EECS 127/227AT Optimization Models in Engineering
Spring 2020 **Homework 11**

This homework is due Friday, April 17, 2020 at 23:00 (11pm).
Self grades are due Friday, April 24, 2020 at 23:00 (11pm).

This version was compiled on 2020-04-19 15:16.

Submission Format: Your homework submission should consist of a single PDF file that contains all of your answers (any handwritten answers should be scanned) as well as your IPython notebook with solutions saved as a PDF.

1. **Median versus average**

For a given vector $\vec{v} \in \mathbb{R}^n$, the average can be found as the solution to the optimization problem

$$\min_{x \in \mathbb{R}} \|\vec{v} - x\vec{\mathbf{1}}\|_2^2, \tag{1}$$

where $\vec{\mathbf{1}}$ is the vector of ones in $\mathbb{R}^n$. Similarly, the median (any value x such that there is an equal number of values in v above or below x) can be found via

$$\min_{x \in \mathbb{R}} \|\vec{v} - x\vec{\mathbf{1}}\|_1. \tag{2}$$

We consider a robust version of the average problem (1):

$$\min_{x} \max_{\vec{u} : \|\vec{u}\|_\infty \leq \lambda} \|\vec{v} + \vec{u} - x\vec{\mathbf{1}}\|_2^2, \tag{3}$$

in which we assume that the components of v can be independently perturbed by a vector u whose magnitude is bounded by a given number $\lambda \geq 0$.

(a) Is the robust problem (3) convex? Justify your answer precisely, based on expression (3), and without further manipulation.

(b) Show that problem (3) can be expressed as

$$\min_{x \in \mathbb{R}} \sum_{i=1}^{n} \left(|v_i - x| + \lambda\right)^2.$$

(c) Express the problem as a QP. State precisely the variables, and constraints if any.

(d) Show that when λ is large, the solution set approaches that of the median problem (2).

(e) It is often said that the median is a more robust notion of “middle” value than the average, when noise is present in $\vec{v}$. Based on the previous part, justify this statement.

Solution:

(a) The robust problem is convex, since the objective function is the pointwise maximum (over $\vec{u}$) of convex functions, $x \to \|\vec{v} + \vec{u} - x\vec{\mathbf{1}}\|_2^2$.

(b) For any vector $\vec{z} \in \mathbb{R}^n$, we have

$$\begin{aligned}
\max_{\vec{u}:\|\vec{u}\|_\infty \leq \lambda} \|\vec{z}+\vec{u}\|_2^2 &= \max_{\vec{u}:\|\vec{u}\|_\infty \leq \lambda} \sum_{i=1}^{n} (z_i+u_i)^2 \\
&= \max_{|u_1|\leq\lambda, |u_2|\leq\lambda, \ldots, |u_n|\leq\lambda} \sum_{i=1}^{n} (z_i+u_i)^2 \\
&= \sum_{i=1}^{n} \max_{u_i \,:\, |u_i|\leq\lambda} (z_i+u_i)^2.
\end{aligned}$$

The last equality follows since the problem is decomposable and it is optimal to maximize each of these terms. Next we have,

$$\begin{aligned}
\sum_{i=1}^{n} \max_{u_i \,:\, |u_i|\leq\lambda} (z_i+u_i)^2 &= \sum_{i=1}^{n} \max_{\eta \,:\, |\eta|\leq\lambda} (z_i+\eta)^2 \\
&= \sum_{i=1}^{n} (|z_i|+\lambda)^2,
\end{aligned}$$

the last line resulting from

$$\forall\, \eta, \quad |\eta| \leq \lambda \;:\; |z_i+\eta| \leq |z_i| + \lambda,$$

with upper bound attained with $\eta = \lambda \text{sign}(z_i)$. Taking $\vec{z} = \vec{v} - x\vec{\mathbf{1}}$ we get the desired form.

(c) A QP formulation is

$$\min_{x,t} \sum_{i=1}^{n} (t_i+\lambda)^2 \;:\; t_i \geq \pm(v_i - x), \quad i = 1, \ldots, n.$$

(d) The objective function takes the form

$$\sum_{i=1}^{n} (|v_i - x| + \lambda)^2 = n\lambda^2 + 2\lambda\|\vec{v} - x\vec{\mathbf{1}}\|_1 + \|\vec{v} - x\vec{\mathbf{1}}\|_2^2,$$

The corresponding optimization problem has the same minimizers as the problem

$$\min_{x} \|\vec{v} - x\vec{\mathbf{1}}\|_1 + \frac{1}{2\lambda}\|\vec{v} - x\vec{\mathbf{1}}\|_2^2,$$

When λ is large, the minimizer will tend to minimize the first term only, which implies the desired result.

