CONVEX OPTIMIZATION PROBLEMS FOR MACHINE LEARNING AND BASIC PROPERTIES OF CONVEXITY

Jianqiang Gao\textsuperscript{a}, Bin Guo\textsuperscript{b} and S. Balan\textsuperscript{c}

\textsuperscript{a}College of Computer and Information, Hohai University, Nanjing 210098, P. R. China
\textsuperscript{b}College of Computer and Information Engineering, Xinjiang Agricultural University, Xinjiang 830052, P. R. China
\textsuperscript{c}Government Arts College (Autonomous), Bharathiar University, Coimbatore 641018, India

Abstract

Some convex optimization problems for machine learning and basic properties of convexity are introduced in this paper. The main purpose is to help readers understand the essence of the convex optimization for machine learning.

Keywords: convex optimization problems, machine learning, basic properties.

\textsuperscript{*}Corresponding author.

E-mail address: dr.jq.gao@gmail.com (Jianqiang Gao), gb@xjau.edu.cn (Bin Guo) and balan.sethumalingam@gmail.com (S. Balan).

Copyright © 2015 Scientific Advances Publishers
2010 Mathematics Subject Classification: 52A, 49K30, 68U.
Received December 3, 2015
1. Introduction

The central objects of our study are convex functions and convex sets in $\mathbb{R}^n$.

**Definition 1** (Convex sets and convex functions). A set $\chi \in \mathbb{R}^n$ is said to be convex if it contains all of its segments, that is,

$$\forall(x, y, \gamma) \in \chi \times \chi \times [0, 1], \ (1 - \gamma)x + \gamma y \in \chi.$$  

A function $f : \chi \rightarrow \mathbb{R}$ is said to be convex if it always lies below its chords, that is,

$$\forall(x, y, \gamma) \in \chi \times \chi \times [0, 1], \ f((1 - \gamma)x + \gamma y) \leq (1 - \gamma)f(x) + \gamma f(y).$$  

We are interested in algorithms that take as input a convex set $\chi$ and a convex function $f$ and output an approximate minimum of $f$ over $\chi$. We write compactly the problem of finding the minimum of $f$ over $\chi$ as

$$\min f(x), \text{ s.t. } x \in \chi.$$  

In the following, we will make more precise how the set of constraints $\chi$ and the objective function $f$ are specified to the algorithm. Before that we proceed to give a few important examples of convex optimization problems in machine learning.

2. Some Convex Optimization Problems for Machine Learning

Many fundamental convex optimization problems for machine learning take the following form:

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{m} f_i(x) + \lambda R(x),$$  

where the functions $f_1, \ldots, f_m, R$ are convex and $\lambda \geq 0$ is a fixed parameter. The interpretation is that $f_i(x)$ represents the cost of using $x$ on the $i$-th element of some data set, and $R(x)$ is a regularization term.
which enforces some “simplicity” in \( x \). We discuss now major instances of Equation (1). In all cases, one has a data set of the form \((w_i, y_i) \in \mathbb{R}^n \times Y, i = 1, \ldots, m\) and the cost function \( f_i \) depends only on the pair \((w_i, y_i)\).

We refer to [1, 2] for more details on the origin of these important problems.

In classification one has \( Y = \{-1, +1\} \). Taking \( f_i(x) = \max(0, 1 - y_i x^T w_i) \) and \( R(x) = \|x\|_2^2 \) one obtains the SVM problem. On the other hand, taking \( f_i(x) = \log(1 + \exp(-y_i x^T w_i)) \) and again \( R(x) = \|x\|_2^2 \), one obtains the logistic regression problem.

In regression one has \( Y = \mathbb{R} \). Taking \( f_i(x) = (x^T w_i - y_i)^2 \) and \( R(x) = 0 \) one obtains the vanilla least-squares problem which can be rewritten in vector notation as

\[
\min_{x \in \mathbb{R}^n} \|Wx - Y\|_2^2, \tag{2.2}
\]

where \( W \in \mathbb{R}^{m \times n} \) is the matrix with \( w_i^T \) on the \( i \)-th row and \( Y = (y_1, \ldots, y_n)^T \). With \( R(x) = \|x\|_2^2 \), one obtains the ridge regression problem, while with \( R(x) = \|x\|_1 \) this is the LASSO problem.

Here our data set consists of observations of some of the entries of an unknown matrix \( Y \), and we want to “complete” the unobserved entries of \( Y \) in such a way that the resulting matrix is “simple”. After some massaging [3], the matrix completion problem can be formulated as follows:

\[
\min_{X \in \mathbb{R}^{n \times n}, X^T = X, X \geq 0, X_{i,j} = Y_{i,j}, \text{ for } (i,j) \in \Omega} \text{Tr}(X), \tag{2.3}
\]

where \( \Omega \subset [n]^2 \) and \((Y_{i,j})_{i,j} \in \Omega \) are given.
3. Basic Properties of Convexity

A basic result about convex sets that we shall use extensively is the separation theorem.

