INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

Stamp / Signature of the Invigilator

EXAMINATION (End Semester)										**SEMESTER (Spring 2024-2025)**			
Roll Number										**Section**		**Name**	
Subject Number	C	S	6	0	0	2	0	**Subject Name**		*FOUNDATIONS OF ALGORITHM DESIGN AND MACHINE LEARNING*			
Department / Center of the Student										**Additional sheets**			

Important Instructions and Guidelines for Students

1. You must occupy your seat as per the Examination Schedule/Sitting Plan.
2. Do not keep mobile phones or any similar electronic gadgets with you even in the switched off mode.
3. Loose papers, class notes, books or any such materials must not be in your possession, even if they are irrelevant to the subject you are taking examination.
4. Data book, codes, graph papers, relevant standard tables/charts or any other materials are allowed only when instructed by the paper-setter.
5. Use of instrument box, pencil box and non-programmable calculator is allowed during the examination. However, exchange of these items or any other papers (including question papers) is not permitted.
6. Write on both sides of the answer script and do not tear off any page. **Use last page(s) of the answer script for rough work.** Report to the invigilator if the answer script has torn or distorted page(s).
7. It is your responsibility to ensure that you have signed the Attendance Sheet. Keep your Admit Card/Identity Card on the desk for checking by the invigilator.
8. You may leave the examination hall for wash room or for drinking water for a very short period. Record your absence from the Examination Hall in the register provided. Smoking and the consumption of any kind of beverages are strictly prohibited inside the Examination Hall.
9. Do not leave the Examination Hall without submitting your answer script to the invigilator. **In any case, you are not allowed to take away the answer script with you.** After the completion of the examination, do not leave the seat until the invigilators collect all the answer scripts.
10. During the examination, either inside or outside the Examination Hall, gathering information from any kind of sources or exchanging information with others or any such attempt will be treated as **'unfair means'**. Do not adopt unfair means and do not indulge in unseemly behavior.

Violation of any of the above instructions may lead to severe punishment.

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

Indian Institute of Technology Kharagpur
Department of Computer Science and Engineering

Foundations of Algorithm Design and Machine Learning (CS60020) **Spring 2024-2025**

April-2025 **End-Semester Examination** Maximum Marks: 100

Instructions:

- Write your answers in the question paper itself. Be brief and precise. Answer *all* questions.
- Write the answers only in the respective spaces provided. The last two blank pages may be used for rough work or leftover answers.
- In case you may need more space/pages, please ask for additional sheets in the exam hall and attach the same with this booklet while submitting.
- If you use any algorithm / result / formula covered in the class, just mention it, do not elaborate (unless the same thing has been explicitly asked to answer in the question).

— Question Paper Starts from Next Page —

Q1. **[Decision Tree Learning]** **8 marks**

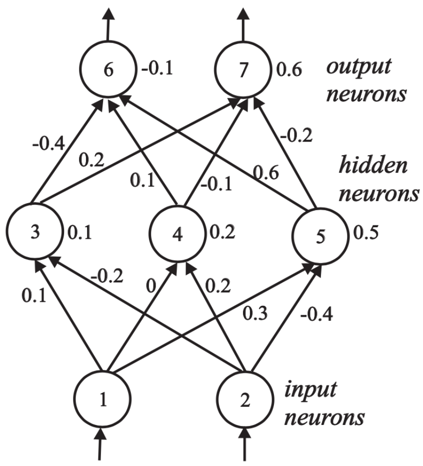
The table (shown right) contains data samples of six patients examined in a hospital. Use entropy based information gain measures to construct a minimal decision tree that can predict whether or not a patient is likely to have a heart attack. *Show each step of your computation.*

Patient IDs	**Attributes**				**Heart Attack**
	Gender	**Smoker**	**Exercise**	**ChestPain**	
1	Female	No	Regular	Yes	YES
2	Male	No	Never	Yes	YES
3	Male	Yes	Never	No	YES
4	Female	No	Never	No	NO
5	Male	Yes	Regular	Yes	YES
6	Male	No	Regular	No	NO

Solution: [Helper Data: $\log_2 3 = 1.585$]

Q2. [Artificial Neural Networks (ANN): Backpropagation] **18 marks**

Consider the 3-layer ANN (having one input, one output and one hidden layer) given in the figure (right) which is being trained to distinguish between `nails` (output encoding as 10) and `screws` (output encoding as 01). Let the learning rate be $\eta = 0.1$ and the initial weights (w_{ij} from a node i to another node j) are mentioned in the directed edges of the figure. Also, the bias (w_{0j} for a node j) is specified beside each node (neuron) directly. Assume that, each node (neuron), n_j ($3 \leq j \leq 7$), applies the default sigmoid activation function $\left(\text{i.e., } o_j = \sigma(s_j) = \frac{1}{1+\exp(-s_j)}\right)$ over the weighted input sum $\left(\text{i.e., } s_j = w_{0j} + \sum_i w_{ij} o_i\right)$.



