15.095: Machine Learning under a Modern Optimization Lens

Recitation 1

Sep 7th, 2018

Recitations

1. Basic knowledge about optimization and machine learning
2. How to use julia, JuMP and other packages like OptImpute and Optimal Trees
3. Review of the course material

Outline

1. Introduction to Optimization
2. Modeling and Logical Constraints
3. Big-M method
4. Duality

Optimization: Definition

The target of the optimization problem is to minimize (maximize) a function $f(x)$ when x satisfies the condition $g(x) \leq q$.

$$\min f(x) \quad s.t. \;\; g(x) \leq q.$$

- $f(x)$: objective function
- $g(x) \leq q$: constraints
- x : decision variables

Optimization: Definition

$$\min f(x) \quad s.t. \;\; g(x) \leq q.$$

- x satisfying all of the constraints is called a **feasible solution**. The set of feasible solutions is called the **feasible set**.
- A feasible solution x^* that minimizes the objective function is called an **optimal feasible solution**, or simply, an **optimal solution**.
- The value $f(x^*)$ is called the **optimal cost**.
- If for every real number K we can find a feasible solution x whose cost is less than K, we say that the optimal cost is $-\infty$, or in other words that the problem is **unbounded**.
- By changing $f(x)$ to $-f(x)$, we can rewrite maximization problems as minimization problems.

Different Types of Optimization Problems

1. **Linear Optimization:** f(x), g(x) are both linear functions. (15.081,15.093)
2. **Nonlinear Optimization:** Some constraints or the objective function are nonlinear.(15.084)
3. **Convex Optimization:** Minimizing convex functions over convex sets.(15.084)
    - **Convex set** C: For all x and y in C and all t in the interval $(0,1)$, the point $(1-t)x+ty$ also belongs to C.
    - **Convex function** $f: X \to \mathrm{R}$:
    $\forall x_1, x_2 \in X, \forall t \in [0,1]: \qquad f(tx_1+(1-t)x_2) \leq tf(x_1)+(1-t)f(x_2)$.
    - Benefits of convexity: Local minimum must be a Global minimum
4. **Integer Optimization:** All of the decision variables x are restricted to be integers. **Mixed Integer Optimization (MIO):** Some decision variables are restricted to be integers, others are not. (15.083)
5. **Robust Optimization:** Optimization under uncertainty. (15.094)

Modeling: Logical Constraints

Assume $x_1, x_2, \ldots x_n$ and $y_1, y_2, \ldots y_n$ are binary decision variables $\{0, 1\}$.

1. Exactly k of $x_1, x_2, \ldots x_n$ are equal to 1.
   $x_1 + x_2 + x_3 + \ldots + x_n = k$
2. At most k of $x_1, x_2, \ldots x_n$ are equal to 1.
   $x_1 + x_2 + x_3 + \ldots + x_n \leq k$
3. If $x_1 = 1$, then $y_1 = 1$.
   $x_1 \leq y_1$
4. If at least k of $x_1, x_2, \ldots x_n$ equals 1, then $y_1 = 1$.
   $x_1 + x_2 + x_3 + \ldots + x_n - (k - 1) \leq n * y_1$

Modeling: Logical Constraints

Assume $x_1, x_2, \ldots x_n$ and $y_1, y_2, \ldots y_n$ are binary decision variables $\{0, 1\}$.

1. If at least k of $x_1, x_2, \ldots x_n$ equal 1, then at least one of $y_1, y_2, \ldots y_n$ equals 1.
   $x_1 + x_2 + x_3 + \ldots + x_n - (k - 1) \leq n * (y_1 + y_2 + y_3 + \ldots + y_n)$
2. If $x_1 = 1$, then we add C to the objective.
   We do not add a constraint. Add the term $C * x_1$ to the objective function.
3. Consider a min problem. If at least one of $x_1, x_2, ...x_n$ equals 1, then we add C to the objective.
   We introduce a new binary variable z. We add $C * z$ to the objective and add the constraint: $x_1 + x_2 + \ldots + x_n \leq n * z$.

One important thing is we can only use the above constraints when x and y are restricted to be **binary variables**.

Big-M Method for Modeling

Suppose that we have a continuous variable α_i, which is a free variable if $z_i = 1$, but if $z_i = 0$ then we must enforce $\alpha_i = 0$. How do we model this?

We add the constraint:
$-Mz_i \leq \alpha_i \leq Mz_i$, where M is a big number.

For example, we may set $M = 1,000,000$ when solving the optimization problem.

Duality: Formulation

Original Problem:

$$
\begin{aligned}
\underset{x}{\text{minimize}} \quad & c^{'}x \\
\text{subject to} \quad & a_i^{'}x \geq b_i, \ i \in M_1, \\
& a_i^{'}x \leq b_i, \ i \in M_2, \\
& a_i^{'}x = b_i, \ i \in M_3, \\
& x_j \geq 0, \ j \in N_1, \\
& x_j \leq 0, \ j \in N_2, \\
& x_j \ \ \text{free}, \ j \in N_3.
\end{aligned}
$$

Dual Problem:

$$
\begin{aligned}
\underset{p}{\text{maximize}} \quad & p^{'}b \\
\text{subject to} \quad & p_i \geq 0, \ i \in M_1, \\
& p_i \leq 0, \ i \in M_2, \\
& p_i \ \ \text{free}, \ i \in M_3. \\
& p_i^{'}A_j \geq c_j, \ i \in N_1, \\
& p_i^{'}A_j \leq c_j, \ i \in N_2, \\
& p_i^{'}A_j = c_j, \ i \in N_3.
\end{aligned}
$$

Duality: Properties

Theorem (Weak Duality)

If x is a feasible solution to the primal problem and p is a feasible solution to the dual problem, then

$$p^{'}b \leq c^{'}x.$$

Theorem (Strong Duality)

If a linear programming problem has an optimal solution, then so does its dual, and the respective optimal costs are equal.

Duality: Usage

1. Dual problem is easier to solve sometimes.
2. Dual Simplex Method to solve linear optimization problems.
3. Tractable reformulations for robust optimization problems.
4. Lagrangian duality in nonlinear optimization.

References

1. Intro to Optimization: Bertsimas and Tsitsiklis, *Introduction to Linear Optimization*, Chapter 1.
2. Duality: Bertsimas and Tsitsiklis, *Introduction to Linear Optimization*, Chapter 4.