Daffodil International University

Faculty of Science & Information Technology
Department of Computer Science and Engineering
Midterm Examination, Summer-2026
Course Code: CSE432, Course Title: Natural Language Processing
Level: 4 Term: 2 Batch: 64
Time: 1.5 Hours **Marks: 25**

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.	a)	Stemming and lemmatization are two preprocessing techniques used in NLP. Discuss their advantages and limitations with proper example.	[3]	
	b)	Consider the following two documents: D1: Chelsea Beats Barcelona but lost to Madrid. D2: Barcelona Beats Madrid but lost to Chelsea.		
	i)	BoW (Using unigram) model is used for finding similarity between documents. Is it possible to apply here? If possible then find similarity to these documents.	[3]	**CO1**
	ii)	Does this approach reflect the context of the sentence? If not, propose a model that can help to overcome the limitations of previous model with proof.	[4]	
2.		A Naïve Bayes classifier is developed to classify short text descriptions into two classes: **Food** and **Technology.** The training dataset is shown below: Test string: **"Blue Phone"**		**CO2**
	a)	Calculate prior probabilities and likelihood for the given problem.	[2]	
	b)	Determine the predicted class for the document using Naïve Bayes.	[3]	
3.		Consider the following training dataset: The test sentence is: **"Furry, Sweet".**		

Description	Class
Sweet Juicy Apple	Food
Fresh Banana Juice	Food
Smart handy Phone	Technology
Fast Computer	Technology

Description	Class
Furry, Tail	Animal
Engine, Wheels	Vehicle
Sweet, Juicy	Fruit

Consider the weight and bias matrix for the hidden layer:

$$W^{(1)} = \begin{bmatrix} 0.5 & 0.4 & -0.3 & -0.2 & 0.6 & 0.1 \\ 0.2 & 0.3 & 0.5 & 0.4 & -0.1 & 0.2 \end{bmatrix}$$

$$b^{(1)} = \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix}$$

Consider the following weight and bias matrix for output layer.

$$W^{(2)} = \begin{bmatrix} 0.6 & 0.2 \\ 0.3 & 0.7 \\ 0.5 & 0.4 \end{bmatrix} \quad b^{(2)} = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.2 \end{bmatrix}$$

The actual class of the test sample is:

$$y = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

	a)	Draw the architecture of the ANN network that can be used to solve the given problem highlighting minimum input nodes, hidden layers, nodes in the hidden layers and output layers with necessary nodes. Complete the necessary tasks prior to the designing.	[2]	CO2
	b)	Calculate the values of the nodes of the hidden layer using sigmoid function.	[3]	
	c)	Compute the final output probabilities using the Softmax function and predict the class.	[3]	
	d)	Compute the Cross-Entropy Loss of the prediction.	[2]	