	INDIAN INSTITUTE OF TECHNOLOGY **KHARAGPUR**	**Stamp/Signature of the Invigilator**

EXAMINATION (Mid Semester)	**SEMESTER (Autumn)**

Roll Number											**Name**	
Subject Number	**C**	**S**	**6**	**0**	**0**	**5**	**0**		**Subject Name**		***Machine Learning***	
Department/Centre 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. Date book, codes or any other materials are allowed only under the instruction from the paper-setter.
4. Use the 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. 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 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 Scrutinizer**

Instructions: ***Answer all FOUR questions in the spaces (box) provided. Do Not Issue Separate Answerscript.***

Time: 2 hrs, Total Marks: 4x10=40, Provide explanations wherever required. The question paper has 12 pages

---

Rough Work

1. Assume that you have a concept learning problem where the input space is a subset of the two dimensional x-y plane with $0 \leq x \leq 10$, and $0 \leq y \leq 10$ (i.e., a 10x10 square in the first quadrant). The hypothesis space consists of axis parallel rectangles that lie completely within the input space and has corners with *integer* co-ordinate values. Rectangles touching the boundaries of input space are also considered to be within the input space. We are provided with the following eight training instances (*x*, *y*, *class*): **[2 + 2 + 2 + 2 + 2 = 10]**
{(2, 2, +), (3, 3, +), (4, 4, +), (3, 2, +), (3, 4, +), (4, 7, -), (8, 8, -), (3, 7, -)}

(a) What is the cardinality of the hypothesis space?

(b) Provide a brief definition of the term "version space":

(c) Sketch the version space for the above set of training instances.

(d) What is the cardinality of the version space for the above set of training instances?

(e) Suppose you are allowed to query the class label of one more point (with unknown class label) such that there is maximal reduction in size of the version space after adding this point to existing training instances. Mark the possible locations of the point that you would query.

Rough Work

2. You have designed a very simple binary classifier that always outputs the class that is in majority in the training set. It does not consider the value of input features in any form. You have a data set consisting of 25 examples from each of the two classes. **[2 + 2 + 4 + 2 = 10]**

(a) We perform cross validation of the classifier by randomly choosing 20% of the examples as holdout set and the remaining examples as training set. What is the expected accuracy of the classifier on the holdout set?

(b) What is the expected leave-one-out cross validation accuracy of the classifier for the given examples?

(c) Consider the following table of training points for a binary classification problem. There are two classes `+' and `-'. There are two binary (T/F) attributes A and B. XOR distance between two points (x_1, y_1) and (x_2, y_2) is defined as d = ((x_1 *XOR* x_2) XOR (y_1 *XOR* y_2)). Using XOR distance as the distance measure, how does the 1-nearest neighbor rule classify the test point (F, T)?

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	-

(d) Briefly explain the condensed nearest neighbor algorithm.

Rough Work

3. We have K classes, C_1, C_2, …, C_K, where each class C_i is uniformly distributed over $-2^{i-2} < x < 2^{i-2}$.

(a) Draw a sketch of the class conditional probability distributions. **[2 + 8 = 10]**

(b) Assume all the classes have equal prior probabilities. Derive the Bayes classification error $P_{Bayes}(\epsilon)$?

DERIVATION:

ANSWER:

Rough Work

4. Rovers, an experienced autonomous vehicle, has landed once again in Mars. It has a battery which powers its antenna. Usually the antenna operates only when the battery is well charged. Rovers communicates with a Mars orbiter satellite. Communication often breaks down when the antenna does not operate. Rovers also sometimes stops communication when there is a storm going on in Mars. We have collected the following data from previous expeditions of Rovers to Mars. **[2 + 4 + 4 = 10]**

Battery Charge	Antenna Operating	Mars Storm	Number of previous expeditions where it *could communicate*	Number of previous expeditions where it *could not communicate*
Hi	Yes	No	5	2
Hi	Yes	Yes	1	2
Hi	No	No	1	5
Hi	No	Yes	0	4
Lo	Yes	No	9	0
Lo	Yes	Yes	5	1
Lo	No	No	1	2
Lo	No	Yes	0	2

(a) Draw the structure of Bayesian network for the above scenario.

(b) Write down the conditional probability tables for each node in the network. Fill up the tables with values estimated from the records of previous expeditions.

(c) It was found that Rovers is not communicating. Use the Bayesian network to compute the probability that there is a storm going on in Mars. Explain your answer.

--- END ----

Rough Work

Rough Work