Worth: 10%

1. **[20 marks]**

   A hospital is scheduling doctors for work. The hospital has a desired number of doctors $a_1,\ldots,a_n$ that they would like to work on each of the next n days. Furthermore, each doctor who works at the hospital has a list $A_j \subseteq \{1,\ldots,n\}$ of days when they would like to work.

   (a) **[5 marks]**
   You are given the numbers $a_1,\ldots,a_n$ and the lists $A_1,\ldots,A_k$ of the doctors' availabilities. Using this data, design an algorithm using network flow to output a schedule $S_1,\ldots,S_k$ for each doctor such that
   - Each doctor works only when they are available (i.e. $S_i \subseteq A_i$), and
   - Exactly a_i doctors work on day i

   or the algorithm reports that no schedule is possible given the doctors' availabilities.

   (b) **[5 marks]**
   Briefly justify correctness and runtime of the algorithm designed in part (a).

   (c) **[5 marks]**
   Due to COVID-19 scenario, the hospital finds that it may be necessary to schedule doctors for times outside when they report they are available to work. They decide that doctors can be scheduled for up to c days outside their availability list, for some integer $c > 0$. Modify the algorithm designed in part (a) so a schedule $S_1,\ldots,S_k$ is produced with:
   - S_i containing at most c days not in A_i, and
   - Exactly a_i doctors work on day i

   or the algorithm reports that no such schedule is possible.

   (d) **[5 marks]**
   Briefly justify correctness and runtime of the modified algorithm designed in part (c).

2. **[20 marks]**

   An edge in a flow network is called **critical** if decreasing the capacity of this edge reduces the maximum possible flow in the network.

   Give an efficient algorithm that finds a critical edge in a network. Give a rigorous argument that your algorithm is correct and analyse its running time.

3. **[20 marks]**

   (a) **[8 marks]**
   Suppose we want to compute a shortest path from node s to node t in a *directed* graph $G = (V, E)$ with integer edge weights $\ell_e > 0$ for each $e \in E$.
   Show that this is equivalent to finding a *pseudo-flow* f from s to t in G such that $|f| = 1$ and $\sum_{e \in E} \ell_e f(e)$ is minimized. There are no capacity constraints.
   Part of this problem requires you to **define** precisely what we mean by "pseudo-flow" in a general, directed graph. This is a natural extension of the notion of flow in a network.

   (b) **[12 marks]**
   Write the shortest path problem as a linear or integer program **where your objective function is *minimized*, based on your answer to the previous part**. Give a detailed justification that your solution is correct.

4. **[20 marks]**

Linear programming is often used to solve statistical and machine learning problems. We consider two examples here. We are given n data points (x_i, y_i) and wish to find a line of best fit $y = ax + b$ for some coefficients a, b that best approximates the behaviour of the data points. The ℓ_1 distance between a point (x_i, y_i) and a line $y = ax + b$ is defined as the quantity $|y_i - ax_i - b|$.

(a) **[10 marks]**

Suppose we want to minimize the ℓ_1 error for the line of best fit, defined as minimum of the sum of the ℓ_1 distances over the set of data points. Give a linear program that produces a line of best fit with minimum ℓ_1 error.

(b) **[10 marks]**

In the *classification* problem, you are given m data points of type 1 and n data points of type 2. We say that a line $y = ax + b$ *separates* the points if all data points of type 1 satisfy $y_i < ax_i + b$ and all data points of type 2 satisfy $y_i > ax_i + b$. Give a linear program that determines if there is a line that separates the points, and if one exists, maximizes the gap δ defined as $\min(e_1, e_2)$, where e_i is the minimum ℓ_1 distance between the line and points of type i.