Can Pure Offline Data Learning Replace Linear Programming Algorithms?

Yinyu Ye
Shaghai JiaoTong University,
Shanghai Institute for Mathematics and Interdisciplinary Sciences
and Stanford University

September 23, 2026

Abstract

In this note we consider a linear program with input data point $(A, \mathbf{b}, \mathbf{c})$ and its optimal solution $\mathbf{x}(A, \mathbf{b}, \mathbf{c})$. We illustrate that, given two data points $(\underline{A}, \underline{\mathbf{b}}, \underline{\mathbf{c}})$ and $(\bar{A}, \bar{\mathbf{b}}, \bar{\mathbf{c}})$ where they can be arbitrarily close and subscribe the identical optimal (unique) solution, that is, $\mathbf{x}(\underline{A}, \underline{\mathbf{b}}, \underline{\mathbf{c}}) = \mathbf{x}(\bar{A}, \bar{\mathbf{b}}, \bar{\mathbf{c}})$; there still exist infinite non-zero-measure data points between $(\underline{A}, \underline{\mathbf{b}}, \underline{\mathbf{c}})$ and $(\bar{A}, \bar{\mathbf{b}}, \bar{\mathbf{c}})$ and they subscribe exponentially many optimal solutions other than $\mathbf{x}(\bar{A}, \bar{\mathbf{b}}, \bar{\mathbf{c}})$.

Consider linear program (LP) and its dual

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ s.t. & A\mathbf{x} = \mathbf{b}; \\ & \mathbf{x} \geq \mathbf{0} \end{array} \qquad\qquad \begin{array}{ll} \max & \mathbf{b}^T\mathbf{y} \\ s.t. & A^T\mathbf{y} \leq \mathbf{c}. \end{array}$$

The LP duality theorem (e.g. [1]) indicates that pair of $\mathbf{x}$ and $\mathbf{y}$ are optimal for the primal and dual if and only if they are satisfy the constraints and have zero duality gap, that is,,

$$\mathbf{c}^T\mathbf{x} = \mathbf{b}^T\mathbf{y}.$$

Denote $\mathbf{x}(A, \mathbf{b}, \mathbf{c})$ and $\mathbf{y}(A, \mathbf{b}, \mathbf{c})$ by the optimal solution pair corresponding to date point $(A, \mathbf{b}, \mathbf{c})$.

It is easy to see that

Proposition 1. *Let two different data point $(A, \mathbf{b}, \underline{\mathbf{c}})$ and $(A, \mathbf{b}, \bar{\mathbf{c}})$ (with same matrix A and right-hand vector $\mathbf{b}$) have an identical (unique) optimal solution*

$$\mathbf{x}^* := \mathbf{x}(A, \mathbf{b}, \underline{\mathbf{c}}) = \mathbf{x}(A, \mathbf{b}, \bar{\mathbf{c}}).$$

Then, for $0 \leq \epsilon_c \leq 1$, $\mathbf{x}^*$ remains an optimal solution for any new data point $(A, \mathbf{b}, \epsilon_c\underline{\mathbf{c}} + (1 - \epsilon_c)\bar{\mathbf{c}})$ between the two data points.

Let the dual optimal solution be $\underline{\mathbf{y}}^*$ and $\bar{\mathbf{y}}^*$ be the optimal dual solutions for the two data points respectively. Then, we have $\underline{\mathbf{c}}^T\mathbf{x}^* = \mathbf{b}^T\underline{\mathbf{y}}^*$ and $\bar{\mathbf{c}}^T\mathbf{x}^* = \mathbf{b}^T\bar{\mathbf{y}}^*$. Thus, the proof is simply follows that $\mathbf{x}^*$ and the dual solutions $\epsilon_c\underline{\mathbf{y}}^* + (1-\epsilon_c)\bar{\mathbf{y}}^*$ are feasible for the primal and dual respectively of the new data point. Furthermore, we have zero duality gap

$$\mathbf{b}^T(\epsilon_c\underline{\mathbf{y}}^* + (1-\epsilon_c)\bar{\mathbf{y}}^*) = (\epsilon_c\underline{\mathbf{c}} + (1-\epsilon_c)\bar{\mathbf{c}})^T\mathbf{x}^*.$$

