Optimization Theory for Machine Learning

ESC 22 Winter

Kim Dueun

Department of Applied statistics

Table of contents

Mathematical optimization

Definition (Mathematical optimization)

$$\textit{minimize } f_0(\mathbf{x})$$

$$\textit{subject to } f_i(\mathbf{x}) \leq b_i, \;\; i = 1, 2, ..., m.$$

$\mathbf{x} \in \mathbb{R}^n$ is a optimization variable.
$f_0 : \mathbb{R}^n \to \mathbb{R}$ is a objective function.
$f_i : \mathbb{R}^n \to \mathbb{R}$ are constraint functions, and b_i are the bounds of them.

$\mathbf{x}^*$ is optimal, or a solution of the optimization problem. for any $\mathbf{z}$ with $f_1(\mathbf{z}) \leq b_1, f_2(\mathbf{z}) \leq b_2, \ldots, f_m(\mathbf{z}) \leq b_m$, $f_0(\mathbf{z}) \geq f_0(\mathbf{x}^*)$.

Least squares and linear programming

Simple optimization problem with no constraints.

Definition (Least squares)

$$minimize\ f_0(\mathbf{x}) = \sum_{i=1}^{k}(a_i^T x + b_i)^2$$

It has known analytical solution, $x = (A^T A)^{-1} Ab$. A current desktop computer can solve a least-squares problem with hundreds of variables, and thousands of terms, in a few seconds.

Weighted least squares, regularization also can be used.

Least squares and linear programming

Definition (Linear programming)

$$\textit{minimize } c^T x$$

$$\textit{subject to } a_i^T x \leq b_i \; i = 1, 2, ...m$$

There is NO analytical solution for linear programming problem. But we can also solve this kind of problem easily with reliable and efficient algorithms.

There are a few simple tricks for converting a problem to linear programming.

Convex optimization

Convex optimization

Definition (Convex optimization problem)

$$\textit{minimize } f_0(x)$$

$$\textit{subject to } f_i(x) \leq b_i,\ i = 1, 2, ...m$$

Where objective and constraints are convex;

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y),\ 0 \leq \theta \leq 1$$

There is no analytical solution but we can solve this kind of problem with reliable and efficient algorithms. The difficult thing is 'recognizing' convex optimization problem.

Convex optimization problem

Example

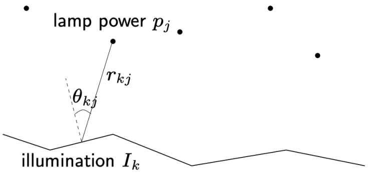
m lamps illuminating n patches, Intensity I_k at patch k depends on lamp powers p_j;

$$I_k = \sum_{j=1}^{m} a_{kj} p_j,\ a_{kj} = r_{kj}^{-2} max\{cos\theta_{kj}, 0\}$$

problem: achieve desired illumination I_{des} with bounded lamp powers.

$$minimize\ max_{k=1,2,\dots,n} |\log I_k - \log I_{des}|$$

$$subject\ to\ 0 \leq p_j \leq p_{max},\ j = 1, \dots, m$$



How to solve?

- use uniform power p; $p_j = p$
- use least-squares:

$$minimize \sum_{k=1}^{n} (I_k - I_{des})^2$$

round if $p_j > p_{max}$ *or* $p_j < 0$

- use weighted least squares:

$$minimize \sum_{k=1}^{n} (I_k - I_{des})^2 + \sum_{j=1}^{m} w_j (p_j - p_{max}/2)^2$$

iteratively adjust weights until $0 \leq p_j \leq p_{max}$

How to solve?

- use linear programming:

$$minimize\ max_{k=1,2,...,n}|I_k - I_{des}|$$

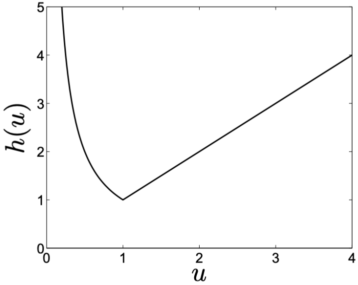
$$subject\ to\ 0 \leq p_j \leq p_{max},\ \ j = 1, ..., m$$

- use convex optimization:

$$minimize\ f_0(p) = max_{k=1,2,...,n}h(I_k/I_{des})$$

$$subject\ to\ 0 \leq p_j \leq p_{max},\ \ j = 1, ..., m$$

How to solve?



$$h(u) = max\{u, 1/u\}$$

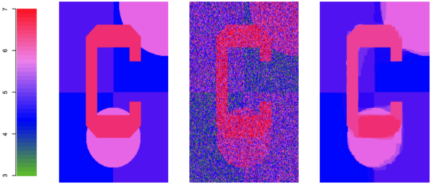
It can be solved with similar complexity to the least squares method.
But when other constraints are added, the problem can become even more complex.
ex) no more than half of the lamps are on($p_j > 0$)

Image denoising problem

Example

In signal processing, particularly image processing, total variation denoising, also known as total variation regularization, is a noise removal process(filter).

$$minimize\ f(\theta) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta_i)^2 + \lambda \sum_{(i,j)\in E} |\theta_i - \theta_j|$$



True image Data Solution

What we will study

Specialized ADMM, Coordinate descent result

- Convex sets, Convex functions
- Convex optimization problems, Duality
- Important algorithms, ML applications
- Dynamic programming?