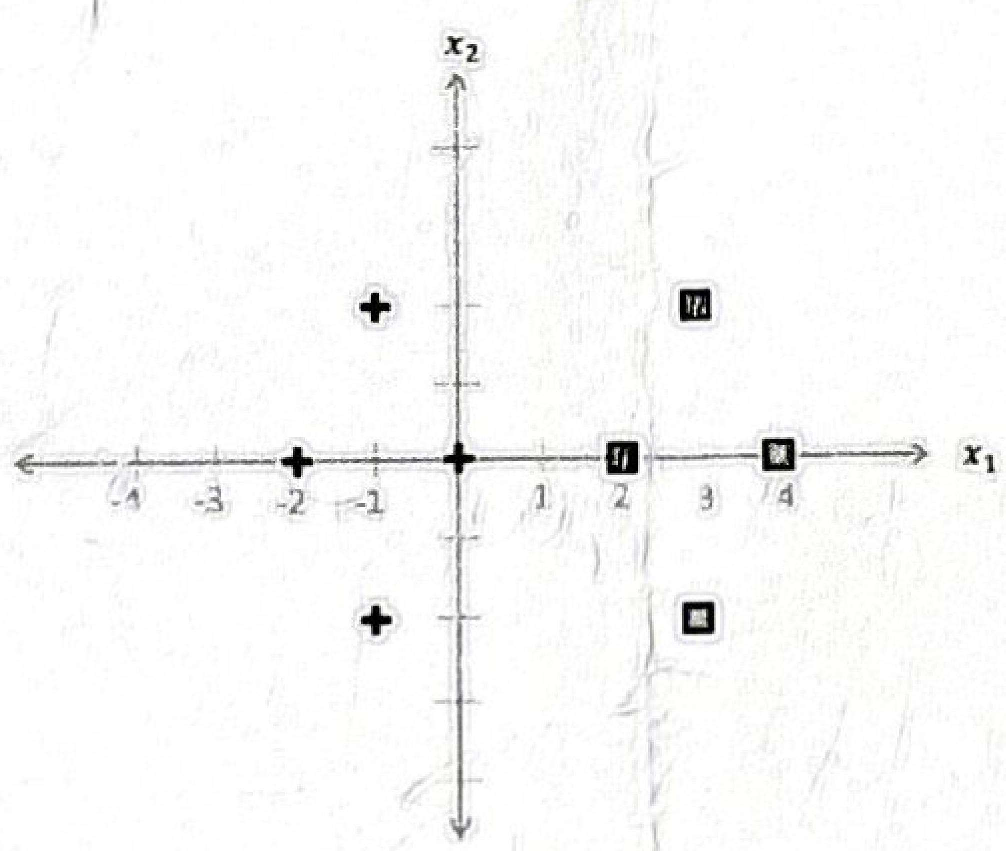
Department of Computer Science and Engineering
Indian Institute of Technology Jodhpur
CSL7620 - Machine Learning

September 3, 2022

Instructions:
1. Read the questions carefully.
2. All questions are mandatory.
3. If you find anything unclear/incorrect in any question, make a reasonable assumption and proceed.

Time: 1 hour — *Minor-I* — Maximum Marks: 15

1. A hypothesis resulted in by a learning algorithm has a very low empirical risk whereas a high true/expected risk. Write any two possible reasons behind this kind of an outcome. [1]
2. Explain the bias-complexity trade off. [2]
3. Consider the linear regression problem $y = w_0 + w_1 \times x_1 + w_2 \times x_2 + w_3 \times x_3$. What is the significance of w_0? [1]
4. Explain the importance of normalization in linear regression. [1]
5. Consider a binary classification problem for 2D Gaussian data. Data points drawn from two Gaussian distributions are shown in the figure below using + and ◘ for class 1 and 2 respectively. Find the Bayes decision boundary using the given data points while assuming that both the classes have equal prior probability. Comment on the nature of the decision boundary. [5]



6. You have been asked to design a spam classifier for the incoming emails using following information. [5]

    1. Out of the 1800 randomly observed emails, 360 are spams.
    2. The word 'offer' is found in 288 spam emails and in 216 non-spam emails.
    3. The word 'discount' is found in 270 spam emails and in 144 non-spam emails.
    4. The word 'rules' is found in 54 spam and in 288 non-spam emails.

    Design a naive Bayes classifier for the given binary classification problem which can be used to predict the category of a new email.

Optimization for Data Science (MAL7070/71)

Dr. Md Abu Talhamainuddin Ansary

Answer all questions.

Counter examples (if required) must be at least 2-dimensional. No marks for one dimensional counter example.

Students are allowed to use scientific calculator

1. Justify the following statements. (4×2.5)

   (i) If S_1 and S_2 are two convex sets then $S_1 \ominus S_2 = \{x = x_1 - x_2 : x_1 \in S_1, x_2 \in S_2\}$ is a convex set.

