CS164 Assignment 2

Quang Tran

April 7, 2020

1 Question 1

$$\min_{\mathbf{x}} f(\mathbf{x}) = \min_{\mathbf{x}} c^T \mathbf{x}$$

subject to:

$$A\mathbf{x} \leq b,$$

We can rewrite the above inequality constraint as:

$$g(\mathbf{x}) = A\mathbf{x} - b \leq 0$$

The KKT conditions are:

1. Stationarity: Let $g(\mathbf{x}) = A\mathbf{x} - b \leq 0$:

$$\nabla \mathbf{x} f(\mathbf{x}) + (\nabla_x g(\mathbf{x}))\mu = \mathbf{0} \ (\mu \in R^{m \times 1})$$

$$c + A^T \mu = \mathbf{0}$$

(Sanity check on the dimension: $c \in R^{n\times 1}, A^T \in R^{n\times m}, \mu \in R^{m\times 1}$)

2. Complementary slackness condition: Let $g_i (i = 1, 2..., m)$ be each row of $g(x)$. Then

$$\mu_i g_i = 0,$$

which can be rewritten as

$$\mathrm{diag}(\mu)(Ax - b) = 0$$

($\mathrm{diag}(\mu) \in R^{m\times m}$)

3. Primal feasibility: This is just the inequality constraint:

$$g(\mathbf{x}) = A\mathbf{x} - b \leq 0$$

4. Dual feasibility: This says that the multipliers μ is non-negative:

$$\mu \geq 0$$

Why will the optimal solution in general never be found in the interior of the feasible region and will always be on a vertex or facet? We have that the objective function and the constraint functions (each row in g corresponds to a constraint function) are linear and thus convex. In one activity in class session 8.2, we know that the necessity of the KKT conditions hold. That is, suppose x^* is an optimal solution, then x^* satisfies those KKT conditions.

We will show that this also means $x*$ "will never be found in the interior of the feasible region and will always be on a vertex or facet". Suppose for the sake of contradiction that $x*$ is in the interior of the feasible region, which means the equality signs in all the inequality constraints do not occur: $g_i < 0, i = 1, 2, ...m$. Furthermore, we have from the Complementary slackness condition that $\mu_i g_i = 0$, which means all the μ_i is 0 (because all the g_i are non-zero; $\mu_i g_i = 0$ means that either μ_i or g_i is 0). When $\mu = 0$, from the stationarity condition $c + A^T \mu = \mathbf{0}$ we get $c = \mathbf{0}$, which does not make sense. The conclusion is that an optimal solution is found in the interior of the feasible region iff $c = \mathbf{0}$ (in which case the optimal solution is actually any point in the feasible region, because the objective function is then $c^T \mathbf{x} = 0$. From the above we also showed there is at least one i such that $g_i = 0$. That said, an optimal solution can lie on one or the intersection of several facets determined by g_i, which means the solution can be on a vertex (if the intersection is one point) or a facet.

2 Question 2

2.1 l_1 regression problem

Problem statement:

$$\min_{\Theta} \|Y - X\Theta\|_1$$

Rewritten:

$$\min_{\Theta,s} \mathbf{1}^T s$$

subject to

$$A = \begin{bmatrix} -X & -\mathbb{I} \\ X & -\mathbb{I} \end{bmatrix} \begin{bmatrix} \Theta \\ s \end{bmatrix} \leq \begin{bmatrix} -Y \\ Y \end{bmatrix}$$

2.2 l_∞ regression problem

Problem statement:

$$\min_{\Theta} \|Y - X\Theta\|_\infty$$

Rewritten problem:

