

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT
TEST-3 EXAMINATION-2022

B.Tech-III Semester (CS/IT/ECE/Civil/BT)

COURSE CODE (CREDITS): 19B1WCI740 (3)

MAX. MARKS: 35

COURSE NAME: An Introduction to Statistical Learning

COURSE INSTRUCTORS: Dr. Hari Singh

MAX. TIME: 2 Hours

Note: All questions are compulsory. Marks are indicated against each question in square brackets.

Q1. Compute R^2 for the following observations? Write all the steps involved.

[CO1](4)

Actual observed (y)	Predicted (y')	Irreducible error (ϵ)
2	3	0.2
4	3	0.4
5	6	0.3
7	8	0.5
9	9	0.3

Q2. For a certain plant come from three different cities A, B and C where 30% are from A, 25% are from B and 45% are from C. Out of those 2% from A, 1% from B and 3% from C are found COVID positive. Find the probability that a worker chosen at random would be tested COVID positive and also find the probability that if a worker is tested COVID positive, what is the probability that he is coming from A or B or C?

[CO2](4)

Q3. (a) Describe the form of a piecewise cubic polynomial with two knots in terms of a mathematical equation.

[CO3](4)

(b) Can any problem be seen while fitting to any dataset in the context of Q3(a)?

[CO3](2)

Q4. Describe a smoothing spline that makes RSS small and is also smooth. Elaborate each term clearly in the mathematical equation representing it. What is the impact of the tuning parameter λ ?

[CO3] (4)

Q5. (a) Compute first, second and third round of residuals in the Gradient Boost Tree for the following dataset, where Weight is the output. Consider the following splitting rule: Gender=Female at the root node; Height < 1.6 at the left root child when Gender=Female is true; Colour is not blue at the right root child when Gender=Female is false.

[CO3] (6)

(b) Make a prediction for the weight(Kg), when Height=1.7, Colour=Green and Gender=Female. [CO3] (3)

Height (m)	Favorite Colour	Gender	Weight (Kg)
1.6	Blue	Male	88
1.6	Green	Female	76
1.5	Blue	Female	56
1.8	Red	Male	73
1.5	Green	Male	77
1.4	Blue	Female	57

Q6. Here we explore the maximal margin classifier on a dataset. Justify your answer with appropriate explanation. [CO3] (2x4=8)

- We are given $n=7$ observations in $p=2$ dimensions. For each observation, there is an associated class label. Sketch the observations.
- Sketch the optimal separating hyperplane, and provide the equation for this hyperplane.
- Describe the classification rule for the maximal margin classifier. It should be something along the lines of "Classify to Red if $\beta_0 + \beta_1 X_1 + \beta_2 X_2 > 0$, and classify to Blue otherwise." Provide the values for β_0 , β_1 , and β_2 .
- On your sketch, indicate the margin for the maximal margin hyperplane.

Observations	X1	X2	Y
1	3	4	Red
2	2	2	Blue
3	4	4	Red
4	1	4	Red
5	2	1	Blue
6	4	3	Red
7	4	1	Blue