**Theorem 1** (Separation theorem). Let \( \chi \subset \mathbb{R}^n \) be a closed convex set, and \( x_0 \in \mathbb{R}^n \setminus \chi \). Then, there exists \( w \in \mathbb{R}^n \) and \( t \in \mathbb{R} \) such that
\[
w^T x_0 < t, \quad \forall x \in \chi, w^T x \geq t.
\]
(2.4)

Note that if \( \chi \) is not closed, then one can only guarantee that \( w^T x_0 \leq w^T x, \forall x \in \chi \) (and \( w \neq 0 \)). This immediately implies the supporting hyperplane theorem:

**Theorem 2** (Supporting hyperplane theorem). Let \( \chi \subset \mathbb{R}^n \) be a convex set, and \( x_0 \in \partial \chi \). Then, there exists \( w \in \mathbb{R}^n, w \neq 0 \) such that
\[
\forall x \in \chi, w^T x \geq w^T x_0.
\]
(2.5)

We introduce now the key notion of subgradients.

**Definition 2** (Subgradients). Let \( \chi \subset \mathbb{R}^n \), and \( f : \chi \to \mathbb{R} \). Then \( g \in \mathbb{R}^n \) is a subgradient of \( f \) at \( x \in \chi \) if for any \( y \in \chi \) one has
\[
f(x) - f(y) \leq g^T(x - y).
\]
(2.6)

The set of subgradients of \( f \) at \( x \) is denoted \( \partial f(x) \). The next result shows that a convex functions always admit subgradients.

**Proposition 1** (Existence of subgradients). Let \( \chi \subset \mathbb{R}^n \) be convex and \( f : \chi \to \mathbb{R} \). If \( \forall x \in \chi, \partial f(x) \neq \emptyset \), then \( f \) is convex. Conversely, if \( f \) is convex, then for any \( x \in \text{int}(\chi), \partial f(x) \neq \emptyset \). Furthermore, if \( f \) is convex and differentiable at \( x \), then \( \nabla f(x) \in \partial f(x) \).
Before going to the proof, we recall the definition of the epigraph of a function \( f : \mathcal{X} \to \mathbb{R} \):

\[
\text{epi}(f) = \{(x, t) \in \mathcal{X} \times \mathbb{R} : t \geq f(x)\}. \quad (2.7)
\]

It is obvious that a function is convex if and only if its epigraph is a convex set.

**Proof.** The first claim is almost trivial: Let \( g \in \partial f((1 - \gamma)x + \gamma y) \), then by definition one has

\[
f((1 - \gamma)x + \gamma y) \leq f(x) + \gamma g^T(y - x), \quad (2.8)
\]

\[
f((1 - \gamma)x + \gamma y) \leq f(y) + (1 - \gamma)g^T(x - y), \quad (2.9)
\]

which clearly shows that \( f \) is convex by adding the two inequalities.

Let us prove that a convex function \( f \) has subgradients in the interior of \( \mathcal{X} \). We build a subgradient by using a supporting hyperplane to the epigraph of the function. Let \( x \in \mathcal{X} \). Then clearly \((x, f(x)) \in \text{epi}(f)\), and \( \text{epi}(f) \) is a convex set. Thus by using the “supporting hyperplane theorem”, there exists \((a, b) \in \mathbb{R}^n \times \mathbb{R}\) such that

\[
a^T x + bf(x) \geq a^T y + bt, \quad \forall (y, t) \in \text{epi}(f). \quad (2.10)
\]

Clearly, by letting \( t \) tend to infinity, one can see that \( b \leq 0 \). Now let us assume that \( x \) is in the interior of \( \mathcal{X} \). Then for \( \varepsilon > 0 \) small enough, \( y = x + \varepsilon a \in \mathcal{X} \), which implies that \( b \) cannot be equal to 0. Thus rewriting Equation (10) for \( t = f(y) \), one obtains

\[
f(x) - f(y) \leq \frac{1}{|b|} a^T(x - y). \quad (2.11)
\]

Thus \( a/|b| \in \partial f(x) \) which concludes the proof of the second claim.
Finally, let \( f \) be a convex and differentiable function. Then by definition

\[
\begin{align*}
  f(y) & \geq f((1 - \gamma)x + \gamma y) - (1 - \gamma)f(x) \\
  \gamma & = f(x) + f(x + \gamma(y - x)) - f(x) \\
  \to f(x) + \nabla f(x)^T(y - x), \text{ (when } \gamma \to 0). 
\end{align*}
\]  

(2.12)

In several cases of interest, the set of constraints can have an empty interior, in which case the above proposition does not yield any information. However, it is easy to replace \( \text{int}(\chi) \) by \( \text{ri}(\chi) \)-the relative interior of \( \chi \)-which is defined as the interior of \( \chi \) when we view it as subset of the affine subspace it generates. Other notions of convex analysis will prove to be useful in some parts of this text. In particular, the notion of closed convex functions is convenient to exclude pathological cases: these are the convex functions with closed epigraphs. Sometimes, it is also useful to consider the extension of a convex function \( f : \chi \to \mathbb{R} \) to a function from \( \mathbb{R}^n \) to \( \mathbb{R} \) by setting \( f(x) = +\infty \) for \( x \notin \chi \). In convex analysis, one uses the term proper convex function to denote a convex function with values in \( \mathbb{R} \cup \{+\infty\} \) such that there exists \( x \in \mathbb{R}^n \) with \( f(x) < +\infty \). The readers can refer [4].

References


