JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -3 EXAMINATION- 2023

M.Tech (DS) I-Semester (CSE/IT/ECE/CE/BT/BI)

COURSE CODE (CREDITS): 22M1WCI133 (3) MAX. MARKS: 35

COURSE NAME: Introduction to Statistical Learning

COURSE INSTRUCTORS: Dr. Hari Singh MAX. TIME: 2 Hours

Note: ***(a)****All questions are compulsory.*

(b)Marks are indicated against each question in square brackets.

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

---

Q1. Calculate the entropy of the following dataset. **[CO3][03 Marks]**

Salary	Age	Purchase
20000	21	Yes
10000	45	No
60000	27	Yes
15000	31	No
30000	30	Maybe
12000	18	No
40000	40	Maybe
20000	20	Maybe

Q2. Assume that a dataset has two classes in the output (yes/No). Describe the minimum and maximum values of probability(yes), probability(no) and entropy. Draw a graph between the entropy and probability(yes).

[CO3][03 Marks]

Q3. Calculate the Information Gain when the decision tree is built on Outlook as the root node from the following dataset.

[CO3][04 Marks]

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Q4. How does an ensemble learning model that is constructed from the same one model (base model) with different data samples behave in terms of bias and variance when the underlying base model has

(a) low bias and high variance

(b) high bias and low variance **[CO3][04 Marks]**

Q5. How can the voting classifier create a stronger model than the three independent weak models used to construct the voting classifier? Support your answer with a mathematical justification. **[CO3][03 Marks]**

Q6. What is bagging? How does bagging impact bias and variance? **[CO4][03 Marks]**

Q7. How does a random forest work for row sampling, column sampling and combined sampling? Assume a dataset of 100 rows and 5 input features and one output feature. Apply three sampling using 50% of the data. **[CO4][03 Marks]**

Q8. Write the differences between bagging and boosting with reference to (a) types of models used (b) sequential vs parallel (c) weightage of base learners **[CO4][04 Marks]**

Q9. Here we explore the maximal margin classifier on a datset. Justify your answer with appropriate explanation. **[CO4] [2x4=08 Marks]**

(a) We are given n=7 observations in p=2 dimensions. For each observation, there is an associated class label. Sketch the observations.
(b) Sketch the optimal separating hyperplane, and provide the equation for this hyperplane.
(c) Describe the classification rule for the maximal margin classifier. It should be something along the lines of "Classify to Red if $\beta 0+\beta 1X1+\beta 2X2 >0$, and classify to Blue otherwise." Provide the values for βo, $\beta 1$, and $\beta 2$.
(d) 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