(66)

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -2 EXAMINATIONS- 2026

M.Tech-II Semester (Data Science)

COURSE CODE (CREDITS): 22M1WCI236 (3) MAX MARKS: 25

COURSE NAME: Natural Language Processing

COURSE INSTRUCTOR: HRI MAX. TIME: 1 Hour 30 Min

Note: (a) All questions are compulsory.

(b) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

(c) Use of calculator is allowed

Q.No	Question	CO	Marks
Q1	**A stock market analyst models the market behavior as having two hidden states: Bull Market (B) and Bear Market (BE). We cannot observe the state directly, but we can observe the daily trade volume, which can be High (H), Medium (M), or Low (L). Let us start on Day 1 in a Bull Market state ($q_1 = B$). The HMM parameters are given below:**	CO2	[4]

Transition Probabilities (a_{ij})

	Bull (B)	Bear (BE)
Bull (B)	0.7	0.3
Bear (BE)	0.4	0.6

Emission Probabilities ($b_j(k)$)

	High (H)	Medium (M)	Low (L)
Bull (B)	0.6	0.3	0.1
Bear (BE)	0.1	0.3	0.6

What is the probability that the observation on Day 2 is "Low" ($O_2 = L$)?

Q.No	Question	CO	Marks
Q2	**Consider the PCFG G':**	CO2	[4]

Phrasal Rules		Lexical Rules	
$S \rightarrow$ NP VP	[0.6]	N $\rightarrow$ *fish*	[0.5]
$S \rightarrow$ VP NP	[0.4]	N $\rightarrow$ *runs*	[0.5]
NP $\rightarrow$ N	[1.0]	V $\rightarrow$ *fish*	[0.5]
VP $\rightarrow$ V	[1.0]	V $\rightarrow$ *runs*	[0.5]

For the sentence "fish runs", how many distinct parse trees exist and what is $P(\text{"fish runs"} \mid G')$?

Q.No	Question	CO	Marks
Q3	Suppose you are training MST Parser for dependency and the sentence, "I like online class" occurs in the training set. The POS tags for these words are Pronoun, Verb, PropNoun and Noun, respectively. Also, for simplicity, assume that there is only one dependency relation, "rel". Thus, for every arc from word w_x to w_y, your features may be simplified to depend only on words w_x and w_y and not on the relation label. Below is the set of features –	CO2	[5]

<table>
<tr><td></td><td>f1: pos(w_x) – Verb and pos(w_y) = Noun | Pronoun
f2: w_x = Root | w_x occurs before w_y in the sentence
f3: w_x = Root and pos(w_y) = Verb
f4: w_y occurs before w_x in the sentence
The feature weights before the start of the iteration are: [5,20,15,12]
Suppose you are also given that after applying the Chu-Liu-Edmonds, you ger the following parse tree { Root ->like, like ->I, I -> online, online -> class}
What would be weights after this iteration?</td><td></td><td></td></tr>
<tr><td>Q4</td><td>Dataset: An automobile is a wheeled motor vehicle used for transporting passengers. A car is a form of transport, usually with four wheels and the capacity to carry around five passengers. Transport for the London games is limited, with spectators strongly advised to avoid the use of cars. The London 2012 soccer tournament began yesterday, with plenty of goals in the opening matches. Giggs scored the first goal of the football tournament at Wembley, North London. Bellamy was largely a passenger in the football match, playing no part in either goal
Target words: <automobile, car, soccer, football>
Term vocabulary: <wheel, transport, passenger, tournament, London, goal, match>
Design a distributional matrix and compute similarity using vector product on the co-occurrence matrix. Which target words are similar?</td><td>CO3</td><td>[5]</td></tr>
<tr><td>Q5</td><td>The distribution of two words A & B is given in 8-dimensional space as
A= [1,0,1,1,0,1,0,1] , B=[0,1,1,1,1,1,1,0]
Compute (a) Dice index (A,B) and (b) Jaccard index (A,B)</td><td>CO3</td><td>[2]</td></tr>
<tr><td>Q6</td><td>A word2vec model has vocabulary $\{w_1, w_2, w_3, w_4\}$ ($V = 4$) and embedding dimension $d = 2$. The two weight matrices are:
$$W_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \\ 0 & 2 \end{pmatrix}, \quad W_2 = \begin{pmatrix} 1 & 1 & 0 & -1 \\ 0 & 2 & 1 & 1 \end{pmatrix}$$
Given the text fragment "… w_3 w_2 w_4 …", where w_2 is the center word and $\{w_3, w_4\}$ is the context
Compute the hidden layer h and the output score vector s for both models CBOW and Skip-gram.</td><td>CO3</td><td>[5]</td></tr>
</table>