Convex Optimization Overview

Conner DiPaolo

Contents

1 Introduction

Convex optimization in a large way influences the way people think about and phrase machine learning problems. Almost all problems we will see in our studies are developed or can be viewed as optimization problems. Some problems, like the Support Vector Machine you will all see in the coming weeks, are almost entirely based in the heart of convex optimization.

We don't plan on bringing you all up to speed completely on the art of Convex Optimization, but hopefully in two sections you will know enough to be able to think about problems in new, interesting ways that will aid your studies in and out of machine learning.

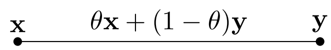

Figure 1: **Convex Combination**. The line segment between $\mathbf{x}$ and $\mathbf{y}$ above represents every possible combination $\{\theta\mathbf{x} + (1-\theta)\mathbf{y} : 0 \leq \theta \leq 1\}$.

Figure 2: **Set convexity.** Intuitively, a set is convex if a line drawn between any two points within the set lies completely in the set. The convex set to the left is a *convex hull* of it's vertices, meaning the set is constructed as all possible convex combinations of its vertices. These sets are subsets of $\mathbb{R}^2$.

The only real prerequisite for this material is a strong confidence in linear algebra and familiarity with matrix calculus.

2 Convex Sets

The first step into examining convexity is defining what a **convex set** is when given, for example, a subset of the real numbers, or the set of matrices.

Definition 2.1 (Convex Combination)**.** *In the $n = 2$ case, the convex combination of points $\mathbf{x}$ and $\mathbf{y}$ in an affine space is*

$$\theta\mathbf{x} + (1-\theta)\mathbf{y}$$

where $0 \leq \theta \leq 1$. Intuitively, this is the line segment between $\mathbf{x}$ and $\mathbf{y}$ (consider $\theta = 0$ and $\theta = 1$), as seen in Figure 1. More generally, a convex combination of points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ in an affine space (vector spaces included) is the combination

$$\boldsymbol{\theta}_1\mathbf{x}_1 + \boldsymbol{\theta}_2\mathbf{x}_2 + \cdots + \boldsymbol{\theta}_n\mathbf{x}_n = \boldsymbol{\theta}^\top\mathbf{x}$$

where $\theta_i \geq 0$ and $\mathbf{1}^\top\theta = 1$. That is, $\boldsymbol{\theta}$ lies on the standard probability simplex.

With this definition of a convex combination we can define a convex set:

Definition 2.2 (Convex Set)**.** *A set C is convex if, given $\mathbf{x}, \mathbf{y} \in C$, every convex combination of $\mathbf{x}$ and $\mathbf{y}$ is still in C. Mathematically,*

$$\theta\mathbf{x} + (1-\theta)\mathbf{y} \in C. \qquad (0 \leq \theta \leq 1)$$

See Figure 2 for an illustration. Intuitively, this means that every line segment between any two points in C is contained entirely within C.

2.1 Examples of Convex and Non-Convex Sets

Example 2.1 (All of $\mathbb{R}^n$ is convex.)**.** *$\mathbb{R}^n$ is convex.*

Proof. As a vector space, for any $\theta_1, \theta_2 \in \mathbb{R}$ and $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$,

$$\theta_1\mathbf{x} + \theta_2\mathbf{y} \in \mathbb{R}^n.$$

Thus, restricting $\theta_2 = 1 - \theta_1$ and $0 \leq \theta_1 \leq 1$ does not change this fact, and any convex combination is also in $\mathbb{R}^n$, making the set convex by definition. ■

Example 2.2 (The Non-Negative Orthant $\mathbb{R}^n_+$)**.** *The set of all vectors*

$$\{\mathbf{x} : \mathbf{x} \in \mathbb{R}^n \text{ and } \mathbf{x}_i \geq 0\}$$

is convex.

Proof. Left as an exercise to the reader. ■

Example 2.3 (Closed Intervals in $\mathbb{R}$ are Convex)**.** *Let $C = [a, b]$ be a subset of the real numbers where $a \leq b$. Then C is convex.*

Proof. Suppose, without loss of generality, that $x_1 \leq x_2$ where $x_1, x_2 \in [a, b]$. Now let $0 \leq \theta \leq 1$. Then

$$\theta x_1 + (1-\theta)x_2 \leq \theta x_2 + (1-\theta)x_2 = x_2$$

because $\theta x_1 \leq \theta x_2$. Similarly,

$$\theta x_1 + (1-\theta)x_2 \geq \theta x_1 + (1-\theta)x_1 = x_1$$

because $(1-\theta)x_2 \geq (1-\theta)x_1$. Thus

$$a \leq x_1 \leq \theta x_1 + (1-\theta)x_2 \leq x_2 \leq b,$$

and hence

$$\theta x_1 + (1-\theta)x_2 \in [a, b].$$

Therefore, by the definition of convexity, $[a, b] \subseteq \mathbb{R}$ is convex for any $a \leq b$. ■

Example 2.4 (The Set of All Complex Hermetian Matrices is Convex)**.** *Let $C = \{A : A \in \mathbb{C}^{n\times n}$ and $A^* = A\}$. C is convex.*

