

18 Ektg

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -2 EXAMINATION- April, 2022

B.Tech. (CSE, IT) VI Semester

COURSE CODE: 18B1WCI634

MAX. MARKS: 25

COURSE NAME: Machine Learning

COURSE CREDITS: 2

MAX. TIME: 1.5 Hrs.

Note: All questions are compulsory. Carrying of mobile phone during examinations will be treated as case of unfair means.

- Q1. CO-3 a. Using the following dataset, predict the class for the record (Color=Green, legs=2, Height=Tall and Smelly= No) using Naïve Bayes algorithm. [3]

Instance	Color	Legs	Height	Smelly	Species
1	White	3	Short	Yes	M
2	Green	2	Tall	No	M
3	Green	3	Short	Yes	M
4	White	3	Short	Yes	M
5	Green	2	Short	No	H
6	White	2	Tall	No	H
7	White	2	Tall	No	H
8	White	2	Short	yes	H

- b. Give any two drawbacks of Naïve Bayes algorithm. [2]

- Q2. CO-2 a. Explain overfitting and underfitting in terms of bias and variance errors. [3]

- b. Suppose you build a model which shows a training accuracy of 95% and a test accuracy of 62%. What could be the possible reasons for the gap between these accuracies? [2]

- Q3. CO-4 a. Differentiate between ID3, C4.5 and CART algorithms. [2]

- b. Consider the following dataset of training examples: [3]

A	B	Class Label
T	T	C0
T	T	C0
T	F	C1
F	F	C0
F	T	C1
F	T	C1

Compute Gini index for both the attributes (Show all steps clearly). Which attribute should be selected as a root node for building a decision tree on the basis of Gini index?

P.T.O.

- Q4. CO-4 a. How does a Support Vector Machine work? Explain with the help of an example. [3]
- b. Suppose 10,000 patients get tested for flu; out of them, 9000 are actually healthy and 1000 are actually sick. For the sick people, a test was positive for 620 and negative for 380. For the healthy people, the same test was positive for 180 and negative for 8,820. Construct a confusion matrix for the data and compute the accuracy. [2]
- Q5. CO-5 a. Give a method to find the optimal value of k for k-NN algorithm. [2]
- b. Given the following dataset, find the class of the test sample using k-NN algorithm. Take k=1 and k=3. Use L2 Norm for distance computations. [3]

Height (CM)	Weight (KG)	Class
167	51	Underweight
173	64	Normal
172	65	Normal
174	56	Underweight
169	58	Normal
168	53	Underweight

Test Sample.

Height (CM)	Weight (KG)	Class
170	57	?

12 Examination