Homework 11 – Due Thursday, December 1, 2016 on Canvas

Please refer to HW guidelines from HW1, course syllabus, and collaboration policy.

Problems to be handed in

1. (**LP formulation of linear regression**) 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:

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

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

   (a) The L_1 error (or total absolute deviation) of the line $y = ax + b$ is defined as follows:

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

   Give an LP whose solution (a, b) describes the line with minimum L_1 error.

   (b) The L_∞ error (or maximum absolute deviation) of the line $y = ax + b$ is defined as follows:

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

   Give an LP whose solution (a, b) describes the line with minimum L_∞ error.

2. (**Integer Programming**) An integer program is a linear program with the additional constraint that the variables must take only integer values.

   (a) Show that weak duality holds for an integer linear program.

   (b) Show that duality does not always hold for an integer linear program.

   (c) Given an LP, let P be its optimal objective value, D be the optimal objective value for the dual, IP be the optimal objective value of the integer version of the primal (that is, the primal with the added constraint that the variables take on integer values), and ID be the optimal objective value for the integer version of the dual. Assuming that both the primal integer program and the dual integer program are feasible and bounded, show that

$$IP \leq P = D \leq ID.$$

   (d) Prove that deciding whether an integer program has a feasible solution is NP-complete.

   (e) Next you will show that the problem of finding the optimal feasible solution for an integer program is hard even if it is known that a feasible solution exists. First, formulate a related decision problem. The input to your problem should be an integer program, a feasible solution to the program and a real value (purported value of the objective function). Prove that your problem is NP-complete.

Hint: Almost any NP-hard decision problem can be formulated as an integer program. Pick your favorite.

3. (**Search vs. Decision Problems**) Consider the (decision version of the) SET COVER problem defined on p. 456 of KT. Let SC-SEARCH be the search version of the problem, where the input is a set U of n elements and a collection $S_1, \ldots, S_m$ of subsets of U, and the goal is to output the smallest set cover, i.e., a collection with the minimum number of sets, such that the union of these sets is equal to all of U. Show that

$$\text{SC-SEARCH} \leq_{p,Cook} \text{SET COVER}.$$

   In other words, show how to solve SC-SEARCH in polynomial time, given an oracle that solves SET COVER. (*Hint:* First, figure out the size of the smallest set cover and then consider one set at a time and decide whether to include it.) Analyze the running time of your algorithm, its space complexity and the number of calls to the oracle.