

1 Overview

In our previous lecture we presented fundamental results from convex analysis and in particular the separating hyperplane theorem. Today, we will begin the first part (out of a total of three) of our course. We will start with linear optimization, which is a special case of convex optimization (every linear function is convex). Linear functions are often easier to think about, yet they are nonetheless a potent modeling tool – many interesting problems can be modeled in terms of linear optimization. Later in the course we will see how to generalize the main concepts we develop in linear optimization to convex optimization. Let's start.

2 Linear Optimization

In linear optimization we seek to solve an optimization problem:

$$\begin{aligned} \max f(\mathbf{x}) \\ \text{s.t. } g_1(\mathbf{x}) \leq b_1 \\ \cdot \\ \cdot \\ \cdot \\ g_m(\mathbf{x}) \leq b_m \end{aligned}$$

where the functions f and g_1, \dots, g_m are all linear. Recall that a function is *linear* if it can be expressed as:

$$f(\mathbf{x}) = \mathbf{c}^\top \mathbf{x} + b = c_1 x_1 + \dots + c_d x_d + b$$

For convenience we will represent linear functions as $f(\mathbf{x}) = \mathbf{c}^\top \mathbf{x}$ (i.e. without the b term): given $\mathbf{x} \in \mathbb{R}^d$ we can transform \mathbf{x} to a vector in \mathbb{R}^{d+1} as $\mathbf{x}' = (\mathbf{x}, 1)$ and for any given $\mathbf{c} = (c_1, \dots, c_d)$ and $b \in \mathbb{R}$ we can instead consider $\mathbf{c}' = (c_1, \dots, c_d, b)$, and thus $f(\mathbf{x}) = \mathbf{c}^\top \mathbf{x} + b = \mathbf{c}'^\top \mathbf{x}'$.

General representation of linear programs. In general a linear optimization problem is a problem where there exist M_1, M_2, M_3, N_1, N_2 and vectors c, a_1, \dots, a_n s.t. the problem is:

$$\begin{aligned} \min \quad & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} \quad & \mathbf{a}_i^\top \mathbf{x} \geq b_i, \quad i \in M_1 \\ & \mathbf{a}_i^\top \mathbf{x} \leq b_i, \quad i \in M_2 \\ & \mathbf{a}_i^\top \mathbf{x} = b_i, \quad i \in M_3 \\ & x_j \geq 0, \quad j \in N_1 \\ & x_j \leq 0, \quad j \in N_2 \end{aligned}$$

Equivalent forms of LPs. We can always represent any linear program as:

$$\begin{aligned} \min \quad & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

The above formulation is called *standard form* and it is a standard way to represent linear programs. Any linear function optimization under linear constraints can be represented in this way using the following conversion rules:

1. $\max \mathbf{c}^\top \mathbf{x} \iff -\min -\mathbf{c}^\top \mathbf{x}$
2. $\mathbf{a}^\top \mathbf{x} \geq b \iff -\mathbf{a}^\top \mathbf{x} \leq -b$
3. $\mathbf{a}^\top \mathbf{x} = b \iff \mathbf{a}^\top \mathbf{x} \leq b \wedge \mathbf{a}^\top \mathbf{x} \geq b$
4. $\mathbf{a}^\top \mathbf{x} + s = b, s \geq 0 \iff \mathbf{a}^\top \mathbf{x} \leq b$
5. $\mathbf{a}^\top \mathbf{x} - e = b, e \geq 0 \iff \mathbf{a}^\top \mathbf{x} \geq b$

2.1 Examples of Linear Programs

Stock portfolio optimization. Suppose there are n companies s_1, \dots, s_n each company s_i has expected net worth of c_i (buying the stocks of the entire company would cost c_i dollars) and investing in the company has expected profit of w_i . Suppose we have a budget B and would like to use the budget to buy stocks in a matter that maximizes our expected profit:

$$\begin{aligned} \max \quad & \sum_{i=1}^n w_i \cdot x_i \\ \text{s.t.} \quad & \sum_{i=1}^n c_i \cdot x_i \leq B \\ & x_i \geq 0, \forall i \in [n] \\ & x_i \leq 1, \forall i \in [n] \end{aligned}$$

Data fitting. Suppose we are given m data points $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_m, y_m)$ and we wish to construct a predictive model. In the first lecture we discussed the residual sum of squares objective, but there are other similar objectives one may wish to consider. One option is to minimize the ℓ_∞ norm:

$$\min \max_i |y_i - \mathbf{a}^\top \mathbf{x}_i|$$

This can be modeled as a linear optimization problem by adding a variable z that captures the value of the absolute difference:

$$\begin{aligned} \min z \\ \text{s.t. } y_i - \mathbf{a}^\top \mathbf{x} \leq z \quad \forall i = 1, \dots, m \\ -y_i + \mathbf{a}^\top \mathbf{x} \leq z \quad \forall i = 1, \dots, m \end{aligned}$$

