Optimization Problems and Convexity

Andrew Nobel

April, 2020

General Optimization Problem

Problem: Minimize a function $f : \mathbb{R}^d \to \mathbb{R}$ over a set $A \subseteq \mathbb{R}^d$ of interest.
Often expressed in the form of a *mathematical program*

$$\min f(x) \text{ subject to } x \in A$$

- f called an *objective function*
- A represents *constraints* on arguments x of interest
- points $x \in A$ called *feasible*
- usually interested in $\min_A f(x)$ and $\operatorname{argmin}_A f(x)$

General Optimization Problem, cont.

Global and local minima

- $u \in A$ is a *global minimum* of f if $f(u) \leq f(v)$ for all $v \in A$
- $u \in A$ is a *local minimum* of f if there exists $r > 0$ such that $f(u) \leq f(v)$ for all $v \in A$ such that $||u - v|| \leq r$.

Note: A global minimum is a local minimum

General optimization problem difficult or impossible to solve

- no closed form solution, solution may not be unique
- many local and/or global minima
- no good iterative or approximate solutions

Review: Convex Sets and Functions

Recall

- A set $C \subseteq \mathbb{R}^d$ is convex if for every $x, y \in C$ and $\alpha \in [0, 1]$ the point $\alpha x + (1 - \alpha)y \in C$.
- An intersection of convex sets is convex
- A function $f : C \to \mathbb{R}$ is convex if for every $x, y \in C$ and $\alpha \in [0, 1]$

  $$f(\alpha x + (1 - \alpha)y) \ \leq \ \alpha f(x) + (1 - \alpha)f(y)$$

  We say f is *strictly convex* if $\leq$ is replaced by strict inequality ($<$) whenever $x \neq y$ and $\alpha \in (0, 1)$.
- The maximum of convex functions is convex

Convexity and Optimization

Fact: If $C \subseteq \mathbb{R}^d$ is convex and $f : C \to \mathbb{R}$ is convex then

- any local minimum is a global minimum
- if f is strictly convex any global minimum is unique

Generally speaking: If $C \subseteq \mathbb{R}^d$ and $f : C \to \mathbb{R}$ are convex then there are efficient iterative methods to find the global minimum of f when it exists

Examples

1. Best linear classification rule. Given $(x_1, y_1), \dots, (x_n, y_n) \in \mathbb{R}^p \times \{0, 1\}$

$$\min f(\beta) = \sum_{i=1}^{n} \mathbb{I}((2y_i - 1)(x_i^t \beta - \beta_0) \geq 0) \text{ over } \beta \in \mathbb{R}^{d+1}$$

2. Least squares regression. Given $(x_1, y_1), \dots, (x_n, y_n) \in \mathbb{R}^p \times \mathbb{R}$

$$\min f(\beta) = \sum_{i=1}^{n} (y_i - x_i^t \beta)^2 \text{ over } \beta \in \mathbb{R}^{d+1}$$

3. Quadratic program. Given matrices $A \in \mathbb{R}^{d \times d}$ symmetric and $B \in \mathbb{R}^{m \times d}$, and vectors $c \in \mathbb{R}^d$ and $b \in \mathbb{R}^m$

$$\min f(x) = x^t A x + c^t x \text{ subject to } Bx \leq b$$