   (ii) Suppose $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then the set $S = \{x : f(x) \geq \alpha, \alpha > 0\}$ is a convex set.

   (iii) The least square problem $\min_{x \in \mathbb{R}^n} \frac{1}{2}\|Ax - b\|^2$, $A_{m \times n}$, $b_{m \times 1}$ has unique solution if column vectors of A are linearly independent.

   (iv) Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function. Then $\partial f(x^0)$ is a convex set for any $x^0 \in \mathbb{R}^n$.

2. If $y = \beta_1 x^2 + \beta_2 x + \beta_3$ be the best fitting curve for the set of points $\{(-1, 3), (-2, 2.5), (2, -1), (1, 2)\}$, then find the value of $(\beta_1, \beta_2, \beta_3)$. (5)

3. Consider $f(x) = x_1^2 - 2x_1x_2 + 4x_2^2 - 2x_1 + x_2$ and $x^0 = (2, 2)^T$. Find d^0, α_0, x^1, and d^1 using steepest descent method with exact line search technique. Show that d^1 is orthogonal to d^0. (5)

4. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function, ∇f is Lipschitz continuous in $\mathbb{R}^n$ with Lipschitz constant $L > 0$, and the level set $M = \{x \in \mathbb{R}^n : f(x) \leq f(x^0)\}$ is bounded. Further suppose $\{x^k\}$ be a sequence generated by $x^{k+1} = x^k + \alpha_k d^k$ where d^k is a descent direction and α_k satisfies Armijo-Wolfe condition. Then show that

$$\lim_{k \to \infty} \cos^2(\theta_k) \|\nabla f(x^k)\|^2 = 0$$

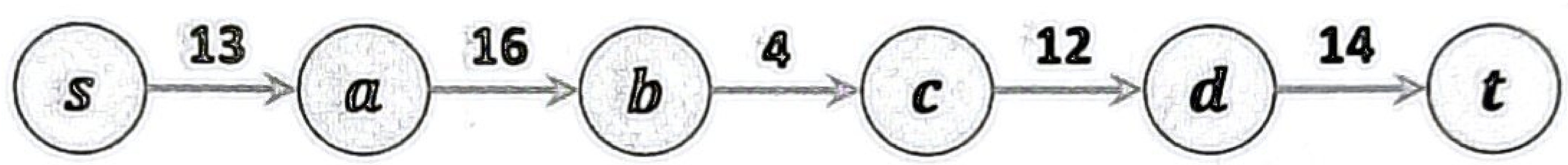
   where θ_k is the angle between $\nabla f(x^k)$ and d^k. (5)

Advanced Data Structures and Algorithms (CSL7560)

Minor Test 2
15 October, 2022

ATTENTION! This test consists of a total of 5 puzzles. You may attempt any 3 puzzles of your choice. You can attempt more than 3 puzzles too; in that case, all your answers will be evaluated and the best 3 will be taken into consideration. Each puzzle also has a hint which you can use if you are stuck. All the best! ☺

Puzzle 1 (Maximum flow). The analysis of the Edmonds-Karp algorithm relies on the fact that each edge is reversed at most n times in the residual graph. For example (see below), the edge (b, c) will be reversed after a flow of 4 is pushed through the path from s to t. Construct a graph where an edge is reversed **twice** by the Edmonds-Karp algorithm (other edges may also be reversed in the process).



Hint 1. *A possible solution is to modify the given path. After a flow of 4 is pushed through it, the edge* (b, c) *will be reversed and become* (c, b)*. To reverse it again, add vertices and edges to the graph, so as to create a new s–t path passing through* (c, b)*. Take care to ensure that your edge capacities are set appropriately, and the path which uses* (c, b) *is a shortest s–t path at the time when a flow of 4 is being pushed through it.*

Puzzle 2 (Knapsack). In class, we saw that greedily selecting the item with the maximum VALUE/WEIGHT ratio (the MDM algorithm) gives just a 2-approximation algorithm for the 0-1 integer knapsack problem.

But now, let us consider a special case. Suppose that the order of the items when they are sorted by increasing weight is exactly the same as their order when the items sorted by decreasing value.

Prove that the MDM algorithm actually gives a perfectly optimal solution in this special case. Show all the steps in your proof. (For simplicity, assume that all the item values are distinct, and so are all the item weights.) You may use the items in Table 1 (which is already sorted) as a guiding example for your proof (with the weight capacity of the knapsack being $W = 100$).

Item	VALUE	WEIGHT	
Item 1	90	5	18
Item 2	80	10	8
Item 3	60	20	3
Item 4	50	25	2
Item 5	35	35	1
Item 6	30	45	
Item 7	20	60	
Item 8	15	75	
Item 9	10	100	

Table 1: A list of items available for selection. Weight capacity of the knapsack is $W = 100$.