(e) The median problem can be interpreted as a robust version of the average problem, when the uncertainty is large.

2. LASSO vs. Ridge

Say that we have the data set $\{(\vec{x}^{(i)}, y^{(i)})\}_{i=1,\ldots,n}$ of features $\vec{x}^{(i)} \in \mathbb{R}^d$ and values $y^{(i)} \in \mathbb{R}$. Define $X = \begin{bmatrix} \vec{x}^{(1)} & \ldots & \vec{x}^{(n)} \end{bmatrix}^\top$ and $y = \begin{bmatrix} y^{(1)} & \ldots & y^{(n)} \end{bmatrix}^\top$. For the sake of simplicity, assume that the data has been centered and whitened so that each feature has mean 0 and variance 1 and the features are uncorrelated, i.e. $X^\top X = nI$.

Consider the linear least squares regression with regularization in the ℓ_1-norm, also known as LASSO:

$$\vec{w}^* = \arg\min_{\vec{w}\in\mathbb{R}^d} \|X\vec{w} - \vec{y}\|_2^2 + \lambda\|\vec{w}\|_1.$$

This problem will compare ℓ_1-regularization with ℓ_2-regularization (ridge regression) to understand their similarities and differences. We will do this by looking at the elements of $\vec{w}^*$ in the solution to each problem.

(a) First, we decompose this optimization problem into d univariate optimization problems over each element of $\vec{w}$. Let $X = \begin{bmatrix}\vec{x}_1 & \dots & \vec{x}_d\end{bmatrix}$ and recall that $X^\top X = nI$.

Solution:

$$\|X\vec{w} - \vec{y}\|_2^2 + \lambda\|\vec{w}\|_1 = \sum_{i=1}^{d}\left[nw_i^2 - 2\vec{y}^\top\vec{x}_i w_i + \lambda|w_i|\right] + \vec{y}^\top\vec{y}.$$

Hence the original problem becomes

$$\min_{\vec{w}\in\mathbb{R}^d}\sum_{i=1}^{d}\left[nw_i^2 - 2\vec{y}^\top\vec{x}_i w_i + \lambda|w_i|\right] + \vec{y}^\top\vec{y}.$$

Since the objective is separable in w_i the problem decomposes into d univariate optimization problems and hence we have

$$\sum_{i=1}^{d}\min_{w_i\in\mathbb{R}}\left[nw_i^2 - 2\vec{y}^\top\vec{x}_i w_i + \lambda|w_i|\right] + \vec{y}^\top\vec{y}.$$

(b) If $w_i^* > 0$, then what is the value of w_i^*? What is the condition on $\vec{y}^\top\vec{x}_i$ for this to be possible?

Solution: If $w_i^* > 0$, then the first order optimality conditions for w_i^* write

$$2nw_i^* - 2\vec{y}^\top\vec{x}_i + \lambda = 0,$$

from which we obtain

$$w_i^* = \frac{2\vec{y}^\top\vec{x}_i - \lambda}{2n},$$

which is positive when

$$\vec{y}^\top\vec{x}_i > \frac{\lambda}{2}.$$

(c) If $w_i^* < 0$, then what is the value of w_i^*? What is the condition on $\vec{y}^\top\vec{x}_i$ for this to be possible?

Solution: If $w_i^* < 0$, then the first order optimality conditions for w_i^* write

$$2nw_i^* - 2\vec{y}^\top\vec{x}_i - \lambda = 0,$$

from which we obtain

$$w_i^* = \frac{2\vec{y}^\top\vec{x}_i + \lambda}{2n},$$

which is negative when

$$\vec{y}^\top\vec{x}_i < -\frac{\lambda}{2}.$$

(d) What can we conclude about w_i^* if $|\vec{y}^\top \vec{x}_i| \le \dfrac{\lambda}{2}$? How does the value of λ impact the individual entries w_i^*?

Solution: From the previous parts we have $w_i^* \neq 0 \Rightarrow |\vec{y}^\top \vec{x}_i| > \dfrac{\lambda}{2}$.