Suppose you train this ANN with only *one* example: $T = \{(0.6, 0.1), \texttt{nail}\}$, which indicates that when the inputs are $o_1 = 0.6$ and $o_2 = 0.1$, the true outcomes become $d_6 = 1$ and $d_7 = 0$ (target class being `nail`). Answer the following questions.

(a) Allowing a *forward pass* with the training example T, compute the values of the outputs of all nodes/neurons. In particular, show your calculation in details for every s_j and o_j ($3 \leq j \leq 7$). **(5)**

Solution:

(c) Present the (delta) weight update rule/equation followed during backpropagation of ANNs. **(2)**

Solution:

(c) Allowing a *backward pass* with the training example T, compute the updates for each weight in the network (as per your equation given in Part (b)). In particular, show your calculation in details for every δ_j and Δw_{ij} (where, $0 \leq i \leq 5$ and $3 \leq j \leq 7$). **(11)**

Solution:

Q3. **[Support Vector Machines (SVM)]** **10 marks**

Consider a set of 2-dimensional training data points (x_1, x_2) belonging to two classes '$\oplus$' and '$\ominus$', respectively, as shown below.

- Class '$\oplus$': $(3,1)$; $(3,-1)$; $(6,1)$; $(6,-1)$
- Class '$\ominus$': $(1,0)$; $(0,1)$; $(0,-1)$; $(-1,0)$

We design a linear hard-margin SVM to classify these linearly separable points. Answer the following.

(a) Pictorially (graphically) represent the constellation of data points and the optimal separating hyperplane. Write the equation of the optimal separator and mention the width of the margin (figuring it out manually from the diagram/graph you have shown). **(2)**

Solution:

(b) Which data points are the support vectors here? **(2)**

Solution:

(c) What weight vector and threshold (bias) value are being learnt using hard-margin SVM training algorithm with these eight training points? Show the detailed calculations. **(4)**

Solution:

(d) Using the learnt weights and threshold values (in part (c)), what is the margin you get for the optimal classifier? Derive mathematically. **(2)**

Solution:

Q4. **[Classifier Evaluation]** **9 marks**

You wrote a spam filtering program by yourself and now you are testing your program on 100 emails among which you already knew that 20% emails are spams. However, upon running your program on those 100 email corpus, it predicted $\frac{1}{2}$ of the 'spam' emails as non-spam. Answer the following:

(a) In order to push the *Accuracy* $\geq 75\%$, how many 'non-spam' emails at most (maximum) you can afford to mis-predict as spams? **(3)**

Solution:

(b) With the derived setup in Part (a), i.e., when your *Accuracy* is exactly 75%, present the confusion matrix (in tabular form). **(3)**

Solution:

(c) As per your confusion matrix that you presented in Part (b), calculate *Precision*, *Recall* and F_1-*score* of your spam filtering program. **(3)**

Solution:

Q5. **[Dimensionality Reduction: Principal Component Analysis (PCA)]** **8 marks**

Given the (x, y)-coordinates of four data points in two-dimensional space: $(4, 1)$, $(2, 3)$, $(5, 4)$ and $(1, 0)$, calculate the first principal component. Show your calculations in details.

Solution:

Q6. **[Learning Theory: VC Dimension]** **15 marks**

(a) What the VC-dimension of axis-aligned rectangles in a 2-dimensional plane? Justify / Prove. **(5)**

Solution:

(b) What is the VC-dimension of axis-aligned squares in a 2-dimensional plane? Justify / Prove. **(5)**

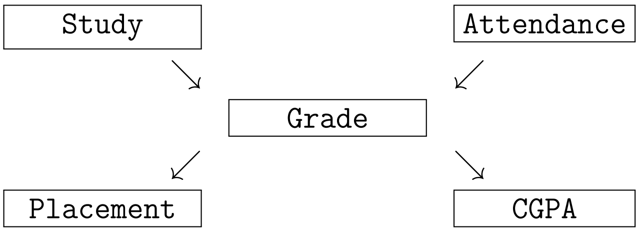
Solution:

(c) Let the VC-dimensions of two hypothesis classes, H_1 and H_2, be $VCDim(H_1) = d_1$ and $VCDim(H_2) = d_2$. Prove that, the VC-Dimension of the union of these hypothesis, i.e. $H = H_1 \cup H_2$, will be at most $(d_1 + d_2 + 1)$, i.e. $VCDim(H) \leq VCDim(H_1) + VCDim(H_2) + 1$. **(5)**

Solution:

Q7. **[Bayesian Learning: Expectation-Maximization (EM) Algorithm]** **16 marks**

Consider the Bayes Network structure shown below. From the figure below, we abbreviate as follows: **S** = `Study` well, **A** = high `Attendance`, **G** = good `Grade`, **P** = better `Placement`, **C** = high `CGPA`.



You are given the following $K = 8$ training examples as shown below, where only two examples contain unobserved values (marked with ?), namely, p_7 and c_8. You have to simulate a few steps of the simplified EM algorithm by hand.