Hint 2. *Claim that, in this special case, the optimal solution must always select the item with the highest* VALUE/WEIGHT *ratio until the knapsack cannot hold any more items. To prove this claim, show that if an optimal solution* OPT *selects an item with a lower* VALUE/WEIGHT *ratio instead of an item with a higher* VALUE/WEIGHT *ratio, then you can find a solution that is even better than* OPT*, which is a contradiction.*

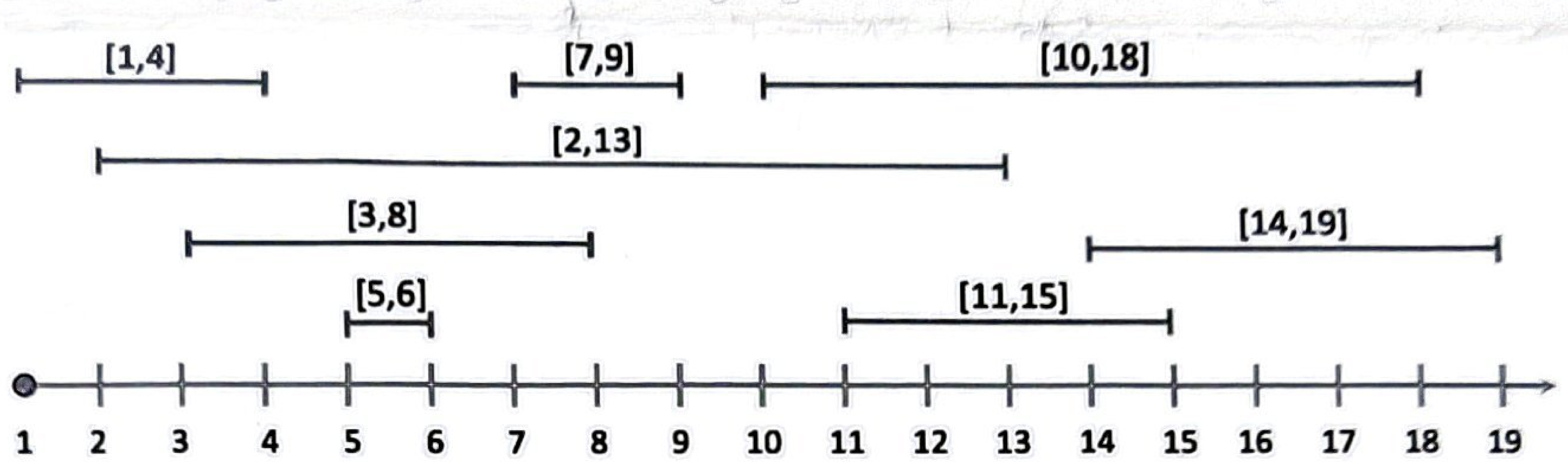
Puzzle 3 (Interval scheduling). In class, we saw that greedily selecting the interval with the earliest end time (GREEDY-EARLIEST-END-TIME) gives an optimal solution for interval scheduling.

But now, let us introduce a mildly different algorithm (DELETE-DOMINATING-INTERVALS) for interval scheduling. An interval A is said to **dominate** another interval B, if A starts before B and A ends after B. For example (see below), $[2,13]$ dominates $[7,9]$, $[3,8]$ dominates $[5,6]$, and $[10,18]$ dominates $[11,15]$. Let us now describe the algorithm DELETE-DOMINATING-INTERVALS. It has two steps.

1. Delete all those intervals that dominate another interval.

2. Find an optimal solution from the remaining intervals using GREEDY-EARLIEST-END-TIME.

In the example below, the intervals $[2,13]$, $[3,8]$, and $[10,18]$ are deleted by the first step of this algorithm, and $[1,4]$, $[5,6]$, $[7,9]$, and $[11,15]$ are selected by the second step as its final solution. Does this algorithm always give an optimal solution? If yes, prove it. If not, show an example where it fails.