Now, the question: does the above proposition remains true if A also changes? Precisely,, let two data points $(\underline{A}, \mathbf{b}, \underline{\mathbf{c}})$ and $(\bar{A}, \mathbf{b}, \bar{\mathbf{c}})$ have an identical (unique) optimal basic solution $\mathbf{x}^*$, does $\mathbf{x}^*$ remain optimal for the new date point $(\epsilon_A\underline{A} + (1-\epsilon_A)\bar{A}, \mathbf{b}, \epsilon_c\underline{\mathbf{c}} + (1-\epsilon_c)\bar{\mathbf{c}})$ for any $0 \le \epsilon_A \le 1$ and $0 \le \epsilon_c \le 1$?

Consider the simple linear program with two variables and one covering constraint

$$\begin{array}{ll} \min & c_1x + c_2x_2 \\ s.t. & a_1x_1 + a_2x_2 \ge 1; \\ & x_1, x_2 \ge 0 \end{array}$$

where that data $(a_1,\ a_2,\ c_1,\ c_2)$ are all positive numbers. Specifically, consider two data point

$$(a_1' = 1,\ a_2' = 1,\ c_1' = 1,\ c_2' = 2) \quad \text{and} \quad (a_1'' = 1,\ a_2'' = N-1,\ c_1'' = 1,\ c_2'' = N)$$

for a positive parameter greater than 2. One can see both data points indicate

$$(x_1^* = 1,\ x_2^* = 0)$$

is the (unique) optimal basic solution or optimal basis since both

$$\frac{c_1'}{a_1'} = 1 < 2 = \frac{c_2'}{a_2'} \quad \text{and} \quad \frac{c_1''}{a_1''} = 1 < \frac{N}{N-1} = \frac{c_2''}{a_2''}.$$

Now consider the convex combination of the two data points

$$(\epsilon_A + (1-\epsilon_A), \quad \epsilon_A + (1-\epsilon_A)(N-1), \quad \epsilon_c + (1-\epsilon_c), \quad 2\epsilon_c + (1-\epsilon_c)N)$$

or

$$(1, \quad \epsilon_A + (1-\epsilon_A)(N-1), \quad 1, \quad 2\epsilon_c + (1-\epsilon_c)N)\,.$$

Thus, if

$$\frac{2\epsilon_c + (1-\epsilon_c)N}{\epsilon_A + (1-\epsilon_A)(N-1)} < 1$$

or

$$\epsilon_c - \epsilon_A > \frac{1}{N-2}$$

then

$$\left(x_1^* = 0,\ x_2^* = \frac{1}{\epsilon_A + (1-\epsilon_A)(N-1)}\right)$$

becomes the unique optimal basis, which can happen when $N > 3$. In particular, let $\epsilon_A = 0$ and $\epsilon_c = \frac{1}{N-3}$. Then, as N increases, the new point would be arbitrarily close to

the second data point ($a''_1 = 1,\ a''_2 = N-1,\ c''_1 = 1,\ c''_2 = N$) but subscribing a difference optimal basis.

One can image this construction can go for n variables and m covering constraints, that is, even $(\underline{A}, \mathbf{b}, \underline{\mathbf{c}})$ and $(\bar{A}, \mathbf{b}, \bar{\mathbf{c}})$ have an identical (unique) optimal basis $\mathbf{x}^*$, there are data point between the two data points to generate exponentially many different optimal basis. Since the problem is scale-invariant, so one can scale the data $(A, \mathbf{b}, \mathbf{c})$ with any small number of $(\epsilon A, \mathbf{b}, \epsilon\mathbf{c})$ where the optima basic solution remains the same. Thus, our finding remains true even when the initial two data points (that subscribe unique optimal basis) can be arbitrarily close. Therefore, the pure offline learning from data, basic on data similarity, is unlikely to accurately predict its optimal basis or optimal solutions for linear programming.

One interesting point is that, if we require the solution to be **integers**, then $(x^*_1 = 1,\ x^*_2 = 0)$ **remains** the unique optimal integral solution for all data points between the original two data points. Therefore, Data learning may be more helpful to Integer Programming due to its solution-Immutability.

References

[1] David Luenberger and Yinyu Ye, Linear and Nonlinear Programming, International Series in Operations Research and Management Science, Sprinter,, Edition 5, 2021.