**Worth:** 10% **Due:** Thursday August 1, 11:59pm.

**Remember to write your *full name*, and *student number*.**

*Please make sure that the work you submit does not contain someone else's words or ideas. Then, to protect yourself, list on the front of your submission* ***every*** *source of information you used to complete this homework (other than your own lecture and tutorial notes, and materials available directly on the course webpage). For example, indicate clearly the* ***name*** *of every student with whom you had discussions, the* ***title*** *of every additional textbook you consulted, the* ***source*** *of every additional web document you used, etc.*

---

For each question, please write up detailed answers carefully. Make sure that you use notation and terminology correctly, and that you explain and justify what you are doing. Marks **will** be deducted for incorrect or ambiguous use of notation and terminology, and for making incorrect, unjustified, ambiguous, or vague claims in your solutions. Note that, depending on the TA times, we may decide not to grade some questions.

In any of your answers you can use any algorithm we discussed in class without proving it solves the problem we discussed in class optimally. If we discussed the runtime of the algorithm you can also use that without reproving it. The same goes for any Lemma, Theorem or Fact we discussed in class.

Note that any algorithm you propose should have a polynomial run time (algorithms with exponential run time such as brute force do not earn any point).

You will receive 20% in each question (except the bonus question) if you leave it completely blank or say **"I don't know"**. The bonus question will be marked harshly.

---

[5] 1. Exercise 7.3 of the DPV textbook.

[9] 2. Exercise 7.7 of the DPV textbook.

[10] 3. Exercise 7.20 of the DPV textbook.

4. You are given $n$ points $(x_1, y_1), ..., (x_n, y_n)$ in the plane. The linear regression problem asks for a line $ax + by = c$ that fits the points as closely as possible according to some criterion (no line is a perfect fit).

[10] (a) Write a linear program whose solution $(a, b, c)$ describes the line that minimizes the maximum absolute error

$$\max_{1 \leqslant i \leqslant n} |ax_i + by_i - c|$$

[10] (b) Write a linear program whose solution $(a, b, c)$ describes the line that minimizes the $L_1$ error

$$\sum_{i=1}^{n} |ax_i + by_i - c|$$

5. For each of the following state if it is true or false and justify your answer:

[5] (a) If $D_1, D_2, D_3$ are decision problems, and $D_1 \leqslant_p D_2$, $D_2 \leqslant_p D_3$, and $D_3 \in$ P then it follows that $D_1 \in$ P.

[5] (b) If $D_1, D_2$ are decision problems, and $D_1 \leqslant_p D_2$ then it cannot be the case that $D_2 \leqslant_p D_1$.

[10] (c) In the 3SAT problem if each literal appears at most once, then the problem is solvable in polynomial time.

6. We want to show the following problems are NP-complete:

(a) We are given a set $A = \{a_1, ..., a_n\}$, a collection $S_1, S_2, ..., S_m$ of subsets of $A$, and a number $k$. Is there a set $B \subseteq A$ such that for each $S_i$, $S_i \cap B \neq \varnothing$ and the size of $B$ is at most $k$?

[5] i. Show that this decision problem is in NP.

[6] ii. Show that this decision problem is NP-hard.

(b) Consider a new version of Job scheduling problem. There are $n$ jobs $J_i = (a_i, d_i, p_i)$ where $a_i$, $d_i$, and $p_i$ are, respectively, the arrival time, the deadline, and the processing time of job $i$. Job $i$ should be scheduled to start not before its arrival time and finish not after its deadline (*i.e.*, $J_i$ must be scheduled to start at a time $t_i$ such that (1) $a_i \leqslant t_i$ and (2) $t_i + p_i \leqslant d_i$). Moreover, the scheduled jobs should not intersect (except at their end points). Is there a valid scheduling with a size of at least $k$ (*i.e.*, can we schedule at least $k$ jobs) ?

[5] i. Show that this decision problem is in NP.

[bonus] ii. Show that this decision problem is NP-hard.