Another reasonable objective is to minimize the ℓ_1 norm:

$$\min \sum_{i=1}^m |y_i - \mathbf{a}^\top \mathbf{x}_i|$$

This can also be modeled as a linear optimization problem:

$$\begin{aligned} \min \sum_{i=1}^m z_i \\ \text{s.t. } y_i - \mathbf{a}^\top \mathbf{x} \leq z_i \quad \forall i = 1, \dots, m \\ -y_i + \mathbf{a}^\top \mathbf{x} \leq z_i \quad \forall i = 1, \dots, m \end{aligned}$$

Linear Classification. Finally, we can also consider the problem of linear classification from the previous lecture as a linear optimization problem. To see this, recall that the challenge was to find a separating hyperplane, i.e. given points $\{(\mathbf{x}_i, y_i)\}_{i=1}^m$ where $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \{0, 1\}$ our goal is to find $\mathbf{a} \in \mathbb{R}^d$ and $\alpha \in \mathbb{R}$ s.t. $\mathbf{a}^\top \mathbf{x} \geq \alpha \iff y_i = 1$. We can actually encode this as the constraints of the linear program, and use some arbitrary objective function:

$$\begin{aligned} \min 221 \cdot a_{2016} \\ \text{s.t. } \mathbf{a}^\top \mathbf{x}_i \geq a_{d+1} \quad \forall i : y_i = 1 \\ \mathbf{a}^\top \mathbf{x}_i \leq a_{d+1} \quad \forall i : y_i = 0 \end{aligned}$$

Here a_{d+1} encodes the constant α and the objective $\min 221 \cdot a_{2016}$ is arbitrary (the implicit assumption is that $d + 1 \geq 2016$). Notice that the constraint

$$\mathbf{a}^\top \mathbf{x}_i \geq a_{d+1} \quad \forall i : y_i = 1$$

is shorthand for stating that in the constraint matrix there are k vectors $\mathbf{a}^\top \mathbf{x}_i \geq a_{d+1}$ for all the \mathbf{x}_i s for which $y_i = 1$, where $k \leq m$ is the number of data points whose label is $y_i = 1$. What is interesting here is that we are not interested in optimizing the objective, but rather in finding a feasible solution.

	<i>CPI</i>	<i>CTR</i>	<i>Conversions</i>	<i>Traffic</i>
Site A	\$0.01	0.02	0.003	4,000,000
Site B	\$0.01	0.01	0.002	∞
Budget	80k	100k	-	-

Table 1: Marketing performance data table. The first column shows that advertising on each site has a Cost Per Impression (CPI) of 1 cent. The budget allocated for the campaign is \$80,000. The second column shows the Click Through Rate (CTR) on each site. The CTR is the ratio between the number of people who see the ad and the number who click on it. The retailer is willing to spend 100,000 coupons on this campaign, in expectation. The third column shows the conversion rate on each site: the ratio between number of people who are shown the ad and those who end up making a purchase. The last column shows the traffic: the number of people who visit the site in the time period which we intend to run the campaign.

3 Example: budget allocation in online advertising

Let's do a simple example to be convinced that linear optimization is interesting and useful to explore. An online advertising agency wishes to optimize conversions (sales) for an online retailer which wishes to advertise a product. The campaign involves placing display ads over two sites, and giving a coupon to every person who clicks on the ad. The retailer is interested in maximizing the number of conversions. Table 1 summarizes data typically collected from such experiments.

We can now write the linear program:

$$\max (3x_1 + 2x_2) \cdot 10^{-3} \tag{1}$$

$$s.t. \ x_1 + x_2 \leq 8 \cdot 10^6 \tag{2}$$

$$2x_1 + x_2 \leq 10 \cdot 10^6 \tag{3}$$

$$x_1 \leq 4 \cdot 10^6 \tag{4}$$

Let's begin by examining a few heuristic solutions and see how well they do.

- **Heuristic 1: Buy as many impressions on site A until exhausting traffic, and spend the rest on site B.** In this case we will buy $4 \cdot 10^6$ impressions on site A and $2 \cdot 10^6$ impressions on site B. The number of conversions in this case is 16,000.
- **Heuristic 2: Spend all budget on site B.** In this case we will buy $8 \cdot 10^6$ impressions and the number of conversions in this case too will be 16,000.
- **Heuristic 3: Split budget between A and B in some arbitrary way.** If we chose $2 \cdot 10^6$ impression on site A, and $6 \cdot 10^6$ impressions on B, the expected number of conversions will be 18,000.

So, even in two dimensions, it's not obvious how to find an optimal solution. To get a better understanding of what solutions to LPs look like, we will need to look at the geometry of LPs. In the figure below we plot the constraints. Note that a solution is a point inside the region defined by the intersection of all these constraints.

