sequence of daily customer interactions** to predict whether a customer will remain satisfied by the end of the week. Each day's interaction affects the next. Earlier negative or positive events continue to influence the customer's mood even after several days. The company initially used a simple neural network (treating each day independently), but it failed because it couldn't remember past interactions.		
		i. Identify **what type of model** should be used for such sequential, memory-dependent data. Explain the complete computational flow of a **sequence-processing neural architecture** that maintains hidden internal states across multiple time steps. Describe how information is **propagated through** time and how weight sharing enables consistent processing of sequential inputs.	5	CO3
		ii. For the same architecture, predict why earlier designs struggle with capturing **long-term dependencies**? Give a clear explanation of the gradient behavior that leads to this difficulty during training.	5	
2.		Given values for that network: X1 = 0.6 X2 = 0.3 X1H1 = W1 = 0.5 X1H2 = W2 = 0.4 X2H2 = W3 = 0.3 X2H3 = W4 = 0.6 H1H4 = W5 = 0.7 H2H4 = W6 = 0.5 H3H4 = W7 = 0.4 H4Y = W8 = 0.9 Target value = 1 $\eta = 0.1$ Sample size = 3  Fig. 01: Neural Network		CO4
	a)	Illustrate the network with weights and apply forward pass to the network.	5	
	b)	Calculate the total error and show how to improve the gap between predicted value and target value with sufficient proof.	10	

			[Marks]	
3.	*a)*	Student X is studying Teacher Y's behavior. Each day, Teacher Y either brings a drink (D) or no drink (N). The actual weather (Cold or Hot) is hidden. Student X builds the following Hidden Markov Model: **Initial state distribution:** $(Cold \quad Hot) = (0.45 \quad 0.55)$ **Row = current state** and **Column = next state**		

Table 1: Transition matrix

	Cold	Hot
Cold	0.70	0.30
Hot	0.25	0.75

Table 2: Emission matrix

	N	D
Cold	0.80	0.20
Hot	0.30	0.70

	i. Using the initial state, what is the probability that the first day was hot?	3
	ii. Using the emission matrix, what is the probability that Teacher Y shows up without a drink on a hot day, and the probability that Teacher Y shows up with a drink on a cold day, which of these two probabilities is higher?	3
	iii. For toal 11 observations [πCold $= 5/11$ and πHOT $= 5/11$], compute the probability of observing N on day 1 and transitioning from Cold on day 1 to Hot on day 2.	4
b)	i. Discuss LLMs by describing their definition, core architectural components with a neat diagram, and discuss the roles of attention mechanisms, and give two clear examples to support your explanation.	2.5
	ii. What is self-attention in a Transformer model, and why is it important for understanding relationships between words in a sentence?	2.5

CO5