CSCI-570 Spring 2023

Practice Midterm 2

INSTRUCTIONS

- The duration of the exam is 140 minutes, closed book and notes.
- No space other than the pages on the exam booklet will be scanned for grading! Do not write your solutions on the back of the pages.
- If you require an additional page for a question, you can use the extra page provided within this booklet. However please indicate clearly that you are continuing the solution on the additional page.

1. True/False Questions

a) (T/F) There is a feasible circulation with demands d_v if $\sum_v d_v = 0$.

False. That is the necessary condition.

b) (T/F) If all capacities in a flow network are integers, then every maximum flow in the network is such that flow value on each edge is an integer.

False. Provide a counter example.

c) (T/F) If you are given a maximum $s-t$ flow in a graph $G = (V, E)$ then you can find a minimum $s - t$ cut in time $O(E)$ where E is the number of the edges in the graph.

False: knowing the max flow value does not help us to find a min cut. The inverse is true.

d) (T/F) The optimization version of the Linear Programming is not in NP.

True. The optimization version of linear programming (LP) is not in NP (nondeterministic polynomial time) but in P (polynomial time) because it can be solved in polynomial time by using algorithms such as the simplex algorithm.
The above explanation is not right. The LP optimization problem is not in NP because we cannot verify its solution in polynomial time

e) (T/F) Maximize

$$\sum_{1 \le i \le n} x_i(1 - x_i)$$

subject to

$$x_i + x_j < 1$$

and

$$x_i \in \{0, 1\}$$

where $1 \le i \le n, 1 \le j \le m$ and n and m are constants, is an integer linear program

True. The objective function is a linear expression of binary variables and is maximized.

f) (T/F) Every optimization problem has an equivalent decision problem.

True: We can always ask if the optimum value is greater/lesser than some estimate.

g) (T/F) If problem A can be reduced to problem B and B can be solved in polynomial time, then so can be A.

False: The reduction may not be done in polynomial time. eg. CNF to DNF

h) (T/F) All the NP-hard problems are in NP.

False

i) (T/F) If SAT $\leq_P$ A, then A is NP-hard.

True

j) (T/F) Every problem in NP can be solved in polynomial time by a nondeterministic Turing machine.

True: this is the definition of NP

2. Multiple Choice Questions

a) What is the role of the objective function in a linear programming problem?
   a) To define the feasible region
   b) To set the constraints
   c) To measure the profit or cost of a decision
   d) To set the decision variables
   c. The goal of the problem is to find the optimal values of the decision variables that satisfy the constraints and optimize the objective function.

b) What is the relationship between NP-hard and NP-complete problems?
   a) All NP-complete problems are also NP-hard
   b) All NP-hard problems are also NP-complete
   c) NP-hard problems are a subset of NP-complete problems
   d) NP-hard problems are not related to NP-complete problems
   a: All NP-complete problems are also NP-hard

c) Assuming P $\neq$ NP, which of the following is true?
   a) NP-complete = NP
   b) NP-complete $\cap$ P $\neq \emptyset$
   c) NP-hard = NP
   d) P = NP-complete
   b: see lecture 11, slide 34

d) Let X be an NP-complete problem and Q and R be two other problems not known to be in NP. Q is polynomial time reducible to X and X is polynomial-time reducible to R. Which one of the following statements is true?
   a) Q is NP-complete
   b) R is NP-complete
   c) R is NP-hard
   d) Q is NP-hard
   c: X $\leq_p$ R, but R is not in NP

e) Which of the following statements is correct ?
   a) Every LP problem has at least one optimal solution.
   b) Every LP problem has a unique solution.

c) If an LP problem has two optimal solutions, then it has infinitely many solutions.
d) If a feasible region is unbounded then LP problem has no solution

c

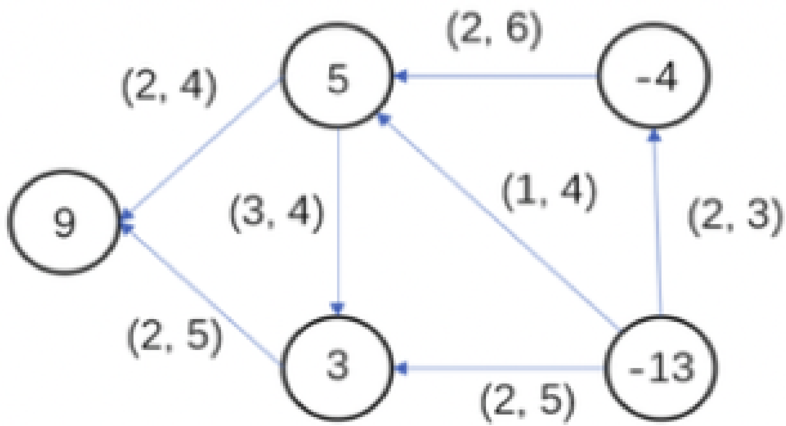

Figure 0.1: Question 3

3. Network Flow

Given the network below with the demand values on vertices and lower bounds on edge capacities, determine if there is a feasible circulation in this graph.