Hence if $|\vec{y}^\top \vec{x}_i| \le \dfrac{\lambda}{2}$ then we must have $w_i^* = 0$. This means that a larger value of λ will force more entries of $\vec{w}$ to be zero — i.e. larger λ will imply higher sparsity.

(e) Now consider the case of ridge regression, which uses the the ℓ_2 regularization $\lambda\|\vec{w}\|_2^2$.

$$\vec{w}^* = \arg\min_{\vec{w}\in\mathbb{R}^d} \|X\vec{w} - \vec{y}\|_2^2 + \lambda\|\vec{w}\|_2^2.$$

Write down the new condition for $\vec{w}_i^*$ to be 0. How does this differ from the condition obtained in part (4) and what does this suggest about LASSO?

Solution: In the case of ridge regression the optimal weight vector $\vec{w}$ is given by

$$w_i^* = \frac{\vec{y}^\top \vec{x}_i}{n+\lambda},\ i = 1,\dots,d.$$

So w_i^* is only zero when $\vec{y}^\top \vec{x}_i = 0$, in contrast to LASSO where w_i^* is zero when $\vec{y}^\top \vec{x}_i \in \left[-\frac{\lambda}{2}, \frac{\lambda}{2}\right]$. This suggest that LASSO forces a lot of coordinates to be zero, i.e. induces sparsity to the optimal weight vector.

3. A slalom problem

A skier must slide from left to right by going through n parallel gates of known position (x_i, y_i) and width c_i, $i = 1,\dots,n$. The initial position (x_0, y_0) is given, as well as the final one, (x_{n+1}, y_{n+1}). Before reaching the final position, the skier must go through gate i by passing between the points $(x_i, y_i - c_i/2)$ and $(x_i, y_i + c_i/2)$ for each $i \in \{1,\dots,n\}$. Figure 1 is a representation. Use values for (x_i, y_i, c_i) from Table 1.

Table 1: Problem data for Problem 2.

i	x_i	y_i	c_i
0	0	4	N/A
1	4	5	3
2	8	4	2
3	12	6	2
4	16	5	1
5	20	7	2
6	24	4	N/A

(a) Given the data $\{(x_i, y_i, c_i)\}_{i=0}^{n+1}$, write an optimization problem that minimizes the total length of the path. Your answer should come in the form of an SOCP.

(b) Solve the problem numerically with the data given in Table 1.

Hint: You should be able to use packages such as `cvxpy` and `numpy`.

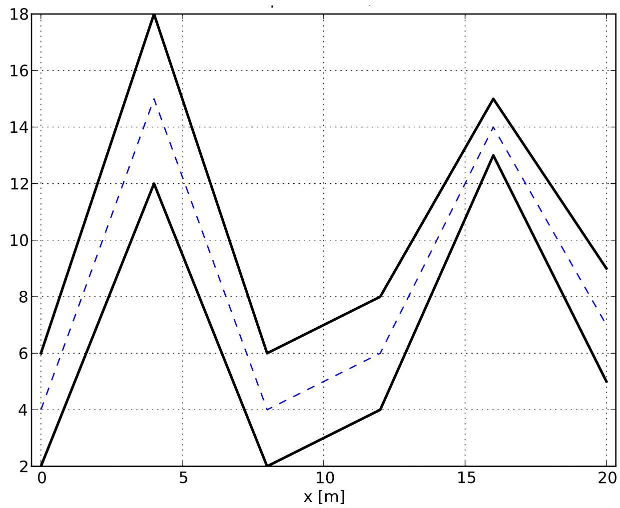

Figure 1: Slalom problem with $n = 6$ gates. The initial and final positions are fixed and not included in the figure. The skier slides from left to right. The middle path is dashed and connects the center points of gates.