Proof. Let $A, B \in \mathbb{C}^{n\times n}$ be Hermitian matrices and $0 \leq \theta \leq 1$. Then

$$\begin{aligned}
(\theta A + (1-\theta)B)^* &= (\theta A)^* + [(1-\theta)B]^* && (1)\\
&= \theta A^* + (1-\theta)B^* && (2)\\
&= \theta A + (1-\theta)B && \text{(because } A^* = A \text{ and } B^* = B\text{)} \quad (3)
\end{aligned}$$

Thus every convex combination of Hermitian matrices is Hermitian, and by the definition of convexity the set is convex. ■

Corollary 2.1 (The Set of All Real Symmetric Matrices is Convex)**.** *Let $C = \{A : A \in \mathbb{R}^{n\times n}$ and $A^\top = A\}$. C is convex.*

Proof. Left as an exercise to the reader. ■

Example 2.5 (The Set of All Linear Matrix Inequalities is Convex)**.** *Let A_i and B be symmetric $n \times n$ matrices and $\mathbf{x} \in \mathbb{R}^n$. Let $C = \{\mathbf{x} : A(\mathbf{x}) \preceq B\}$ where $A(\mathbf{x}) = \mathbf{x}_1 A_1 + \cdots + \mathbf{x}_k A_k$. C is convex.*

Proof. Let $0 \leq \theta \leq 1$ and $\mathbf{x}, \mathbf{y} \in C$. Then

$$\begin{aligned}
A(\theta\mathbf{x} + (1-\theta)\mathbf{y}) &= \sum_i [\theta\mathbf{x}_i + (1-\theta)\mathbf{y}_i]\, A_i\\
&= \theta\sum_i \mathbf{x}_i A_i + (1-\theta)\sum_i \mathbf{x}_i A_i\\
&= \theta A(\mathbf{x}) + (1-\theta)A(\mathbf{y})\\
&\leq \theta B + (1-\theta)B\\
&= B.
\end{aligned}$$

Thus any convex combination of elements of C is contained in C and by definition C is convex. ■

Example 2.6 (The Space of Probability Distributions is Convex)**.** *Let $\mathcal{P}$ be the space of continuous probability distributions over $\mathbb{R}^n$. That is, every element of $\mathcal{P}$ defines a unique probability density function $\mathbb{P}(x) \geq 0$ such that*

$$\int_{\mathbb{R}^n} \mathbb{P}(x)dx = 1.$$

$\mathcal{P}$ is convex.

Proof. Let f and h be valid probability distributions from $\mathcal{P}$. That is,

$$f, h \in \left\{ \mathbb{P}(x) : \int_{\mathbb{R}^n} \mathbb{P}(x)dx = 1 \text{ and } \mathbb{P}(x) \geq 0 \right\}.$$

Now let $0 \leq \theta \leq 1$. Then

$$\theta f(x) + (1-\theta)h(x) \geq 0$$

as a positive combination of positive functions. Similarly,

$$\begin{aligned} \int_{\mathbb{R}^n} \left[\theta f(x) + (1-\theta)h(x)\right] dx &= \theta \int_{\mathbb{R}^n} f(x)dx + (1-\theta) \int_{\mathbb{R}^n} h(x)dx \\ &= \theta + (1-\theta) = 1. \end{aligned}$$

Thus every convex combination of probability distributions over $\mathbb{R}^n$ is a valid distribution (often called a mixture), and therefore the set of all valid probability distributions over $\mathbb{R}^n$ is convex itself. ∎

Example 2.7 (Disjoint Intervals in $\mathbb{R}$ are Not Convex)**.** *Let $N = [a,b] \cup [c,d]$ where $a \leq b < c \leq d$. N is* ***not*** *convex.*

Proof. Let $b, c \in N$ be as described above. Then for $0 < \theta < 1$ (not we aren't including inequality),

$$\theta b + (1-\theta)c \notin N.$$

Thus not *every* convex combination of elements in N is in N, and N is not convex. ∎

3 Convex Functions

3.1 First Order Conditions: $f(x) \geq f(y) + \nabla f(y)^\top (x - y)$

3.1.1 Examples On Determining Convexity

3.2 Second Order Conditions: $\nabla^2 f \succeq 0$

3.2.1 More Examples On Determining Convexity

3.3 Concavity and Linearity

4 Optimization Problems

4.1 Convex Optimization Problems

4.1.1 Linear Programs

4.1.2 Quadratic Programs

4.1.3 Global Optimality of a Convex Optimum

4.2 Optimal Flow As a Convex Optimization Problem

5 Algorithms for Solving Convex Problems

5.1 Gradient Descent

5.2 Newton's Method

6 Duality

6.1 Duality Gap

6.2 The Lagrange Dual Function

6.3 Finding the Dual

6.3.1 The Dual of The Linear Program

6.4 Solving Problems Using Duality

6.4.1 Minimize a Quadratic Under Quadratic Constraints

6.5 Complementary Slackness of Duality

6.5.1 Solving The Dual Linear Program via a Primal Solution

6.6 The KKT (Karush-Kuhn-Tucker) Conditions