a) Turn the circulation with lower bounds problem into a circulation problem without lower bounds

b) Turn the circulation with demands problem into the max-flow problem

c) Does a feasible circulation exist? Explain your answer.

solution: Yes feasible circulation exits if and only if maxflow = 4+7+8+4 = 23.

4. Duality

Consider the following linear programming problem:

$$\max(3x_1 + 8x_2)$$

subject to:

$$\begin{aligned} x_1 + 4x_2 &\leq 20 \\ x_1 + x_2 &\geq 7 \\ x_1 &\geq -1 \\ x_2 &\leq 5 \end{aligned}$$

Write the dual associated to the above problem.

solution:

$$\min(y_1 - 3y_2)$$

subject to:

$$\begin{aligned} y_1 - y_2 &\geq 3 \\ -4y_1 + y_2 &\geq -8 \\ y_1, y_2 &\geq 0 \end{aligned}$$

5. **Linear Programming**

A paper company is building warehouses to supply its $m \geq 1$ customers. The company is considering whether to build a warehouse at each of $n \geq 1$ potential construction sites. The cost of building a warehouse at site i is denoted by $b_i \geq 0$. After the warehouses are built, the company assigns each customer to a warehouse that will supply it. Each warehouse that is built may supply paper to any number of customers, but each customer is supplied by exactly one warehouse. The cost of supplying customer j from a warehouse built at site i is denoted by $s_{ij} \geq 0$. Find a subset of construction sites and an assignment of customers to warehouses such that the total cost of building a warehouse at each selected site and supplying each customer from their assigned warehouse is minimized. Formulate this problem as a LP problem

a) Describe what your LP variables represent.

b) Show your objective function.

c) Show your constraints.

a) Let the variables $x_i \epsilon \{0, 1\} (1 \leq i \leq n)$ indicate whether construction site i was selected, and let the variables $y_{ij} \epsilon \{0, 1\} (1 \leq i \leq n, 1 \leq j \leq m)$ indicate whether a warehouse built at site i supplies customer j.

b) Min $C_B + C_S$ s.t.

i. $C_B = \sum_{1 \leq i \leq n} x_i b_i$ (C_B is the total cost of building the warehouses)

ii. $C_S = \sum_{1 \leq i \leq n} \sum_{1 \leq j \leq m} y_{ij} s_i$ (C_S is the total cost of supplying the customers.)

c) The constraints are as follows

i. $\sum_{1 \leq i \leq n} y_{ij} = 1, \quad 1 \leq j \leq m$ (Each customer is supplied by exactly one warehouse.)

ii. $x_i - y_{ij} \geq 0, \quad 1 \leq i \leq n \;\; \& \;\; 1 \leq j \leq m$ (A warehouse built at site i may only supply customer j if site i was selected.)

iii. $x_i \epsilon \{0, 1\}, \quad 1 \leq i \leq n$ (Construction site i is either selected or not.)

iv. $y_{ij} \epsilon \{0, 1\} \quad 1 \leq i \leq n \;\; \& \;\; 1 \leq j \leq m$ (A warehouse at site i supplies customer j, or does not.)

6. NP Completeness

Let there be a set ground set G and $S_i \subset G$, $i = 1...n$ be subsets of G. The hitting set problem is defined as finding a set $H \subset G$ of size at-most k such that it intersects all the sets S_i. Prove that this problem is NP complete by reducing from 3SAT/Vertex Cover.

First we want to show that the problem is in NP by verifying a given solution in polynomial time. Clearly, given a solution H, we can check if $|H| = k$ in linear time. Now we can also check if all the subsets S_i intersect with the given set H in at-most O(k l n) where $l = \max_i |S_i|$ which is clearly polynomial time.

