Convex Sets

Guy Lebanon

May 25, 2006

Convex sets form the basis of convex analysis and optimization which is a very important topic for machine learning and statistics. It is important because it leads to computationally efficient ways of solving a wide class of optimization problems and because it may be used to derive powerful theoretical results. In this note we concentrate on the most basic layer of convex analysis - convex sets. Despite the fact that we assume that the embedding space is $(\mathbb{R}^n, \|\cdot\|_2)$, much of what we say here can be generalized to Banach spaces (with some added complexity for non-reflexive spaces).

The affine hull of a set $C \subset \mathbb{R}_n$ is defined as $\text{aff}(C) = \{\sum_{i=1}^k \theta_i x_i : x_1, \ldots, x_n \in C, \sum \theta_i = 1\}$. It is also the smallest affine set (the zero set of an affine function which is a linear function plus a constant) that contains C. The relative interior of C is the interior of C in the subset topology of $\text{aff}(C) \subset \mathbb{R}^n$.

Example: consider $C = \{(0, y) : y \in [1, 2]\} \subset \mathbb{R}^2$. The affine hull of C is the one dimensional line $\{(0, y) : y \in \mathbb{R}\}$ and the relative interior of C is $C = \{(0, y) : y \in (1, 2)\} \subset \mathbb{R}^2$.

A set $C \subset \mathbb{R}^n$ is convex if every line segment between two points in C is in C i.e. $\forall x, y \in C$, $\forall \theta \in [0, 1]$, $\theta x + (1 - \theta)x \in C$. We refer to $\sum \theta_i x_i$ as an affine combination of $x_1, \ldots, x_n$ if $\sum x_i = 1$ and a convex combination if in addition $x_i \geq 0$. It can be shown that a set is convex iff it contains all its convex combinations. The convex hull of a set C is the set of all convex combinations of points in C: $\text{conv}(C) = \{\sum \theta_i x_i : x_1, \ldots, x_k \in C, \theta_i \geq 0 \forall i, \sum \theta_i = 1\}$.

Example: The simplex $\mathbb{P}_n = \{x \in \mathbb{R}^{n+1} : \forall i\, x_i > 0, \sum x_i = 1\}$ is the subset of $\mathbb{R}^{n+1}$ that contains the set of all possible positive probabilities on a set of $n + 1$ elements. Its affine hull is $\mathbb{R}^{n+1}$ and its convex hull is its closure $\overline{\mathbb{P}}_n$. Every convex combination of elements in the simplex represents a mixture model from a probabilistic point of view.

A set C is a cone if for every $x \in C$ and $\theta \geq 0$, $\theta x \in C$. If it is also convex it is called a convex cone in which case, the definition is equivalent to $\forall x, y \in C$, $\forall \theta_1, \theta_2 \geq 0$, $\theta_1 x + \theta_2 y \in C$. Such a combination is called a conic combination and it can be used to define the conic hull, in a similar way to the affine hull and convex hull.

A hyperplane in $\mathbb{R}^n$, characterized by $\alpha \in \mathbb{R}^n, \beta \in \mathbb{R}$ is a set $\{x \in \mathbb{R}^n : \alpha^\top x = \beta\}$. A hyperplane divides the space $\mathbb{R}^n$ into two half spaces $\{x : \alpha^\top x < \beta\}$, $\{x : \alpha^\top x > \beta\}$ (and the hyperplane itself which is the decision boundary). β represents the vertical offset of the hyperplane and α represents the normal vector, translated to the origin.

Example: The set of symmetric positive semi-definite matrices is a convex cone since a conic combination of spd matrices is spd as well as $v^\top(\alpha H + \beta T)v = \alpha\, v^\top H v + \beta\, v^\top H v$.

Listed below are some operations that preserve convexity of sets

- Intersection: An arbitrary intersection of convex sets is a convex set.
- Scaling: If C is convex, $aC = \{ax : x \in C\}$ is convex.
- Translation: If C is convex, $C + a = \{x + a : x \in C\}$ is convex.
- Affine transformation: If f is an affine function (a sum of a linear function and a constant) and C convex, then $\{f(x) : x \in C\}$ is convex.
- Sum: If C, D are convex than $C + D = \{x + y : x \in C, y \in D\}$ is convex.

We now turn to explore some relationship between convex sets and cones and hyperplanes.

For every two disjoint convex sets C, D it can be shown that there is a hyperplane (α, β) that separates them i.e. $\forall x \in C$, $\alpha^\top x + \beta \leq 0$ and $\forall x \in D$, $\alpha^\top x + \beta \geq 0$.

A hyperplane (α, β) is a said to be a supporting hyperplane for a set C (not necessarily convex) if $\forall x \in C, \alpha^\top x + \beta \leq 0$ and $\alpha^\top x' + \beta = 0$ for some x' on the boundary of C. It can be shown that for every point x on the boundary of an arbitrary convex set there exists a supporting hyperplane.