NAME: ________________________

ISC 4221: Discrete Algorithms

December 12th, 2023

7:30am–9:30am

Final Exam

INSTRUCTIONS

- By printing your name above, you agree to abide by the Honor Code.
- Turn cell-phones off.
- You may not communicate or collaborate with any other person (talk, email, chat, etc.) throughout the duration of this exam.
- This is an open notes, open internet, open "everything" exam.
- For every answer submit your code and provide an explanation of your answer. No credit will be given for an answer without description.
- You can submit some/part of your answers in paper and your code and the rest of the answers through github.

SCORE

Question	Points	Max
1		25
2		25
3		25
4		25
Total		100

Question 1: Lines and Planes

For this problem, consider the line passing through the two points

$$p_1 = (2, 3, 5) \quad \text{and} \quad p_2 = (4, 7, 8).$$

(a) Write a formula for a point $p(s)$ on the line.

(b) Use the dot product method to compute the s coordinate of the point $p = (5, 9, 9.5)$ which is on the line.

(c) Find the point p on the line that is closest to $q = (0, 2, 1)$.

(d) Compute how far q is from the line.

(e) The two-dimensional analog of a line is the plane. An example of a plane in three-dimensional space is the set of points (x,y,z) such that

$$(n_1, n_2, n_3) \cdot (x, y, z) = 0,$$

where (n_1, n_2, n_3) is the vector that is perpendicular to the plane, and the "·" is the dot product. Find any two points on the plane when $(n_1, n_2, n_3) = (-1, 2, -3)$.

Question 2: Decision Tree

Suppose you are making a decision tree to predict whether a driver will file an insurance claim in the next year. You hypothesize that the most important attributes are:

- The number of years of driving experience;
- Whether the vehicle is classified as a sports car;
- The annual income.

The 10 records you collect are the following.

Record	Years Driving	Sports Car	Income	Claim
1	2	No	25K	Yes
2	5	Yes	50K	Yes
3	20	No	85K	No
4	15	No	90K	No
5	40	No	45K	No
6	25	Yes	95K	Yes
7	1	No	120K	No
8	30	No	100K	No
9	32	No	75K	No
10	4	Yes	30K	Yes

Suppose you decide to make all the attributes binary by dividing the *Years Driving* attribute into less than 10 years and greater than 10 years, and dividing the *Income* attribute into less than 80K and more than 80K.

(a) Compute the impurity of the root node with the classification error of a node t

$$E(t) = 1 - \max_{1 \le i \le k} p(i \mid t).$$

(b) A table that gives the split for the child nodes corresponding to the *Years Driving* attribute is below. Construct similar tables for the other two attributes.

Years Driving	Claim	No Claim
< 10	3	1
≥ 10	1	5

(c) Use the classification error to compute the impurity of **all six** child nodes of the root node. To do this for the *Years Driving* attribute, you need to compute

$$p(\texttt{Claim} \mid < 10), \quad p(\texttt{No Claim} \mid < 10), \quad p(\texttt{Claim} \mid \geq 10), \quad p(\texttt{No Claim} \mid \geq 10),$$

and then use these results to compute the impurities.

$$E(< 10) \quad \text{and} \quad E(\geq 10).$$

(d) Use the impurity measurements to compute the gain of the three possible splits.

(e) According to Hunt's Algorithm, what split should first be applied to the root node?