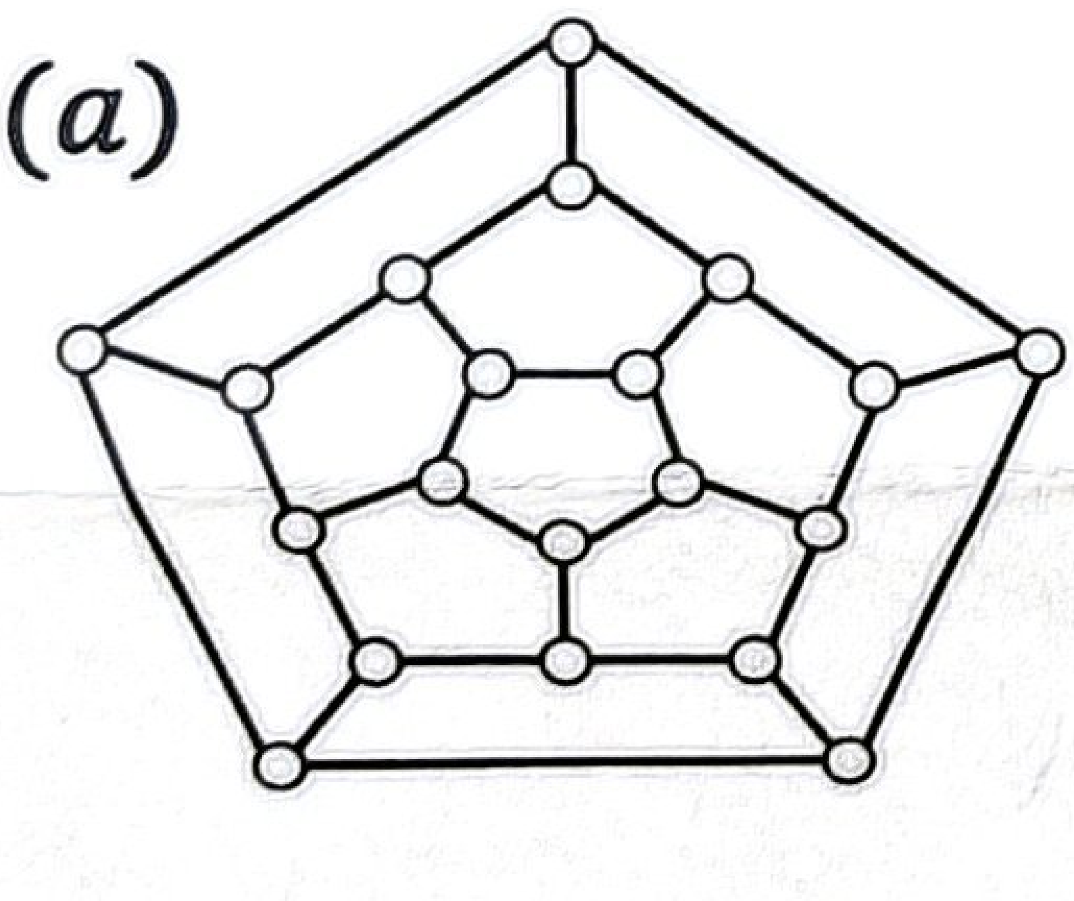
Hint 3. *If OPT has an interval which was deleted by the first step of* DELETE-DOMINATING-INTERVALS, *can you replace that interval by a new interval while ensuring a valid solution remains after its inclusion?*

Puzzle 4 (Longest common subsequence). Suppose that the English alphabet contains only 6 letters {a,b,c,d,e,f}. Design an efficient algorithm that finds the longest alphabetically-ordered subsequence of a given string. For example if the input to your algorithm is the string (eabfcadfbe), then the output of your algorithm should be (abcde) or (abcdf). Find the longest subsequence of (bafdecafdc) that is in alphabetical order. Show all the steps in the working of your algorithm.

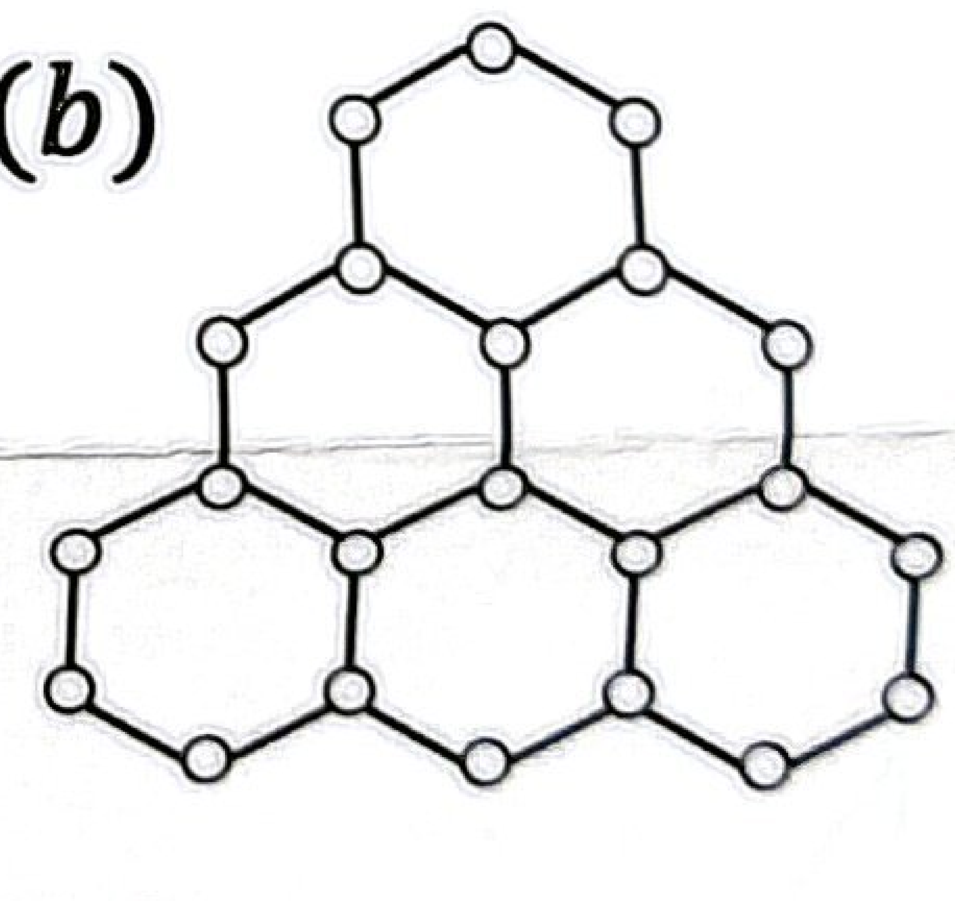
Hint 4. *Solve for longest common subsequence, or better still, longest increasing subsequence. Your choice!*

