Chapter 1

Convex optimization

1.1 Introduction

Convex optimization is important for two reasons: a) the local minimum is also the global minimum (there is only one minimum) and b) there are efficient algorithms (that is polynomial time) to find the minimum values. So, if a learning problem can be formulated as a convex optimization problem then it can be used for large data sets. Non-convex optimization problems are in general hard - that is NP-hard. Many learning problems can be formulated as convex optimization problems.

The material here is meant as a refresher and compiles the relevant definitions and results. There are no examples or proofs. For more details, examples and proofs of results see the book by Boyd, Vandenberghe, 'Convex optimization'. An online copy is available.

1.2 Convex sets

We start with some definitions related to convex sets and functions as they are basic for any convex analysis:

Definition 1 (Line)**.** *A line through* $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$*, written* $L[\mathbf{x}_1, \mathbf{x}_2]$*, is the set of points defined by* $\{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{x} = \theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2,\ \theta \in \mathbb{R}\}$

Note that a line has infinite extent in both directions.

Definition 2 (Line segment)**.** *A line segment through* $x_1, x_2 \in \mathbb{R}^n$*, written* $l[\mathbf{x}_1, \mathbf{x}_2]$*, is the set of points defined by* $\{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{x} = \theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2, 0 \le \theta \le 1\}$

A line segment has finite extent and exists only between the points $\mathbf{x}_1, \mathbf{x}_2$.

Definition 3 (Convex set)**.** S *is a convex set if for any distinct* $\mathbf{x}_1, \mathbf{x}_2 \in S$*,* $l[\mathbf{x}_1, \mathbf{x}_2] \in S$*. Equivalently,* S *is convex if for* $0 \le \theta \le 1$ *and* $\mathbf{x}_1, \mathbf{x}_2 \in S$ $\mathbf{x} = \theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in S$*.*

S *is strictly convex if all points of* $l[\mathbf{x}_1, \mathbf{x}_2]$ *except the end points* $\mathbf{x}_1, \mathbf{x}_2$ *are in* S *or we have* $0 < \theta < 1$ *in the definition of the convex set.*

Definition 4 (L_p ball)**.** *The* L_p *ball* $B_p[\mathbf{x}_0, r]$*, is the set of points defined by* $\{\mathbf{x} \,|\, \|\mathbf{x} - \mathbf{x}_0\|_p \le r\}$*, where* $\|x\|_p = (\sum_{i=1}^d |x|^p)^{1/p}$*.*

Definition 5 (Euclidean ball)**.** *A Euclidean ball is an L_p ball where $p = 2$, that is, all points in the set $B_2[\mathbf{x}_0, r] = \{\mathbf{x} \,|\, \|\mathbf{x} - \mathbf{x}_0\|_2 \leq r\}$. We will write $B()$ instead of $B_2()$.*

Definition 6 (Half-space)**.** *A half-space $HS(\mathbf{w}, w_0) = \{\mathbf{x} \in \mathbb{R}_n \,|\, \mathbf{w}^T\mathbf{x} \leq w_0\}$, that is the set of points on one side of the hyper-plane $\mathbf{w}^T + w_0 = 0$.*

Definition 7 (Cone)**.** *A cone C is the set of points defined by $\{\theta\mathbf{x} \,|\, \theta \geq 0,$ and $\mathbf{x} \in C\}$.*

Definition 8 (Affine set)**.** *A set S is an affine set if for any $\mathbf{x}_1, \mathbf{x}_2 \in S$ the line through $\mathbf{x}_1, \mathbf{x}_2$ is in S - that is $L[\mathbf{x}_1, \mathbf{x}_2] \in S$.*

Definition 9 (Convex hull)**.** *The convex hull of set S, written $Cnvxh[S]$, $Cnvxh[S] = \{\mathbf{x} \,|\, \mathbf{x} = \theta_1\mathbf{x}_1 + \ldots + \theta_k\mathbf{x}_k,\, \mathbf{x}_i \in S, \theta_i \geq 0, i = 1..k$ and $\sum_{i=1}^{k} \theta_i = 1\}$.*

A convex hull is some times defined as the set of all convex combinations of elements in S. A convex combination is what is defined for $\mathbf{x}$above.

Definition 10 (Affine hull)**.** *The affine hull of the set S is the smallest affine set containing S. Or equivalently it is the intersection of all affine sets that contain S. It is defined as the set of points $Affh[S] = \{\mathbf{x} = \theta_1\mathbf{x}_1, \ldots, \theta_k\mathbf{x}_k \,|\, \mathbf{x} \in S,\, \theta_i \in \mathbb{R}$ and $\sum_{i=1}^{k} \theta_i = 1\}$.*

An affine hull is also defined as the set of all affine combinations of elements of S. An affine combination is the one defined for $\mathbf{x}$ above. Note that the only difference between an affine combination and convex combination is the restriction ($\theta \geq 0$) placed on θ.

Definition 11 (Conic hull)**.** *A conic hull of a set S, $Coneh[S]$, is the set of points $Coneh[S] = \{\mathbf{x} \,|\, \mathbf{x} = \theta_1\mathbf{x}_1 + \ldots. + \theta_k, \theta_i \geq 0, i = 1..k$ and $\mathbf{x}_i \in S\}$.*

The conic hull of S is the intersection of all convex cones that contain S. The conic hull is a convex set.