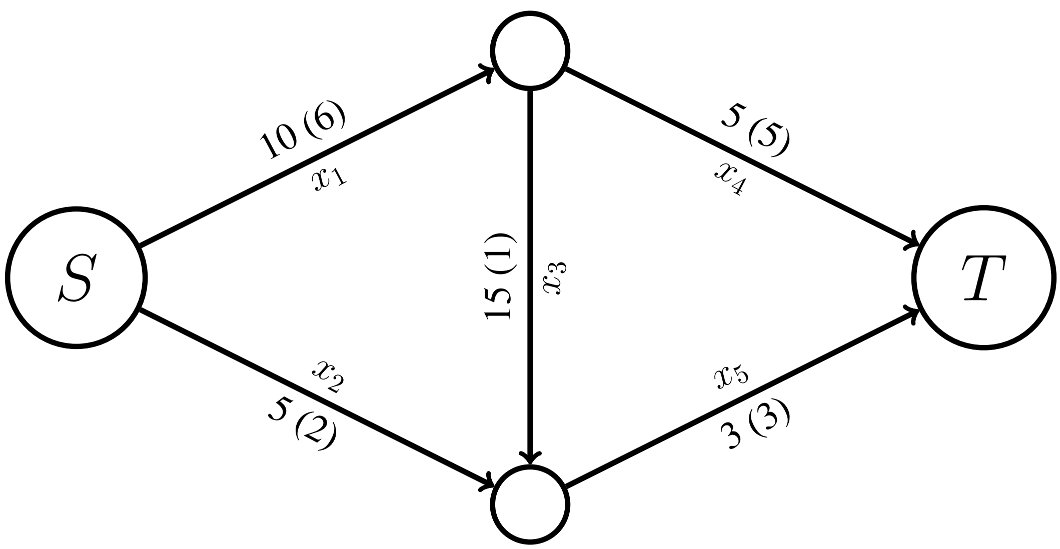
Question 3: Linear Programming

Consider a weighted directed graph $G = (V, E)$, where V are the vertices and E are the edges with weights that correspond to their capacity. Therefore, the weights are a limitation of some quantity, such as a fluid or electrical current, that is flowing between adjacent vertices, which might represent houses or cities. Moreover, flow only goes in the direction indicated by the directed edge. Finally, one vertex, called the source and labelled S, only has outgoing edges, and another vertex, called the target and labelled T, only has incoming edges.

The goal is to decide how much flow to put through each of the edges to maximize the total amount of flow coming out of the source S. The two rules for the amount of flow through each edge are:

- The flow through an edge can not exceed the edge's capacity;
- For every vertex except the source and target, the net flow must be zero. That is, the amount of inflow must equal the amount of outflow.

In the following example, the maximum flow of each edge is the value not in parentheses. One configuration that maximizes the flow out of the source S is given by the parenthetic values.



(a) Considering only integer-valued flow rates (ie. no fractions), list the four flow configurations that result in a maximum flow of 8. I have provided four blank templates for you to fill in. The figure above is one of the four configurations.

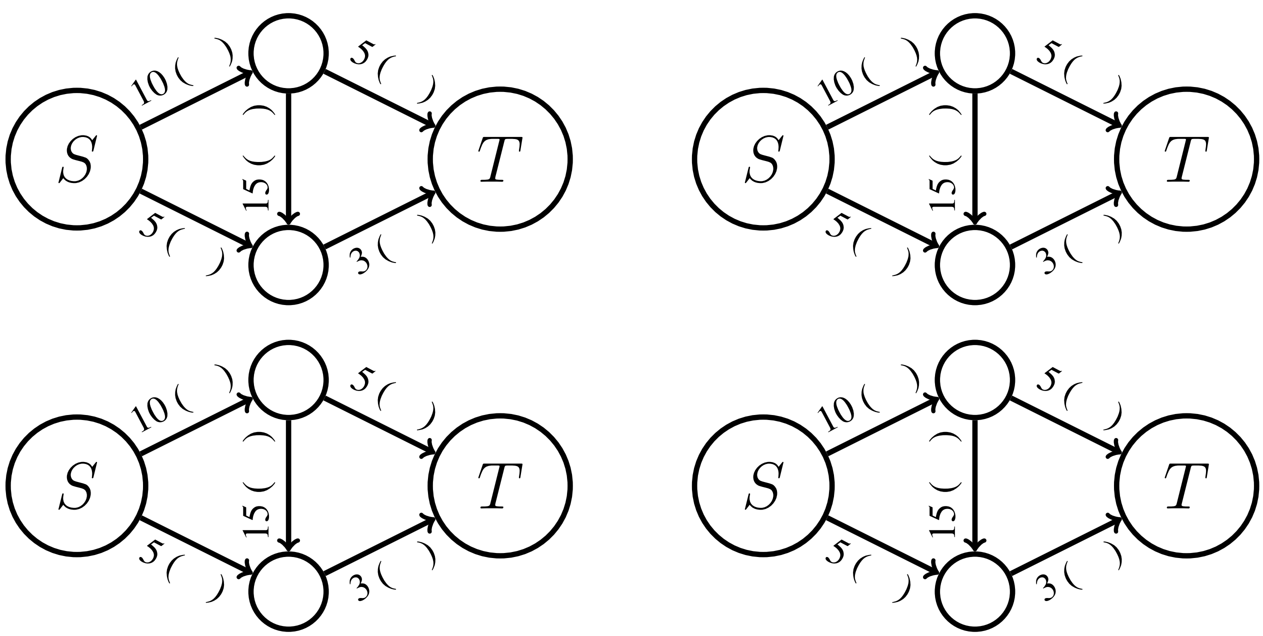
(b) Explain how the weights of the edges starting at S, and how the weights of the edges ending at T can be used to provide an upper bound (ie. a limit) for the maximum flow.

