CS229 convex optimization notes

James Chuang

January 30, 2017

Contents

My notes on CS229 Convex Optimization Overview notes by Zico Kolter and Honglak Lee.

1. Intro

- Many situations in machine learning require ***optimization*** of the value of some function
- I.e., given $f : \mathbb{R}^n \to \mathbb{R}$, want to find $x \in \mathbb{R}^n$ that minimizes/maximizes $f(x)$
- least-squares, logistic regression, and support vector machines can all be framed as optimization problems
- in general, finding the global optimum of a function is very difficult
  - for a ***convex optimization problems***, we can efficiently find the global solution in many cases

2. Convex Sets

Convex sets:

- A set C is convex if, for any $x, y \in C$ and $\theta \in \mathbb{R}$ with $0 \leq \theta \leq 1$,

$$\theta x + (1 - \theta)y \in C$$

- This means that for any two elements in C, every point on the line segment between those points also belongs to C.
- The point $\theta x + (1 - \theta)y$ is called a ***convex combination*** of the points x and y.
- Examples of convex sets:
  - all of $\mathbb{R}^n$
  - the non-negative orthant, $\mathbb{R}_+$:
    - all vectors in $\mathbb{R}^n$ whose elements are all non-negative: $\mathbb{R}^n_+ = \{x : x_i \geq 0 \ \forall \ i = 1, \ldots, n\}$
  - norm balls
  - affine subspaces and polyhedra
  - intersections of convex sets
  - positive semidefinite matrices

3. Convex Functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is ***convex*** if its domain (denoted $\mathcal{D}(f)$) is a convex set, and if, for all $x, y \in \mathcal{D}(f)$ and $\theta \in \mathbb{R}, 0 \geq \theta \geq 1,$

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

Intuitively, the way to think about this definition is that if we pick any two points on the graph of a convex function and draw a straight line between them, then the portion of the function between these two points will lie below this straight line.

We say a function is ***strictly convex*** if this definition holds with strict inequality for $x \neq y$ and $0 < \theta < 1$. We say that f is ***concave*** if $-f$ is convex, and likewise that f is ***strictly concave*** if $-f$ is strictly convex.

3.1 First Order Condition for Convexity

Suppose a function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable (i.e., the gradient $\nabla_x f(x)$ exists at all points $x \in \mathcal{D}(f)$). Then f is convex iff $\mathbb{D}(f)$ is a convex set and for all $x, y \in \mathcal{D}(f)$,

$$f(y) \geq f(x) + \nabla_x f(x)^T (y - x)$$

- $f(x) + \nabla_x f(x)^T (y - x) \leftarrow$ ***first-order approximation*** of the function f at the point x.
  - I.e., the line tangent to f at x is a global underestimator of the function f
- Similarly, f is:
  - strictly convex if this holds w/ strict inequality
  - concave if the inequality is reversed
    - strictly concave if this reverse inequality is strict

3.2 Second Order Condition for Convexity

Suppose a function $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable (i.e., the Hessian $\nabla_x^2 f(s)$ is defined for all points x in the domain of f). Then f is convex iff $\mathcal{D}(f)$ is a convex set and its Hessian is positive semidefinite: i.e., for any $x \in \mathcal{D}(f)$,

$$\nabla_x^2 f(x) \succeq 0.$$

- In one dimension, this is equivalent to the condition that the second derivative $f''(x)$ always be non-negative
- if Hessian is:
  - positive definite, f is strictly convex
  - negative semidefinite, f is concave
  - negative definite, f is negative definite

3.3 Jensen's Inequality

Suppose we start with the inequality in the basic definition of a convex function

$$f(\theta x + (1-\theta)\, y) \leq \theta f(x) + (1-\theta)\, f(y) \quad \text{for } 0 \leq \theta \leq 1.$$

This can be extended (by induction) to convex combinations of more than one point,

$$f\left(\sum_{i=1}^{k} \theta_i x_i\right) \leq \sum_{i=1}^{k} \theta_i f(x_i) \quad \text{for } \sum_{i=1}^{k} \theta_i = 1, \theta_i \geq 0 \;\; \forall i$$

This can be further extended to infinite sums or integrals:

$$f\left(\int p(x)x dx\right) \leq \int p(x)f(x)dx \quad \text{for} \quad \int p(x)dx = 1, p(x) \geq 0 \ \ \forall x$$

Since $\int p(x)dx = 1$, it can be interpreted as a probability density, in which case the above can be written as expectations:

$$f\left(\mathsf{E}\left[x\right]\right) \leq \mathsf{E}[f(x)].$$

This is ***Jensen's inequality***.

3.4 Sublevel Sets

Convex functions give rise to an important of convex set called an α***-sublevel set***. Given a convex function $f : \mathbb{R}^n \to \mathbb{R}$ and a real number $\alpha \in \mathbb{R}$, the α-sublevel set is defined as

$$\left\{x \in \mathcal{D}(f) : f(x) \leq \alpha\right\}.$$

I.e., the α-sublevel set is the set of points x s.t. $f(x) \leq \alpha$.

