



INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR  
Mid-Autumn Semester Examination 2022-23

Date of Examination \_\_\_\_\_ Session: (FN/AN) Duration: 2 hrs Full Marks: 20

Subject No.: CS60050

Subject Name: Machine Learning

Department/ Center/ School: Computer Science and Engineering

Special Instructions (if any): Write answers to all questions of a Section together.

Section A ( $5 \times 1$ )

**QUESTION 1:**

What is teacher noise?

Type: MCQ

- a) Noise due to hyper-parameter selection of the model
- b) Error in labeling
- c) Noise due to transmission
- d) None of the above

**QUESTION 2:**

What is the cardinality of inductive bias set for a rote learner?

Type: MCQ

- a) 2
- b) Dataset dependent
- c) 0
- d) Infinite

**QUESTION 3:**

Assume that the source population has 1 crore samples and follows Binomial distribution. 2 lakh samples are taken randomly from the population. According to the Central Limit Theorem, what is the distribution of the samples mean?

Type: MCQ

- a) Binomial
- b) Cannot be determined
- c) Skewed
- d) Normal

**QUESTION 4:**

John speaks truth 3 out of 5 times. An unbiased dice is tossed. John reports that there is a six. What is the chance that actually there was a six?

**QUESTION 5:**

A training example has the 6 attributes, out of which 2 attributes can take 3 values each, and the rest can take 2 values each. What is the size of the input space and the number of distinct hypotheses, respectively?

Handwritten notes:  
P.O. p  
plus 1 million  
plus 1 million  
reports.

### Section B (5 × 3)

#### QUESTION 6:

It is estimated that 30% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 99% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 5%. Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

#### QUESTION 7:

The following table has 5 attributes. Each attribute can take 3 values a; b; c. i) Use Find-S algorithm to find the final hypothesis. ii) Use Candidate elimination algorithm to find the final hypothesis at which specific and general hypothesis converge. Show the Version space.

Attribute 1	Attribute 2	Attribute 3	Attribute 4	Attribute 5	Label
a	c	b	a	b	1
a	a	a	a	a	0
a	c	b	a	a	0
a	a	b	b	c	1
a	b	a	c	a	0
b	b	a	a	a	0

#### QUESTION 8:

Consider the following Table. Find the root node along with the corresponding Information Gain using Entropy.

Outlook	Temperature	Play
sunny	hot	0
rainy	cool	0
cloudy	hot	1
humid	hot	0
sunny	cool	1
cloudy	cool	1

#### QUESTION 9:

Solve the question on Bayesian network:

i) There are 4 features present, A, B, C and D. Feature B and C are influenced by Feature A. However, no other features can influence A. Both Feature B and C influence Feature D. Form the Bayesian network for this relation.

ii) Using the Bayesian Network formed in i), compute  $P(A, \sim B, C, D)$  using the following information.  $P(A) = 0.2$ ,  $P(A|B) = 0.4$ ,  $P(A|C) = 0.3$ ,  $P(A|D) = 0.6$ ,  $P(A|C, B) = 0.3$ ,  $P(A|\sim C, B) = 0.1$ ,  $P(A|\sim B, C) = 0.2$ ,  $P(A|B, C) = 0.4$   $P(B) = 0.4$ ,  $P(B|A) = 0.2$ ,  $P(B|C) = 0.8$ ,  $P(B|D) = 0.1$ ,  $P(B|C, D) = 0.5$ ,  $P(B|\sim C, A) = 0.3$ ,  $P(B|\sim D, C) = 0.6$   $P(C) = 0.6$ ,  $P(C|A) = 0.5$ ,  $P(C|B) = 0.7$ ,  $P(C|D) = 0.3$ ,  $P(C|A, B) = 0.3$ ,  $P(C|\sim A, B) = 0.1$ ,  $P(C|\sim B, D) = 0.2$ ,  $P(C|B, D) = 0.4$   $P(D) = 0.2$ ,  $P(D|A) = 0.4$ ,  $P(D|B) = 0.6$ ,  $P(D|C) = 0.8$   $P(D|A, B) = 0.4$ ,  $P(D|\sim A, B) = 0.9$ ,  $P(D|\sim B, C) = 0.6$ ,  $P(D|B, C) = 0.7$

#### QUESTION 10:

Draw the decision tree with its height not more than 3, to represent the boolean function:

$$[A \wedge D] \vee [B \wedge C]$$