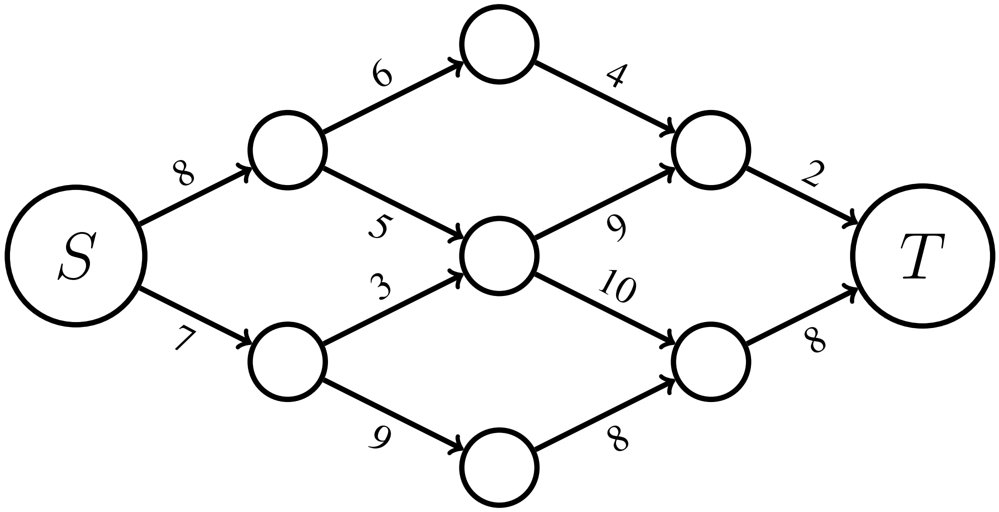
(c) As labelled above, let x_i, $i = 1, \ldots, 5$, be the amount of flow through each of the five edges. Then the flow optimization problem amounts to maximizing the linear function $x_1 + x_2$. The constraints can be written as a set of linear equalities and inequalities, and two of them are

$$0 \leq x_3 \leq 15,$$
$$x_1 - x_3 - x_4 = 0.$$

Write all the other linear constraints required to write this maximum flow rate problem as a linear program. You do not need to solve the linear program.

(d) For the nine-vertex flow optimization problem illustrated below, write the objective function, 3 of the inequality constraints, and 3 of the equality constraints. Be sure to define x_i, $i = 1, \ldots, 12$.





S
T
8
7
6
5
3
9
4
9
10
8
2
8

Question 5: Image Manipulation

For this problem, consider the black and white image represented with the unsigned 4-bit (max value is 15) integer matrix

4	4	5	6	8	10	10
4	5	5	15	0	10	11
4	0	6	7	9	0	11
5	0	15	6	8	15	10
4	6	4	6	9	0	11
5	7	4	7	8	10	12

(a) This image contains some salt and pepper noise. Apply the denoising algorithm we introduced in class where each pixel is replaced with the median of its nine neighbors. Do not apply the algorithm to the boundary pixels. I have started the algorithm.

4	4	5	6	8	10	10
4	4	5	6			11
4						11
5						10
4						11
5	7	4	7	8	10	12

(b) Enhance the contrast of the result from part (a) using an s value of 2. Use only the 4 adjacent pixels to compute the necessary average. Be sure to convert the result back to an unsigned 4-bit integer. Do not apply the algorithm to the boundary pixels.

(c) Compute the edge statistic E of the result from part (b). Do not apply the algorithmn to the boundary pixels.