Puzzle 5 (Polynomial-time verification). A Hamiltonian cycle in a graph is a cycle that visits every vertex of the graph exactly once. How will you convince your friend in reasonably quick time that graph (a) below contains a Hamiltonian cycle, and graph (b) does not contain a Hamiltonian cycle?

(a)



(b)



Hint 5. *In graph (a), show a Hamiltonian cycle. For graph (b), show that it is* **bipartite**. *That is, partition its vertices into two sets: left and right, such that all edges of (b) connect a vertex in the left set to a vertex in the right set. Is the number of vertices in both sets the same? What do you conclude from this, and why?*

Computational Game Theory
MAL 7400
Minor 2

Name :

Max. Marks : 20

Roll. No. :

Q1 Consider the two person game given by matrix $\begin{pmatrix} (0,2) & (4,1) \\ (2,4) & (5,4) \end{pmatrix}$ and represent it in coalational form. Comment on the information lost during change in representation. Represent the derive derived coalational form into strategic form. Is the strategic form derived same as the previously given form? Justify. [2+1+2+1=6]

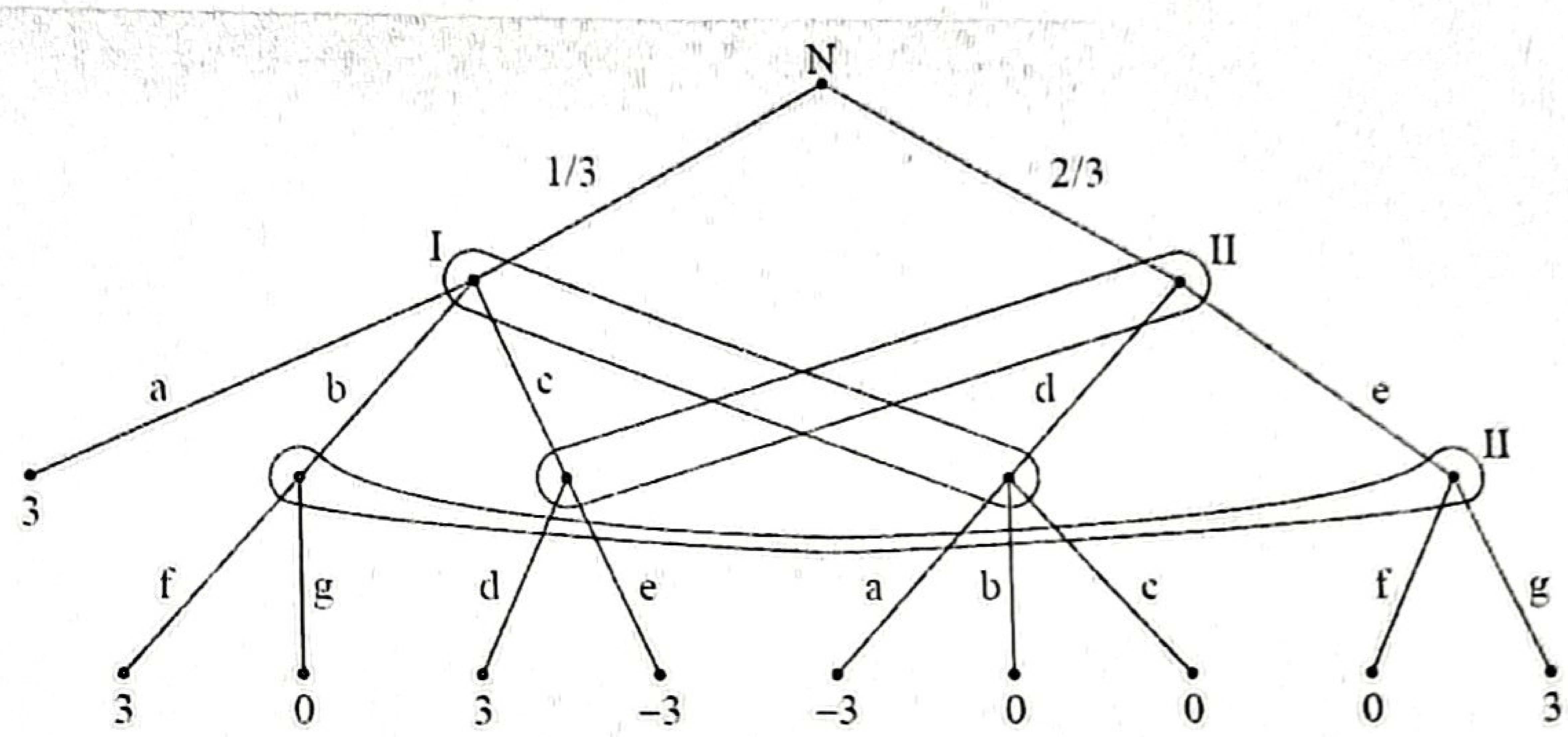
Q2 Consider the game $G=(N,V)$ where V is given by: $V(\phi)=0$, $V(S)=0$ if $|S|<2$, $V(S)=100$ if $|S|=2$ and $V(N)=125$ and $N=3$. Prove that the core is empty. Also find the ε^* determining the least core and cost of stability of the given game. [2+2+2=6]