K	S	A	G	P	C
$k = 1$	1	0	1	1	1
$k = 2$	0	1	1	1	0
$k = 3$	1	1	1	1	1
$k = 4$	0	0	0	0	1
$k = 5$	0	0	0	1	0
$k = 6$	0	0	0	0	0
$k = 7$	1	1	1	?	1
$k = 8$	1	1	1	1	?

Notation: Here, s_k, a_k, g_k, p_k, and c_k indicate the values of **S**, **A**, **G**, **P**, and **C**, respectively, as seen in the k-th example/row. For example, $s_1 = 1$, $a_1 = 0$, $g_1 = 1$, $p_1 = 1$, and $c_1 = 1$.

Answer the following questions:

(a) Given that *all variables are Boolean*, how many basic parameters you need to estimate for the given Bayes Network?
For example, one parameter will be $\theta(g \mid 11)$, which stands for $\mathbb{P}(G = 1 \mid S = 1, A = 1)$. **(1)**
Solution:

(b) Now, you need to simulate the first E-step of the EM algorithm. Before you start, you initialize all the parameters as 0.5, and then proceed to execute the E-step. What are the following expectation values that will get calculated in this E-step? In particular, calculate the following: **(5)**

- $\mathbb{E}(p_7 = 1 \mid s_7, a_7, g_7, c_7, \theta) =?$
- $\mathbb{E}(c_8 = 1 \mid s_8, a_8, g_8, p_8, \theta) =?$

(*Note that, only two examples (k=7 and k=8) contains unobserved variables, where* $p_7 =?$, *but* $s_7 = a_7 = g_7 = c_7 = 1$*; and* $c_8 =?$, *but* $s_8 = a_8 = g_8 = p_8 = 1$, *respectively.*)

Solution:

(c) Now, you need to simulate the first M-step of the EM algorithm. What will be the estimated values of all the model parameters (which you identified in Part (a)) after this M-step? **(5)**

(*Note that, you can use the expected count only when the variable is unobserved in an example*)

Solution:

(d) Lastly, you again simulate the second E-step of the EM algorithm. What are the following expectation values that will get calculated in this E-step? In particular, calculate the following: **(5)**

- $\mathbb{E}(p_7 = 1 \mid s_7, a_7, g_7, c_7, \theta) = ?$
- $\mathbb{E}(c_8 = 1 \mid s_8, a_8, g_8, p_8, \theta) = ?$

Solution:

Q8. [Unsupervised Learning: K-Means Clustering] **16 marks**

Suppose the following dataset (consisting of (x,y)-coordinates of eight data points in a 2-dimensional plane) is given: $(1,2)$, $(2,5)$, $(2,10)$, $(4,9)$, $(5,8)$, $(6,4)$, $(7,5)$, and $(8,4)$. You need to run K-Means algorithm (till termination) with $K=3$ to cluster these points. Assume that, Euclidean distance measure is used as the distance computing function for the dataset. Answer the following.

(a) Assuming the initial centroids as $(1,2)$, $(2,10)$, and $(5,8)$, show in details the execution of K-Means algorithm (till termination) with $K=3$. In particular, indicate the set of points that come under each cluster after every iteration and also compute their centroid to be used for the next iteration. Indicate when and how you decided to terminate/stop. **(8)**

Solution:

Iteration	Output	Cluster-1	Cluster-2	Cluster-3
0 (init.)	Points	–	–	–
	Centroid	$(1,2)$	$(2,10)$	$(5,8)$
1 (cont.)	Points			
	New Centroid			
2 (cont.)	Points			
	New Centroid			
3 (cont.)	Points			
	New Centroid			
4 (stop)	Points			
	New Centroid			

Decision to terminate:

(b) Upon termination, compute the average silhouette coefficient (SC) of the overall clustering only for the three clusters formed. Show your calculations in details. **(8)**

Note: For your convenience, pairwise distances between points are shown in the following table.

Pairwise Distance		Data Points							
		$(\mathbf{1},\mathbf{2})$	$(\mathbf{2},\mathbf{5})$	$(\mathbf{2},\mathbf{10})$	$(\mathbf{4},\mathbf{9})$	$(\mathbf{5},\mathbf{8})$	$(\mathbf{6},\mathbf{4})$	$(\mathbf{7},\mathbf{5})$	$(\mathbf{8},\mathbf{4})$
Data Points	$(\mathbf{1},\mathbf{2})$	0.000							
	$(\mathbf{2},\mathbf{5})$	3.162	0.000						
	$(\mathbf{2},\mathbf{10})$	8.062	5.000	0.000					
	$(\mathbf{4},\mathbf{9})$	7.616	4.472	2.236	0.000				
	$(\mathbf{5},\mathbf{8})$	7.211	4.243	3.606	1.414	0.000			
	$(\mathbf{6},\mathbf{4})$	5.385	4.123	7.211	5.385	4.123	0.000		
	$(\mathbf{7},\mathbf{5})$	6.708	5.000	7.071	5.000	3.606	1.414	0.000	
	$(\mathbf{8},\mathbf{4})$	7.280	6.083	8.485	6.403	5.000	2.000	1.414	0.000

Solution:

— Question Paper Ends Here —