Solution:

(a) Assume that (x_i, z_i) is the crossing point of gate i, the path length minimization problem is thus

$$\begin{array}{ll} \min\limits_{z} & \sum\limits_{i=1}^{n+1} \left\| \begin{pmatrix} x_i \\ z_i \end{pmatrix} - \begin{pmatrix} x_{i-1} \\ z_{i-1} \end{pmatrix} \right\|_2 \\ \text{subject to} & y_i - c_i/2 \le z_i \le y_i + c_i/2, \ \text{for } i = 1, \ldots, n \\ & z_0 = y_0, z_{n+1} = y_{n+1}, \end{array}$$

which is equivalent to

$$\begin{array}{ll} \min\limits_{z} & \sum\limits_{i=1}^{n+1} t_i \\ \text{subject to} & y_i - c_i/2 \le z_i \le y_i + c_i/2, \ \text{for } i = 0, \ldots, n+1 \\ & \left\| \begin{pmatrix} x_i \\ z_i \end{pmatrix} - \begin{pmatrix} x_{i-1} \\ z_{i-1} \end{pmatrix} \right\|_2 \le t_i, \ \text{for } i = 1, \ldots, n+1. \end{array}$$

with the convention $c_0 = c_{n+1} = 0$. Hence, the problem is an SOCP.

(b) The code can be found in the corresponding IPython notebook.

4. Robust linear programming

In this problem we will consider a version of linear programming under uncertainty.

(a) Consider vector $\vec{x} \in \mathbb{R}^n$. Recall that $\vec{x}^\top \vec{y} \le \|\vec{x}\|_1$ for all $\vec{y}$ such that $||\vec{y}||_\infty \le 1$. Further this inequality is tight, since it holds with equality for $\vec{y} = \text{sgn}(\vec{x})$. You saw this in the previous homework, just remind your self of the solution, no need to turn anything in.

Let us focus now on a LP in standard form:

$$\begin{aligned} &\min_{\vec{x}} \vec{c}^\top \vec{x} \\ &\text{s.t. } \vec{a}_i^\top \vec{x} \le b_i, \quad i = 1, ..., m. \end{aligned} \tag{4}$$

Consider the set of linear inequalities in (4). Suppose you don't know the vectors $\vec{a}_i$ exactly. Instead you are given nominal values $\vec{\overline{a}}_i$, and you know that the actual vectors satisfy $\|\vec{a}_i - \vec{\overline{a}}_i\|_\infty \le \rho$ for a given $\rho > 0$. In other words, the actual components a_{ij} can be anywhere in the intervals $[\overline{a}_{ij} - \rho, \overline{a}_{ij} + \rho]$. Or equivalently, each vector $\vec{a}_i$ can lie anywhere in a hypercube with corners $\vec{\overline{a}}_i + \vec{v}$ where $\vec{v} \in \{-\rho, \rho\}^n$. We desire that the set of inequalities that constrain problem (4) be satisfied for all possible values of $\vec{a}_i$; i.e., we replace these with the constraints

$$\vec{a}_i^\top \vec{x} \le b_i \ \ \forall \vec{a}_i \in \{\vec{\overline{a}}_i + \vec{v} \ \mid \|\vec{v}\|_\infty \le \rho\} \ \ i = 1, ..., m. \tag{5}$$

Note that the above defines an *infinite* number of constraints (of the form $\vec{\overline{a}}_i^\top \vec{x} + \vec{v}^\top \vec{x} \le b_i, \ \ \forall \vec{v}$ satisfying $\|\vec{v}\|_\infty \le \rho$, $i = 1, 2, \ldots, m$).

(b) Argue why for our LP we can replace the infinite set of constraints as above to a finite set of $2^n m$ constraints of the form,

$$\vec{\overline{a}}_i^\top \vec{x} + \vec{v}^\top \vec{x} \le b_i \ \ \forall \vec{v} \in \{-\rho, \rho\}^n \ \ i = 1, ..., m.$$

Hint: What do you know about the optimal solutions of LPs?

Solution:

We know from LP theory that the maximum of an affine function on a bounded convex set must occur at an extreme point of the set. Thus we only need to consider the maximum value the LHS of Equation (5) could take. Consequently for each i we can go from infinite constraints to 2^n constraints since the set $\|\vec{v}\|_\infty \le \rho$ set has 2^n extreme points (corners of the n-dimensional hypercube). Doing this we obtain the constraint set,

$$\vec{\overline{a}}_i^\top \vec{x} + \vec{v}^\top \vec{x} \le b_i \ \ \forall \vec{v} \in \{-\rho, \rho\}^n \ \ i = 1, ..., m.$$

(c) Use result from part (a) to show that the constraint set in Equation (5) is in fact equivalent to the much more compact set of m nonlinear inequalities

$$\vec{\overline{a}}_i^\top \vec{x} + \rho \|\vec{x}\|_1 \le b_i, \quad i = 1, ..., m.$$