Q3 Compute the Nucleolus of $G=(N,V)$ where $N=\{A,B,C\}$ and V is given by:

$$V(\phi)=0 \quad \begin{array}{l} V(\{A\})=-1 \\ V(\{B\})=0 \\ V(\{C\})=1 \end{array} \quad \begin{array}{l} V(\{AB\})=3 \\ V(\{AC\})=4 \\ V(\{BC\})=2 \end{array} \quad V(\{ABC\})=5$$

[4 Marks]

Q4 Convert the following extensive form game into strategic form and solve



MINOR-I EXAMINATION
ARTIFICIAL INTELLIGENCE

Max Duration: 1 hour **Max Marks: 20**

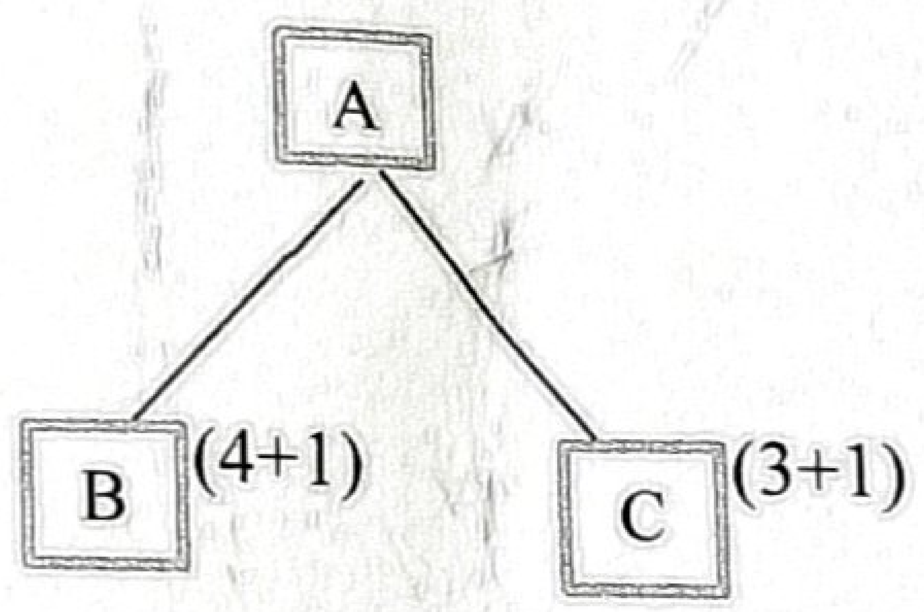
Instructions:

- This is a closed book exam.
- Ensure neatness, crispness, and relevance.
- In case you make any assumption, clearly write the same along with the reasons.
- There are 5 questions. Turn around the sheet.

