JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

(39)

TEST -1 EXAMINATION- 2025

B.Tech-VII Semester (CSE/IT)

COURSE CODE (CREDITS): 19B1WCI731 (2) MAX. MARKS: 15

COURSE NAME: Computational Data Analysis

COURSE INSTRUCTORS: Ekta Gandotra MAX. TIME: 1 Hour

Note: (a) All questions are compulsory.
(b) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems
(c) Calculator is allowed.

Q. No.	Question	CO	Marks
Q1.	Differentiate between generative and discriminative models in machine learning with respect to (i) the probability distribution they use, (ii) use cases, and (iii) examples of algorithms that fall under each category.	1	3
Q2.	Show that applying Maximum Likelihood Estimation, under the assumption that the error terms in linear regression model follow a normal distribution, leads to the Ordinary Least Squares objective.	1	4
Q3.	Consider the following data on the number of hours of study (x1) and the number of online tutorials watched (x2) by students, along with their exam scores (y). Using this data, find the regression equation by applying the normal equation.	2	4
Q4.	In the training data of a Naïve Bayes text classifier, the word "discount" appears 5 times in Spam emails and 1 time in Ham emails, while the word "meeting" appears 2 times in Spam emails and 8 times in Ham emails. The total word counts are N_{Spam}=200, N_{Ham}=250, and the vocabulary size is V=1000. Using Laplace smoothing (α=1), answer the following: a. Compute P(discount|Spam) and (discount|Ham). b. Compute P(meeting|Spam) and P(meeting|Ham). c. If the prior probabilities are P(Spam)=0.4 and P(Ham)=0.6, classify a test document containing the words "discount meeting".	2	4

Hours Studied (x1)	Tutorials Watched (x2)	Exam Score (y)
2	1	65
3	2	70
5	3	80
7	3	85
9	4	95