31 Jan 2022

Convexity and Gradient Descent

Announcement:

TA office hours begin today.
See office hours calendar on website.

https://cs.cornell.edu/courses/cs4850/2022sp

Fill out OH modality poll. (Pinned post on Ed.)
Write an Ed post requesting CMS access if you don’t have it yet.

Def. If $x_1, \ldots, x_m$ are vectors in vector space $V$, an affine combination is any linear combination

$$a_1 x_1 + \cdots + a_m x_m$$

such that $a_1 + a_2 + \cdots + a_m = 1$.

A convex combination is an affine combination with $a_1, \ldots, a_m \geq 0$. (weighted average)

A subset of $V$ is convex if it is closed under taking convex combinations. (Suffices to just test that the line segment joining any 2 vectors in the set remains in the set.)
Ex. A halfspace in $V$ is a set of the form

$$H = \{ x | f(x) \leq \theta \}$$

where $f \in V^*$, $f \neq 0$. Equivalently, if $V$ has a non-degenerate inner product,

$$H = \{ x | \langle w, x \rangle \leq \theta \}$$

where $w \neq 0$.

Prop. In a finite dimensional vector space, a closed subset $K$ is convex if and only if it is representable as an intersection of (potentially infinitely many) halfspaces.

Def. A function $h : K \to \mathbb{R}$ is convex if

$$h(tx + (1-t)y) \leq t h(x) + (1-t) h(y) \quad \forall x, y \in K$$
Def: For $K\subseteq V$ convex, $h: K \to \mathbb{R}$,

1. The epigraph of $h$ is
   $$\{(x,y) \in V\times \mathbb{R} \mid y \geq h(x)\}$$

2. The subdifferential of $h$ at $x \in K$ is the subset of $V^*$ defined as
   $$\partial h(x) = \{f \in V^* \mid h(y) \geq h(x) + f(y-x), \forall y \in K\}$$

   - $\partial h(\emptyset) = \{f \in \mathbb{R}^* \mid \forall y \ |y| \geq f(y)\}$
   - $= \{f(x) = ax \mid -1 \leq a \leq 1\}$.

   - $\partial h(1) = \{f \in \mathbb{R}^* \mid \forall y \ |y| \geq 1 + f(y-1)\}$
   - $= \{f(x) = x^2\}$.

   $\partial h(x) = \text{"set of slopes of supporting hyperplanes"}
   \text{to the graph of } h \text{ at } (x, h(x)).$
Theorem. (Proved in lecture notes)
For K ε V convex and \( h: K \rightarrow \mathbb{R} \)

The following are equivalent.

(i) \( h \) is a convex function
(ii) The epigraph of \( h \) is a convex set.
(iii) The subdifferential of \( h \) is nonempty at every point.

**Differentiable Functions**

Def. The norm of a vector \( \mathbf{x} \) in a space with pos. def. inner product is \( \| \mathbf{x} \| = \langle \mathbf{x}, \mathbf{x} \rangle^{\frac{1}{2}} \).

E.g., in \( \mathbb{R}^n \) with standard inner prod,
\[
\| \mathbf{x} \| = (x_1^2 + \cdots + x_n^2)^{\frac{1}{2}} = \text{Euclidean length of } \mathbf{x} \\
= \text{2-norm of } \mathbf{x}.
\]

Def. A function \( g: V \rightarrow \mathbb{R} \) vanishes to first order at \( \mathbf{0} \) if \( \forall \varepsilon > 0 \exists \delta > 0 \) st.
\[
\frac{g(\mathbf{x})}{\| \mathbf{x} \|} < \varepsilon \quad \text{whenever } \| \mathbf{x} \| < \delta.
\]
Def: \( f : V \to \mathbb{R} \) is differentiable at \( x \) if there exists an element of \( V^* \) called the **differential** of \( f \) denoted \( df_x \), such that

\[
\forall y \quad f(x+y) = f(x) + df_x(y) + \theta(y)
\]

where \( \theta \) vanishes to 1st order at 0.

If \( h \) is convex and differentiable at \( x \) then

\[
\nabla h(x) = \left\{ \frac{dh}{dx} \right\}
\]

For differentiable \( f : \mathbb{R}^n \to \mathbb{R} \) the differential is

\[
\frac{df}{dx} = \left[ \frac{\partial f}{\partial x_1} \middle| \frac{\partial f}{\partial x_2} \middle| \ldots \frac{\partial f}{\partial x_n} \right].
\]

The gradient of \( F \) is defined when \( V \) has a non-degenerate inner product and \( f \) is differentiable at \( x \). Then \( \nabla f_x \) is the image of \( df_x \) under the isomorphism \( V^* \to V \).
Ex. $\mathbb{R}^n$ with standard inner prod.

$$(\mathbb{R}^n)^* = \text{row vectors length } n$$

$\mathbb{R}^n = \text{col vectors}$

Isomorphism $= \text{transpose}$

$$\nabla f_x = \left[ \begin{array}{c} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{array} \right]$$