Optimization Theory for Machine Learning

ESC 22 Spring

Kim Dueun
Department of Applied statistics

Table of contents

Statistical estimation

Statistical estimation

Definition (Maximum likelihood estimator)

$$\hat{\theta} = \arg\max_{\theta} L(\theta|\mathbf{X})$$

$\theta \in \mathbb{R}^n$ is parameters of distribution.
$\mathbf{X} \in \mathbb{R}^m$ is a random vector.
$L(\theta|\mathbf{X}) = \prod_{i=1}^{N} f(\mathbf{x}_i|\theta)$, where $f(\mathbf{x}|\theta)$ is a density of $\mathbf{X}$.

Given N iid samples of $\{\mathbf{X}_i\}$, find $\hat{\theta}$ that maximizes likelihood function.

In addition to this, many statistical problems can be expressed as optimization problems with prior information as constraints.

Unconstrained minimization

Unconstrained minimization

Definition (Unconstrained minimization)

$$\textit{minimize } f(x)$$

Where $f : \mathbb{R}^n \to \mathbb{R}$ is convex and twice continuously differentiable.

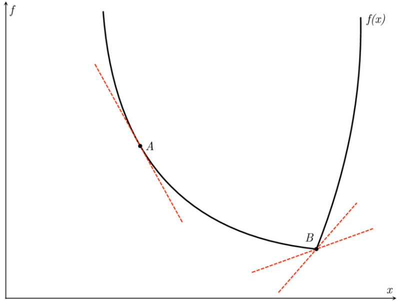
There exists an optimal point x^* and optimal value $p^* = \inf_x f(x)$ where $\nabla f(x^*) = 0$.

We will learn how to find a solution numerically, when we cannot derive the solution analytically.

- keyword: descent, gradient descent, steepest descent, newton's method

Subgradient methods

Subgradient methods



How can we find solution when objective function is non-differentiable?
We say a vector $g \in \mathbb{R}^n$ is a subgradient of $f : \mathbb{R}^n \to \mathbb{R}$ at $x \in \textbf{dom}\, f$ if for all $z \in \textbf{dom}\, f$,

$$f(z) \geq f(x) + g^T(z - x).$$

There are various extensions, such as with duality, to speed up convergence or handle constraints.

Robust optimization

Robust optimization

- Suppose we have data with some uncertainty. For example, when there are outliers or missing values in the data, or when there is incorrectly classified data.
- Robust optimization is an important sub-field of optimization that deals with uncertainty in the data of optimization problems. Under this framework, the objective and constraint functions are only assumed to belong to certain sets in function space (the so-called "uncertainty sets").

$$
\begin{array}{ll}
\text{minimize} & f_0(\mathbf{x}) \\
\text{subject to} & f_i(\mathbf{x}, u_i) \leq 0, u_i \in \mathcal{U}
\end{array}
$$

- Over two weeks, we will first take a closer look at the concept and then apply it to machine learning problems.