Solution: From Equation (5) we have the constraints,

$$\vec{a}_i^\top \vec{x} \le b_i \ \ \forall \vec{a}_i \in \{\vec{\overline{a}}_i + \vec{v} \ \mid \|\vec{v}\|_\infty \le \rho\} \ \ i = 1, ..., m.$$

This inequality can be written as:

$$\vec{\overline{a}}_i^\top \vec{x} + \vec{v}^\top \vec{x} \le b_i \ \ \forall \vec{v} \in \{v \ \mid \|\vec{v}\|_\infty \le \rho\} \ \ i = 1, ..., m.$$

or equivalently as,

$$\vec{\overline{a}}_i^\top \vec{x} + \rho \vec{v}^\top \vec{x} \le b_i \ \ \forall \vec{v} \in \{\vec{v} \ \mid \|\vec{v}\|_\infty \le 1\} \ \ i = 1, ..., m.$$

This is equivalent to ,

$$\vec{\bar{a}}_i^\top \vec{x} + \rho \max_{\|\vec{v}\|_\infty \le 1} \vec{v}^\top \vec{x} \le b_i \ \ \forall i = 1, ..., m.$$

because if the inequality is satisfied when the LHS is maximized over $\vec{v}$ such that $\|\vec{v}\|_\infty \le 1$ then it is satisfied for all $\vec{v}$ such that $\|\vec{v}\|_\infty \le 1$. From part (a), we have $\max_{\|\vec{y}\|_\infty \le 1} \vec{x}^\top \vec{y} = \|\vec{x}\|_1$, which gives us the equivalent constraint set:

$$\vec{\bar{a}}_i^\top \vec{x} + \rho \|\vec{x}\|_1 \le b_i, \ \ i = 1, ..., m.$$

We now would like to formulate the LP with uncertainty introduced. We are therefore interested in situations where the vectors $\vec{a}_i$ are uncertain, but satisfy bounds $\|\vec{a}_i - \vec{\bar{a}}_i\|_\infty \le \rho$ for given $\vec{\bar{a}}_i$ and ρ. We want to minimize $\vec{c}^\top \vec{x}$ subject to the constraint that the inequalities $\vec{a}_i^\top \vec{x} \le b_i$ are satisfied for *all* possible values of $\vec{a}_i$.

We call this a *robust LP* :

$$\begin{aligned} &\min_{\vec{x}} \vec{c}^\top \vec{x} \\ &\text{s.t. } \ \vec{a}_i^\top \vec{x} \le b_i, \ \ \forall \vec{a}_i \in \{\vec{\bar{a}}_i + \vec{v} \mid \|\vec{v}\|_\infty \le \rho\} \quad i = 1, ..., m. \end{aligned} \tag{6}$$

(d) Using the result from part (c), express the above optimization problem as an LP.

Solution:

From part (c), We can rewrite the problem as

$$\begin{aligned} &\min_{\vec{x}} \vec{c}^\top \vec{x} \\ &\text{s.t. } \ \vec{\bar{a}}_i^\top \vec{x} + \rho \|\vec{x}\|_1 \le b_i, \ \ i = 1, ..., m. \end{aligned}$$

We can express this optimization problem as an LP by introducing variables t_i:

$$\begin{aligned} &\min_{\vec{x}, \vec{t}} \vec{c}^\top \vec{x} \\ &\text{s.t. } \ \vec{\bar{a}}_i^\top \vec{x} + \rho \sum_i t_i \le b_i, \ \ i = 1, ..., m. \\ &x_i \le t_i \ \ i = 1, ..., m. \\ &- x_i \le t_i \ \ i = 1, ..., m. \end{aligned}$$

5. Formulating problems as LPs or QPs

Formulate the problem

$$p_j^* \doteq \min_{\vec{x}} \ f_j(\vec{x}),$$

for different functions f_j, $j = 1, \ldots, 4$, as convex QPs or LPs, or, if you cannot, explain why. In our formulations, we always use $\vec{x} \in \mathbb{R}^n$ as the variable, and assume that $A \in \mathbb{R}^{m \times n}$, $\vec{y} \in \mathbb{R}^m$. If you obtain a convex LP or QP formulation, state precisely what the variables, objective, and constraints are.

