CS 473: Algorithms, Fall 2016
HW 9 (due Tuesday, November 15 at 8pm)

This homework contains three problems. **Read the instructions for submitting homework on the course webpage**.

Collaboration Policy: For this home work, each student can work in a group with up to three members. Only one solution for each group needs to be submitted. Follow the submission instructions carefully.

For problems that ask for a linear-programming formulation of some problem, a full credit solution requires the following components:

- A list of variables, along with a brief English description of each variable. (Omitting these English descriptions is a Deadly Sin.)
- A linear objective function (expressed either as minimization or maximization, whichever is more convenient), along with a brief English description of its meaning.
- A sequence of linear inequalities (expressed using $\leq$, $=$, or $\geq$, whichever is more appropriate or convenient), along with a brief English description of each constraint.
- A proof that your linear programming formulation is correct, meaning that the optimal solution to the original problem can always be obtained from the optimal solution to the linear program. This may be very short.

It is ***not*** necessary to express the linear program in canonical form, or even in matrix form. Clarity is much more important than formality.

For problems that ask to prove that a given problem X is NP-hard, a full-credit solution requires the following components:

- Specify a known NP-hard problem Y, taken from the problems listed in the notes.
- Describe a polynomial-time algorithm for Y, using a black-box polynomial-time algorithm for X as a subroutine. Most NP-hardness reductions have the following form: Given an arbitrary instance of Y, describe how to transform it into an instance of X, pass this instance to a black-box algorithm for X, and finally, describe how to transform the output of the black-box subroutine to the final output. A cartoon with boxes may be helpful.
- Prove that your reduction is correct. As usual, correctness proofs for NP-hardness reductions usually have two components (one for each f).

---

1. Given points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ in the plane the *linear regression problem* asks for real numbers a and b such that the line $y = ax + b$ fits the points as closely as possible according to some criterion. The most common fit criterion is minimizing the L_2 error, defined as follows:

$$\epsilon_2(a, b) = \sum_{i=1}^{n} (y_i - ax_i - b_i)^2.$$

   But there are several other fit criteria, some of which can be optimized by linear programming.

   - The L_1 error of the line $y = ax + b$ is defined as follows

   $$\epsilon_1(a, b) = \sum_{i=1}^{n} |y_i - ax_i - b|.$$

     Describe a linear program whose solution (a, b) describes the line with the minimum L_1 error.

   - The L_∞ error of the line $y = ax + b$ is defined as follows.

   $$\epsilon_\infty(a, b) = \max_{i=1}^{n} |y_i - ax_i - b|.$$

     Describe a linear program whose solution (a, b) describes the line with the minimum L_∞ error.

   Comment: In general the points can be in $\mathbb{R}^d$ for some d and in that case one fits a hyperplane. We are here considering the simple case of $d = 2$.

2. The facility location problem is the following. There is a set of n facilities F and m clients C. The cost of openining a facility i is f_i. There is a cost $c(i, j)$ to connect client j to facility i. The goal is to open a susbet of the facilities and connect each client to an open facility. The objective function is to minimize the sum of the facility opening costs and the connection costs. Write an integer linear programming formulation for this problem using two sets of decision variables, one set for opening facilities, and one set for assigning clients to open facilities. Prove that it is sufficient to constrain only the facility opening variables to be integer valued. Such a problem is called a mixed-integer programming problem.

3. A directed graph $G = (V, E)$ is *strongly connected* if for all $u, v \in V$ there is a path in G from u to v and from v to u. The STRONGLY-CONNECTED SPANNING SUBGRAPH problem is the following: given a directed graph $G = (V, E)$ and an integer k is there a spanning subgraph H of G with at most k edges such that H is strongly connected? A subgraph is spanning if it contains all the vertices of the original graph. Describe a reduction to show that STRONGLY-CONNECTED SPANNING SUBGRAPH is NP-Complete.

The remaining problems are for self study. Do *NOT* submit for grading.

- Suppose we have a linear program $\max cx, A_1x \leq b_1, A_2x = b_2, A_3x \geq b_3, x \geq 0$. What is its dual?
- See Problem 3 in HW 8 from Jeff's home work last spring. `https://courses.engr.illinois.edu/cs473/sp2016/hw/hw8.pdf`
- See HW 5 from Chandra's graduate algorithms course in Fall 2011. `https://courses.engr.illinois.edu/cs573/fa2011/Homework/hw5.pdf`
- See HW 8 from Sariel's course in Fall 2015. `https://courses.engr.illinois.edu/cs473/fa2015/w/hw/hw_08.pdf`.