Reduction from 3SAT: Let the ground set G contain all the literals (total 2k, if there are k variables) and their negations, i.e

$$G = \{x_1, x_2.....x_k, \neg x_1, \neg x_2, ...\neg x_k\}$$

Now for the i^{th} clause in 3SAT, we create a subset S_i out of the literals. For example if the 11th clause is $x_2 \vee x_6 \vee \neg x_9$, then our set S_{11} is $\{x_2, x_6, \neg x_9\}$. In addition to these sets, we also define k sets of the form $T_i = \{x_i, \neg x_i\}$ for $i = 1, 2....k$. Clearly all of this can be done in polynomial time.

We claim that if there is a set $H \subset G$ of size k such that it hits all the sets S_i and T_i, then we can extract a solution for the 3SAT.

Since there is a hitting set H of size k, it must be true that we select one of x_i or $\neg x_i$ from the set T_i. Now, if $x_i \in H$, we set x_i to true else we set it to false. This means all the elements in H evaluate to true and since each of S_i is hit by H, at least one of the term in the i^{th} clause evaluates to true and hence all the clauses are satisfied. Hence we have $3SAT <_p$ hitting set and if we can solve the hitting set problem, we can solve the 3SAT problem too. Hence the hitting set problem must be at least as hard as 3SAT and hence is NP Hard.

Reduction from Vertex Cover: Let there be a graph (V, E) for which we want a vertex cover. We construct the ground set as $G = V$ and for the ith edge $e_i = (u, v) \in E$, we introduce $S_i = \{u, v\}$. Finding a hitting set of size k provides us the vertices which form the vertex cover.

7. Approximation Algorithm

There are k restaurants and each restaurant receives various numbers of customer ratings, ranging from 1 star to 5 stars. Specifically, the i-th restaurant get n_{ij} ratings of j star, $1 \leq j \leq 5$. The average rating of the i-th restaurant is

$$\frac{\sum_{j=1}^{5} j \cdot n_{ij}}{\sum_{j=1}^{5} n_{ij}}$$

and it is guaranteed that $\sum_{j=1}^{5} n_{ij} > 0$. To find the highest rating among all the restaurants, we consider the following approximation algorithms

a) Prove that randomly choosing 1 restaurant is a 1/5-approximation algorithm.

b) Prove that choosing $\arg\max_{1 \leq i \leq k} \frac{n_{i5}}{\sum_{j=1}^{5} n_{ij}}$, a restaurant with the greatest portion of 5-star ratings is a 1/4-approximation algorithm.

c) Similarly, we can choose $\arg\min_{1 \leq i \leq k} \frac{n_{i1}}{\sum_{j=1}^{5} n_{ij}}$, a restaurant with the *least* portion of 1-star ratings. Does this method produce a better approximation? If so, prove that this is a ρ-approximation for some $\rho > 1/4$. Otherwise, find an example showing that the output of this method is a β-approximation for some $\beta \leq 1/4$.

a) Let OPT be the highest rating and r be the ratings of the chosen restaurant. Since $1 \le r \le \text{OPT} \le 5$, we have

$$\frac{r}{\text{OPT}} \ge \frac{r}{5} \ge \frac{1}{5}$$

b) Let $x = \max_{1 \le i \le k} \frac{n_{i5}}{\sum_{j=1}^{5} n_{ij}}$ and x^* be the portion of 5-star of the restaurant with the highest ratings.

$$\begin{aligned}
\frac{r}{\text{OPT}} &\ge \frac{r}{5x^* + 4(1 - x^*)} && \text{(upper bound OPT)} \\
&\ge \frac{5x + (1 - x)}{5x^* + 4(1 - x^*)} && \text{(lower bound } r\text{)} \\
&= \frac{4x + 1}{x^* + 4} \\
&\ge \frac{4x^* + 1}{x^* + 4} && (x^* \le x) \\
&= \frac{1}{4} + \frac{15x^*/4}{x^* + 4} \ge \frac{1}{4}
\end{aligned}$$

c) We show that $\rho = \frac{2}{5}$. Let $x = \min_{1 \le i \le k} \frac{n_{i1}}{\sum_{j=1}^{5} n_{ij}}$ and x^* be the portion of 1-star of the restaurant with the highest ratings.

$$\begin{aligned}
\frac{r}{\text{OPT}} &\ge \frac{r}{x^* + 5(1 - x^*)} && \text{(upper bound OPT)} \\
&\ge \frac{x + 2(1 - x)}{x^* + 5(1 - x^*)} && \text{(lower bound } r\text{)} \\
&= \frac{2 - x}{5 - 4x^*} \\
&\ge \frac{2 - x}{5 - 4x} && (x^* \ge x) \\
&= \frac{2}{5} + \frac{0.6x}{5 - 4x} \ge \frac{2}{5}
\end{aligned}$$