1. Give algorithm for generate and test strategy. Whether it is depth first or breadth first search procedure. Give reasons for your answer. (2 + 2)

2. When would be greedy best first search be worse than simple breadth first search? Give example. (2 + 2)

3. Suppose that the first step of the operation of a greedy best-first search variant algorithm results in the following situation. (a + b, where a is the heuristic evaluation of a node and b is the cumulative path cost).

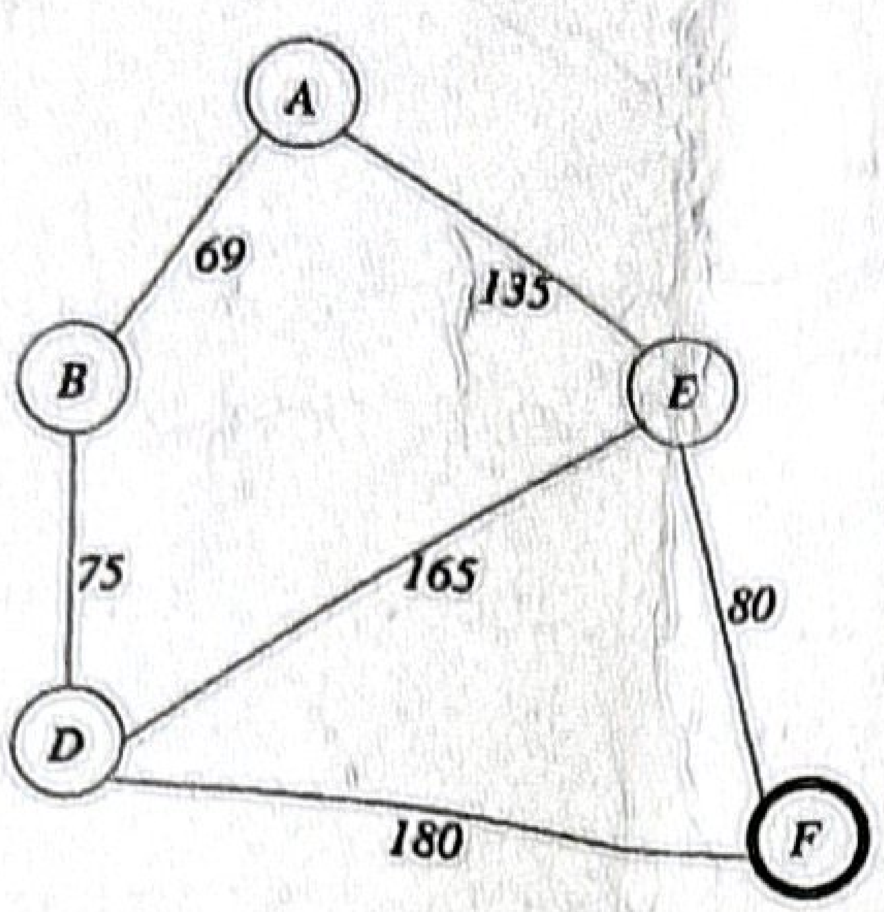


Any path between parent and immediate child has uniform cost b = 1. The lesser the cost a+b for a node, the better it is. D (with a = 4) is the child of C, and E (with a = 2) is the child of B.

Give the order of expansion of nodes and show the next two positions of the tree. Also mark the costs (a+b). Write pseudocode for the algorithm. (2+3)

4. If the above variant solves any deficiencies of greedy best first search algorithm. If yes, how? Give reasons for your answer. (2)

5. What is a heuristic? How it is important in a search process? Given is a graph with the true costs between nodes shown, and below are two possible heuristics h_1 and h_2.



$$
\begin{aligned}
h_1(A) &= 200 & h_2(A) &= 205\\
h_1(B) &= 247 & h_2(B) &= 270\\
h_1(D) &= 162 & h_2(D) &= 175\\
h_1(E) &= 72 & h_2(E) &= 82\\
h_1(F) &= 0 & h_2(F) &= 0
\end{aligned}
$$

Indicate if both/any of these is/are admissible heuristics. If yes, which one(s). Give reasons and explain why?

(3 + 2)

