INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

Stamp / Signature of the Invigilator

EXAMINATION (End Semester)	SEMESTER (Autumn)

Roll Number										Section		Name	

Subject Number	C	S	6	0	0	5	0	Subject Name	*MACHINE LEARNING*

Department / Center of the Student		Additional sheets	

Instructions and Guidelines to Students Appearing in the Examination

1. Ensure that you have occupied the seat as per the examination schedule.
2. Ensure that you do not have a mobile phone or a similar gadget with you even in switched off mode. Note that loose papers, notes, books should not be in your possession, even if those are irrelevant to the paper you are writing.
3. Data book, codes or any other materials are allowed only under the instruction of the paper-setter.
4. Use of instrument box, pencil box and non-programmable calculator is allowed during the examination. However, exchange of these items is not permitted.
5. Additional sheets, graph papers and relevant tables will be provided on request.
6. Write on both sides of the answer script and do not tear off any page. Report to the invigilator if the answer script has torn page(s).
7. Show the admit card / identity card whenever asked for by the invigilator. It is your responsibility to ensure that your attendance is recorded by the invigilator.
8. You may leave the examination hall for wash room or for drinking water, but not before one hour after the commencement of the examination. Record your absence from the examination hall in the register provided. Smoking and consumption of any kind of beverages is not allowed inside the examination hall.
9. After the completion of the examination, do not leave the seat until the invigilator collects the answer script.
10. During the examination, either inside the examination hall or outside the examination hall, gathering information from any kind of sources or any such attempts, exchange or helping in exchange of information with others or any such attempts will be treated as adopting 'unfair means'. Do not adopt 'unfair means' and do not indulge in unseemly behavior as well.

Violation of any of the instructions may lead to disciplinary action.

Signature of the Student

To be filled in by the examiner

Question Number	1	2	3	4	5	6	7	8	9	10	Total
Marks Obtained											

Marks obtained (in words)	Signature of the Examiner	Signature of the Scrutineer

Instructions: Answer all FOUR questions. Time = 2hrs. Total marks = 4x10 = 40. Write your answers only in the space provided. Show the solution steps as required. The question paper has total 8 pages.

Rough Work

1. Assume that you have a concept learning problem that only allows pure conjunctive formulas and we have a domain with the following attributes. A don't care is represented by the "?" symbol.

Attr1	A, B	*Attr2*	A, B, C	*Attr3*	1, 2, 3	*Attr4*	a, b	*Attr5*	1, 2, 3, 4

Given the following seven examples:

Example	*Class*	*Example*	*Class*
<A,B,2,a,3>	+	<A,B,2,a,4>	+
<B,B,1,a,3>	-	<A,C,3,b,3>	-
<A,C,2,a,1>	+	<A,C,2,a,2>	-
<B,A,1,a,3>	-		

(a) What is the size of the hypothesis space? [2]

(b) What is the hypothesis learned by the FIND-S algorithm? [2]

(c) Are the examples consistent? If yes, what is the specific (S) and general (G) boundary of the version space as obtained by the candidate elimination algorithm? [4]

S:

G:

(d) The candidate elimination algorithm is an example of unsupervised learning (True/False)? [2]

2. Consider the following data for a binary classification problem. There are two binary attributes A and B.

A	B	Class	A	B	Class
T	F	+	F	F	-
T	T	+	F	F	-
T	T	+	F	F	-
T	F	-	T	T	-
T	T	+	T	F	-

(a) What is the information gain for splitting on attribute A as root? [2]

(b) Which attribute will be chosen as root of the decision tree built using information gain? [2]

(c) What is the Boolean formula learnt by the complete tree above for the + class? Few misclassifications are allowed. For a mixed class leaf, majority rule is adopted for class decision. [4]

(d) Define *description length* of a decision tree: [2]

3. Consider the following 2-class (w_1, w_2) classification problem involving a single feature x. Assume equal class priors. The class conditional distributions are:

$$p(x|w_1) = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \qquad p(x|w_2) = \begin{cases} 2-2x & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Derive the Bayes decision boundary? [4]

(b) What is the Bayes classification error? [2]

(c) What is the new decision boundary if the apriori probability of class w_1 is changed to *0.7*? [2]

(d) K-NN rule approximates the MAP classifier (True/False)? [2]

4. As part of a comprehensive study of people's happiness we have been collecting important data from past students of the CS60050 course. We ask the following eight objective questions:

i. Do you party frequently [*Party*: Yes/No]? ii. Are you smart [*Smart*: Yes/No]? iii. Are you creative [*Creative*: Yes/No]? iv. Did you do well on all your homework assignments? [*HW*: Yes/No] v. Do you use a Mac? [*Mac*: Yes/No] 4 vi. Did your term project succeed? [*Project*: Yes/No] vii. Did you succeed in your exam? [*Success*: Yes/No] viii. Are you currently Happy? [*Happy*: Yes/No]

After consulting a behavioral psychologist we obtained the following complete set of conditional relationships: • HW depends only on Party and Smart • Mac depends only on Smart and Creative • Project depends only on Smart and Creative • Success depends only on HW and Project • Happy depends only on Party, Mac, and Success.

(a) Draw the Bayesian network for the above scenario. [2]

(b) Write joint distribution of all the variables in this network as product of conditional probabilities: [2]

(c) Draw the Bayesian networks (of four variables) that correspond to the following factorization of the joint probability distribution $P(A, B, C, D)$. [4]

(i) $P(B|A, C)P(A)P(C|D)P(D)$

(ii) $P(A|B)P(C|B)P(B)P(D)$

(iii) $P(D|C)P(C|B)P(B|A)P(A)$

(iv) $P(B|A)P(A)P(C|D)P(D)$

(d) Suppose the fraction of rural population of India who owns a car is 15% and the fraction of urban population of India who owns a car is 23%. If $1/5^{th}$ of the Indian population is urban and the rest are rural, what is the probability that a person who owns a car is an urban person? [2]

--- END ----

Rough Work

Rough Work