(a) $f_1(\vec{x}) = \|A\vec{x} - \vec{y}\|_\infty + \|\vec{x}\|_1$

(b) $f_2(\vec{x}) = \|A\vec{x} - \vec{y}\|_2^2 + \|\vec{x}\|_1$

(c) $f_3(\vec{x}) = \|A\vec{x} - \vec{y}\|_2^2 - \|\vec{x}\|_1$

(d) $f_4(\vec{x}) = \|A\vec{x} - \vec{y}\|_2^2 + \|\vec{x}\|_1^2$

Solution:

(a) For p_1^*, we replace the ℓ_∞ norm with constraints on the maximum absolute value of each element of $Ax - y$, and rewrite absolute values as linear constraints. This gives us the convex LP formulation:

$$p_1^* = \min_{\vec{x},t,\vec{z}} t + \vec{1}^T\vec{z} \ : \ \begin{array}{l} z_i \geq x_i \geq -z_i, \quad i = 1, \ldots, n \\ t \geq (A\vec{x} - \vec{y})_i \geq -t, \quad i = 1, \ldots, m. \end{array}$$

(b) For p_2^*, we obtain the convex QP

$$p_2^* = \min_{\vec{x},\vec{z}} \vec{x}^\top(A^\top A)\vec{x} - 2\vec{y}^\top A\vec{x} + \vec{y}^\top\vec{y} + \vec{1}^T\vec{z} \ : \ z_i \geq x_i \geq -z_i, \quad i = 1, \ldots, n.$$

(c) For p_3^*, the problem is not convex. Consider for instance the special case with $n = 1$, $A = 1$, $y = 0$: plot $f_3(x) = x^2 - |x|$ to verify it is not convex. In general, the objective of the *difference* of two convex functions is not necessarily convex.

(d) For p_4^*, we have the convex QP

$$p_4^* = \min_{\vec{x},\vec{z}} \vec{x}^\top(A^\top A)\vec{x} - 2\vec{y}^\top A\vec{x} + \vec{y}^\top\vec{y} + \left(\sum_{i=1}^n z_i\right)^2 \ : \ z_i \geq x_i \geq -z_i, \quad i = 1, \ldots, n.$$

Notice that $\left(\sum_{i=1}^n z_i\right)^2 = \vec{z}^T Q\vec{z}$, where we define Q as an $n \times n$ matrix of all ones. Thus, our problem can be written as:

$$p_4^* = \min_{\vec{x},\vec{z}} \vec{x}^\top(A^\top A)\vec{x} - 2\vec{y}^\top A\vec{x} + \vec{y}^\top\vec{y} + \vec{z}^T Q\vec{z} \ : \ z_i \geq x_i \geq -z_i, \quad i = 1, \ldots, n.$$

6. Sphere enclosure

Let B_i, $i = 1, \ldots, m$, be m Euclidean balls in $\mathbb{R}^n$, with centers $\vec{x}_i$, and radii $\rho_i \geq 0$. We wish to find a ball B of minimum radius that contains all the B_i, $i = 1, \ldots, m$. Cast this problem as an SOCP.

Solution: Let $\vec{c} \in \mathbb{R}^n$ and $r \geq 0$ denote the center and radius of the enclosing ball B, respectively. We express the given balls B_i as

$$B_i = \{\vec{x} : \vec{x} = \vec{x}_i + \vec{\delta}_i, \ \|\vec{\delta}_i\|_2 \leq \rho_i\}, \quad i = 1, \ldots, m.$$

We have that $B_i \subseteq B$ if and only if

$$\max_{\vec{x} \in B_i} \|\vec{x} - \vec{c}\|_2 \leq r.$$

Note that

$$\max_{\vec{x} \in B_i} \|\vec{x} - \vec{c}\|_2 = \max_{\|\vec{\delta}_i\|_2 \leq \rho_i} \|\vec{x}_i - \vec{c} + \vec{\delta}_i\|_2 = \|\vec{x}_i - \vec{c}\|_2 + \rho_i.$$

The last step follows by choosing $\vec{\delta}_i$ in the direction of $\vec{x}_i - \vec{c}$. The problem is then cast as the following SOCP

$$\begin{array}{ll} \min\limits_{\vec{c},r} & r \\ \text{s.t.} & \|\vec{x}_i - \vec{c}\|_2 + \rho_i \leq r, \quad i = 1, \ldots, m. \end{array}$$

7. Homework process

Whom did you work with on this homework? List the names and SIDs of your group members.