Definition 12 (Boundary)**.** *The boundary of a set S, written $\delta[S]$ are all points $\mathbf{x}$ such that $\forall\epsilon > 0$, $S \cap B[\mathbf{x}, \epsilon] \neq \emptyset$ and $S^c \cap B[\mathbf{x}, \epsilon] \neq \emptyset$, where S^c is the complement of set S. $\mathbf{x}$ is often called a boundary point.*

Definition 13 (Interior, relative interior)**.** *The interior of a set S, written $int[S]$, are all points $\mathbf{x}$ such that $\exists\epsilon > 0$ and $B[\mathbf{x}, \epsilon] \subset S$. $\mathbf{x}$ is called an interior point of S.*

The relative interior of S, written $rint[S]$, is all those $\mathbf{x} \in int[S]$ such that the $B[\mathbf{x}, \epsilon]$ restricted to $Affh[S]$ is a subset of S - that is $B[\mathbf{x}, \epsilon] \cap Affh[S] \subset S$.

To see why we need $rint[S]$ consider a line segment in 2D. Then $int[S]$ is $\emptyset$ because any ball with its centre on the line segment will always contain points in S^c. But if we consider $rint[S]$ then it is non-empty and contains all points in S except the end points since the affine hull of S is the line containing S.

Definition 14 (Closure)**.** *The closure of S, $Closure[S] = S \cup \delta[S]$. S is said to be closed if $\delta[S] \subset S$. It is open if $S \cap \delta[S] = \emptyset$. S is compact if it is closed and bounded in $\mathbb{R}^d$.*

1.2.1 Examples of convex sets, representation, showing a set is convex

The following are some examples of convex sets:

- Lines, line segments, hyper-plane, half-spaces, norm balls (for $p \geq 1$).

- $\mathbb{R}^n$, affine spaces[1].
- Polygons, polyhedra.
- $int[S]$ where S is a convex set.
- S^{n+} the set of all positive semi-definite matrices (psd matrices). [2]

There are two representations for convex sets - S. The primal representation of S is by specifying the points in the set in some way. The number of points could be finite (e.g. polygons, polyhedra) or infinite (e.g. circle). The dual representation which represents it as a set of half planes containing S. Actually, S (a closed convex set) is the intersection of all the closed half-spaces that contain S.

To show that a set is convex we can use the following strategies:

1. Use the definition of convexity.
2. Construct the set from other convex set using convexity preserving operations. Let $S, T \subset \mathbb{R}^n$, $a \in \mathbb{R}$, $\mathbf{b} \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ then the following are convex function:
   (a) $S + \mathbf{b} = \{\mathbf{x} + \mathbf{b} \,|\, \mathbf{x} \in S\}$.
   (b) $aS = \{a\mathbf{x} \,|\, \mathbf{x} \in S\}$.
   (c) $S + T = \{\mathbf{x} + \mathbf{y} \,|\, \mathbf{x} \in S, \mathbf{y} \in T\}$.
   (d) $S \times T = \{(\mathbf{x}, \mathbf{y}) \,|\, \mathbf{x} \in S, \mathbf{y} \in T\}$.
   (e) $AS + \mathbf{b} = \{A\mathbf{x} + \mathbf{b} \,|\, \mathbf{x} \in S\}$.
3. Represent it as a convex hull. A convex set is equal to its convex hull.
4. Represent it as an intersection of half-planes.

1.2.2 Properties of convex sets

Below we state some properties of convex sets that are useful in the context of classification - especially linear classifiers.

Theorem 1 (Separating hyper-plane). *Let S, T be disjoint convex sets ($S \cap T = \emptyset$) then $\exists \mathbf{w} \in \mathbb{R}^n$, $w_0 \in \mathbb{R}$ such that $\forall \mathbf{x} \in S$, $\mathbf{w}^T\mathbf{x} \leq w_0$ and $\forall \mathbf{x} \in T$, $\mathbf{w}^T\mathbf{x} \geq w_0$. That is there is a hyper-plane (symbolized by $hp[\mathbf{w}, w_0]$) that separates S and T.*

Definition 15 (Strong separation). *Convex sets S and T are strongly separated by $hp[\mathbf{w}, w_0]$ if $\exists \epsilon > 0$ such that $\mathbf{w}^t(S + B[0, \epsilon]) < w_0$ and $\mathbf{w}^t(T + B[0, \epsilon]) > w_0$. This says that S and T are strongly separable when they remain separable after being displaced by a small amount ϵ.*

Theorem 2 (Strong separation). *Convex sets S and T are strongly separable iff the distance between any $\mathbf{x} \in S$ and $\mathbf{y} \in T$ is greater than 0.*

[1] If $A \in \mathbb{R}^{m \times n}, \mathbf{b} \in \mathbb{R}^m$ then $S = \{\mathbf{x} \in \mathbb{R}^n \,|A\mathbf{x} = \mathbf{b}\}$ is an affine space, that is all vectors $\mathbf{x}$ that can be transformed to a fixed vector $\mathbf{b}$.

[2] $A \in \mathbb{R}^{n \times n}$ is a psd matrix if A is symmetric and $\forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x}^T A\mathbf{x} \geq 0$.

Theorem 3 (Supporting hyper-plane)**.** *For any boundary point* $\mathbf{x} \in \delta[S]$ *, there exists a* $hp[\mathbf{x}, w_0]$ *that lies on one side of the set* S*.*

Theorem 4 (A characterization of convex set)**.** *If* S *is closed with* $int[S] \neq \emptyset$ *and there exists a supporting hyper-plane (previous theorem) for all* $\mathbf{x} \in \delta[S]$ *then* S *is convex.*