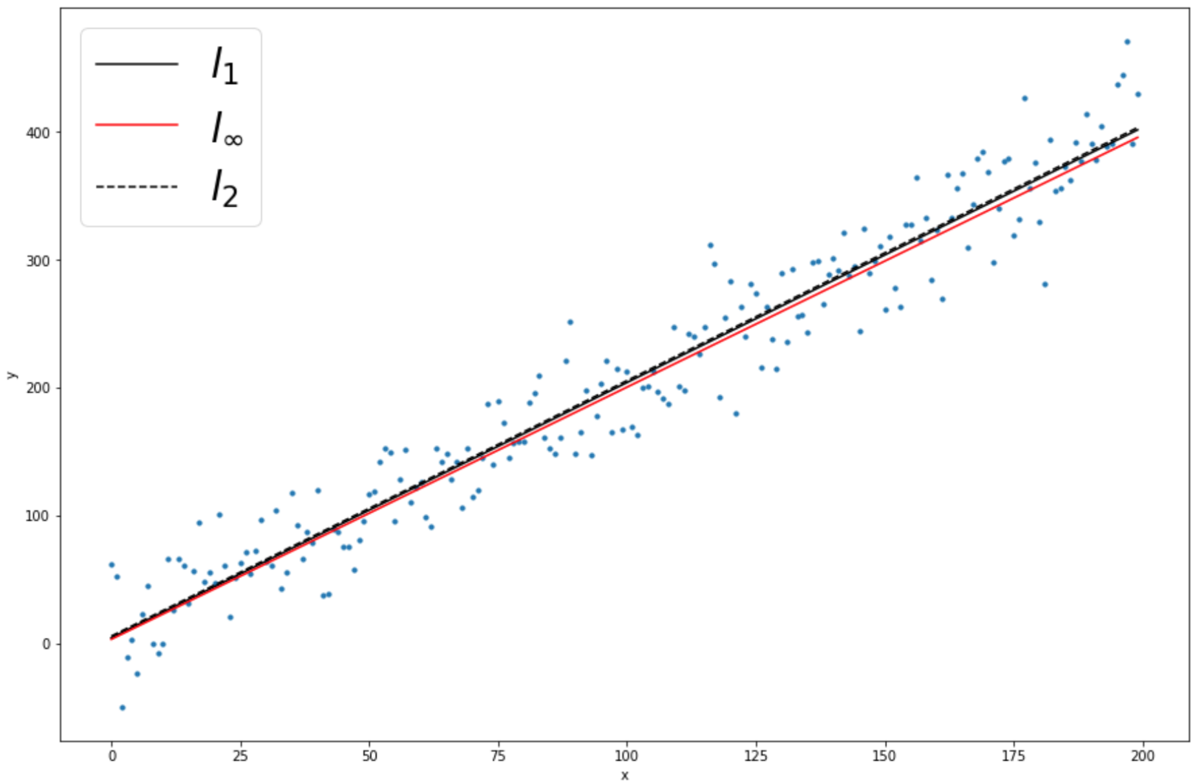
$$\min_{\Theta,s} s$$

subject to

$$A = \begin{bmatrix} -X & -\mathbf{1} \\ X & -\mathbf{1} \end{bmatrix} \begin{bmatrix} \Theta \\ s \end{bmatrix} \leq \begin{bmatrix} -Y \\ Y \end{bmatrix}$$

3 Question 3

Link to code: https://gist.github.com/quangntran/a7e74ecdaa3ab2275102df16121105ce



	Actual	l_1 regression	l_2 regression	l_∞ regression
θ_1	2	1.995	1.998	1.971
θ_1	5	4.334	5.558	3.113

We see that the lines for l_1 and l_2 regression are really close to each other, almost indistinguishable. l_∞ is somewhat further away from the other two lines. Regardless, the three parameters optimized from the three types of regression are pretty close to the actual parameters (see the table above), with l_2 regression is the closest to the truth.

4 Appendix

4.1 Explanation for the rewritten l_1 regression

Problem statement:

$$\min_{\Theta} \|Y - X\Theta\|_1$$

Because the l_1 norm of a vector is the sum of the absolute values of elements in the vector, if we introduce a new variable (which is a vector) s, set $-s \leq Y - X\Theta \leq s$, and try to make each element of s as small as possible, then we get s as a vector of the absolute values of elements in $Y - X\Theta$. Summing elements of s ($\mathbf{1}^T s$) will get us exactly $\|Y - X\Theta\|_1$. With this motivation in mind, we can rewrite the problem as:

$$\min_{\Theta, s} \mathbf{1}^T s \quad (s \in R^{N\times 1})$$

subject to

$$-s_i \leq y_i - x_i^T \Theta \leq s_i, i = 1, 2, ...N$$

where x_i is a row in X, y_i is a row in Y, and s_i is an element in s. The constraint above is not in the standard form. We will transform it into the form $Ax \le b$. First, $Y - X\Theta \le s$ means $-X\Theta - s \le -Y$ and $-s \le Y - X\Theta$ means $X\Theta - s \le Y$. Therefore:

$$A = \begin{bmatrix} -X & -\mathbb{1} \\ X & -\mathbb{1} \end{bmatrix} \begin{bmatrix} \Theta \\ s \end{bmatrix} \le \begin{bmatrix} -Y \\ Y \end{bmatrix}$$

4.2 Explanation for the rewritten l_∞ regression

Problem statement:

$$\min_{\Theta} \|Y - X\Theta\|_\infty$$

Because the l_∞ norm of a vector is the maximum of the absolutes of the elements in the vector, if we introduce a new variable (which is a scalar), and set each and every element of the vector to be in the interval $[-s; s]$, then a the smallest s can get is the l_∞ norm of the vector. That said, we should set $Y - X\Theta \le s\mathbf{1}$ and $-s\mathbf{1} \le Y - X\Theta$. Therefore, our rewritten problem is:

$$\min_{\Theta, s} s \quad (s \in R)$$

subject to

$$-s\mathbf{1} \le Y - X\Theta \le s\mathbf{1}$$

The constraint above is not in the standard form. We will transform it into the form $Ax \le b$. $Y - X\Theta \le s\mathbf{1}$ means $-X\Theta - s\mathbf{1} \le -Y$ and $-s\mathbf{1} \le Y - X\Theta$ means $X\Theta - s\mathbf{1} \le Y$. Therefore:

$$A = \begin{bmatrix} -X & -\mathbf{1} \\ X & -\mathbf{1} \end{bmatrix} \begin{bmatrix} \Theta \\ s \end{bmatrix} \leq \begin{bmatrix} -Y \\ Y \end{bmatrix}$$