3.5 Examples

4. Convex Optimization Problems

A ***convex optimization problem*** is of the form

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & x \in C \end{aligned}$$

- f: a convex function
- C: a convex set
- x: the optimization variable

The same problem written more explicitly:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{aligned}$$

- f: a convex function
- $g_i(x)$ are convex functions
- $h_i(x)$ are affine functions
- x: the optimization variable

The ***optimal value*** of an optimization problem is denoted p^* (sometimes f^*) and is equal to the minimum possible value of the objective function in the feasible region

$$p^* = \min\left\{f(x) : \ \ g_i(x) \leq 0, i = 1, \ldots, m, \ \ h_i(x) = 0, 1, \ldots, p\right\}.$$

p^* can take on the values $+\infty$ or $-\infty$ when the problem is *infeasible* (feasible region is empty) or *unbounded below* (feasible points exist s.t. $f(x) \to -\infty$), respectively. We say that x^* is an ***optimal point*** if $f(x^*) = p^*$. Note that there can be more than one optimal point, even when the optimal value is finite.

4.1 Global Optimality in Convex Problems

Intiuitive definitions:

- ***locally optimal*** - no "nearby" feasible points that have a lower objective value
- ***globally optimal*** - no feasible points at all that have a lower objective value

Formal definitions:

- a point x is ***locally optimal*** if it is feasible (i.e., satisfies the constrains of the optimization problem) and if there exists some $R > 0$ s.t. all feasible points z with $\|x - z\|_2 \leq R$, satisfy $f(x) \leq f(z)$.
- a point x is ***globally optimal*** if it is feasible and for all feasible points z, $f(x) \leq f(z)$.

The key idea is that **for a convex optimization problem, all locally optimal points are globally optimal**.

Proof by contradiction:

- Suppose x is locally optimal but *not* globally optimal.
  - I.e., there exists a feasible point y s.t. $f(x) > f(y)$.
  - By definition of local optimality, there exist no feasible points z s.t. $\|x - z\|_2 \leq R$ and $f(z) < f(x)$.
- Then, suppose we choose the point

$$z = \theta y + (1-\theta)\, x \quad \text{with} \quad \theta = \frac{R}{2\|x - y\|_2}.$$

- then,

$$\begin{aligned}
\|x - z\|_2 &= \left\| x - \left(\frac{R}{2\|x-y\|_2} y + \left(1 - \frac{R}{2\|x-y\|_2}\right) x \right) \right\|_2 \\
&= \left\| \frac{R}{2\|x-y\|_2} (x - y) \right\|_2 \\
&= \frac{R}{2} \leq R
\end{aligned}$$

- By convexity of f, we have:

$$f(z) = f(\theta y + (1-\theta)x) \leq \theta f(y) + (1-\theta) f(x) < f(x)$$

- the feasible set is a convex set
  - x and y are both feasible
  - therefore, $z = \theta y + (1-\theta)x$ is also feasible
- since z is feasible and:
  - $\|x - z\|_2 < R$ (i.e. z is within the radius R neighborhood of x)
  - $f(z) < f(x)$
- therefore, the point x which is not globally optimal, cannot be locally optimal

4.2 Special Cases of Convex Problems

Some special cases of the general convex optimization problem have efficient algorithms to solve very large problems

- **Linear Programming**. A convex optimization problem is a **linear program** (LP) if both the objective function f and inequality constraints g_i are affine functions. In other words, these problems have the form

$$\begin{aligned}
\text{minimize} \quad & c^T x + d \\
\text{subject to} \quad & Gx \preceq h \\
& Ax = b
\end{aligned}$$

- $x \in \mathbb{R}^n$: the optimization variable
- $c \in \mathbb{R}^n, d \in \mathbb{R}, G \in \mathbb{R}^{m \times n}, h \in \mathbb{R}^m, A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p$
- '$\preceq$' denotes elementwise inequality
- **Quadratic Programming**. A convex optimization problem is a **quadratic program** (QP) if the inequality constraints g_i are still all affine, but if the objective function f is a convex quadratic function. In other words, these problems ahve the form

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} x^T P x + c^T x + d \\ \text{subject to} \quad & G(x) \preceq h \\ & Ax = b \end{aligned}$$

- $x \in \mathbb{R}^n$: the optimization variable
- $c \in \mathbb{R}^n, d \in \mathbb{R}, G \in \mathbb{R}^{m \times n}, h \in \mathbb{R}^m, A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p$
- $P \in \mathbb{S}_+^n$, a symmetric positive semidefinite matrix
- **Quadratically Constrained Quadratic Programming**. A convex optimization problem is a **quadratically constrained quadratic program** (QCQP) if both the objective f and the inequality constraints g_i are convex quadratic functions