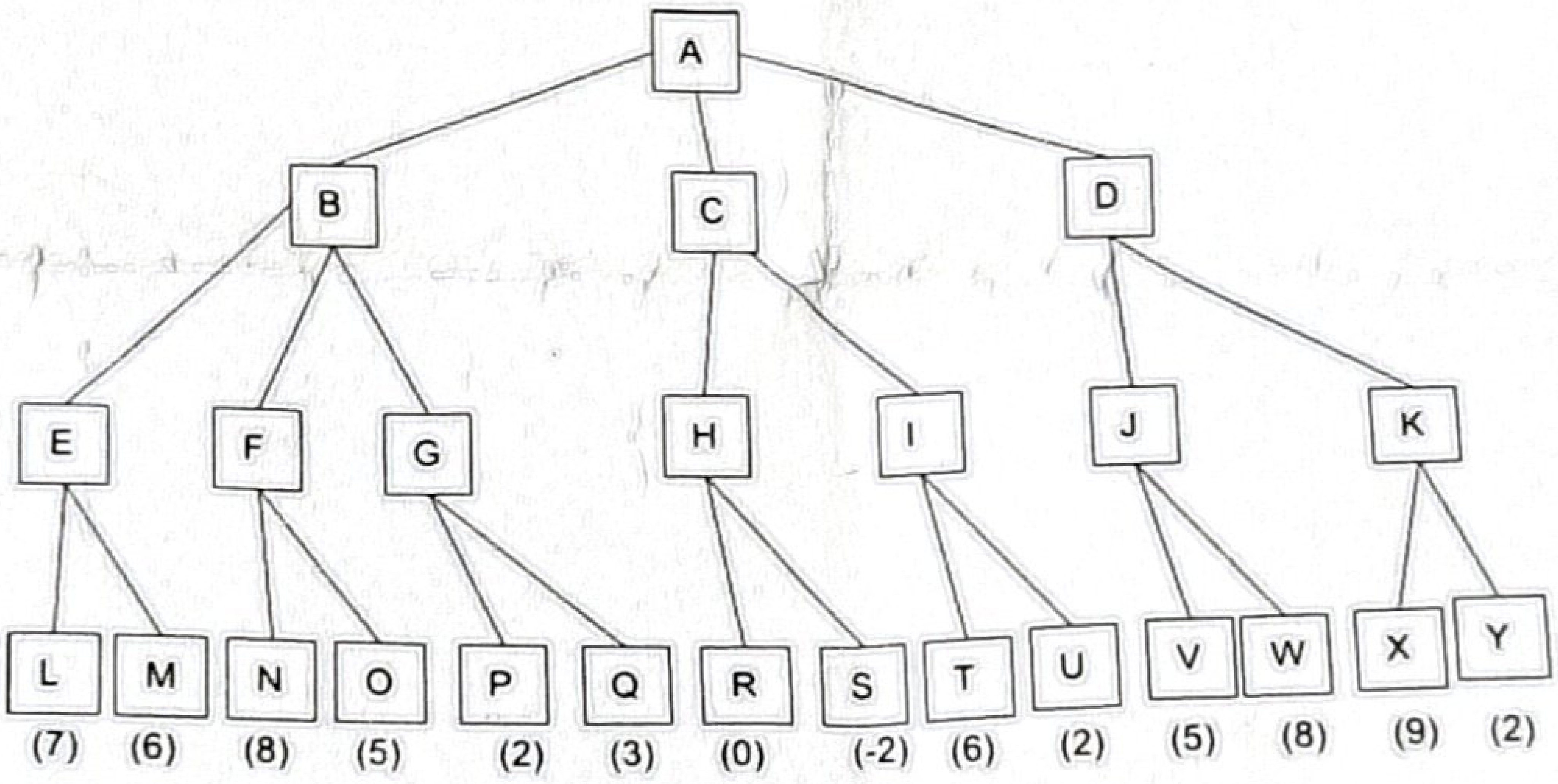
MINOR-2 EXAMINATION
ARTIFICIAL INTELLIGENCE

Max Duration: 1 hour **Max Marks: 20**

Instructions:

- This is a closed book exam.
- Ensure neatness, crispness, and relevance.
- In case you need to make any assumption(s), clearly write the same along with the reasons.

Q. 1 Consider the following game tree with the given static scores that are from the first player's point of view. The first player is the maximizing player, while the opponent is the minimizing one.



Give the move that should be chosen by the first player. Show the steps to reach your answer. (1 + 3)

Q.2 Write pseudo-code/steps for implementing alpha-beta search procedure. Apply your steps/pseudo-code to prune the tree as shown in the Question 1. Show the application of steps and the pruned tree. (2 + 3)

Q3. Consider the following cryptarithmetic problem.

$$\begin{array}{r} \text{C R O S S} \\ +\ \text{R O A D S} \\ \hline \text{D A N G E R} \\ \hline \end{array}$$

Contd...

Trace the constraint satisfaction problem answering the following:-

Describe the variables, their domains and the constraints. (2)

Show the constraint graph. (2)

Solve the puzzle as a constraint satisfaction problem. (2)

Q4. How Semantic Tableaux system operates to establish consistency (or inconsistency) of a formula in Propositional logic.

Check if

$\alpha : (P \wedge Q \rightarrow R) \wedge (\sim P \rightarrow S) \wedge Q \wedge \sim R \wedge \sim S$

is consistent using the Semantic Tableaux method. (2 + 3)