Statistical Learning and Machine Learning

Lecture 19 - Introduction to Mathematical Optimization and Lagrangian

November 5, 2021

What will be presented today?

- Mathematical Optimization Problem
- Convex Optimization Problem
- Lagrange Multipliers and Dual Functions
- KKT Conditions
- Example: Robust classification

Mathematical Optimization Problem

$$
\begin{aligned}
\underset{x}{\text{minimize}} \quad & f_0(x) \\
\text{subject to} \quad & f_i(x) \leq b_i, \ \text{i=1,...,m}, \\
& h_j(x) = 0, \ \text{j=1,...,p}
\end{aligned}
$$

$x = (x_1, \ldots, x_n)$ is *optimization variable*
$f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$ is *objective function*
$f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ are *inequality constraint functions*
$h_j : \mathbb{R}^n \rightarrow \mathbb{R}$ are *equality constraint functions*

x^* is the solution of the above problem, if it has the smallest objective value among all possible x that satisfy the constraints.

Examples/Applications

- Data fitting (e.g., Least squares)
  1. variables: model parameters
  2. constraints: prior information, parameter limits
  3. objective: prediction error
- Device sizing in electronic circuits
  1. variables: device width and length
  2. constraints: manufacturing limits, timing requirements, maximum area
  3. objective: power consumption of the device
- Density Estimation (e.g. Maximum Likelihood)
- Supervised classification (e.g., SVM)

Solving Optimization Problem

In general, optimization problem are

- hard to solve
- methods involve some compromise, e.g., very long computation time, or lack of guarantees in terms of reaching right solution

exceptions: classes of problems that can be solved accurately and efficiently

- least-squares problems
- linear programming problems
- convex optimization problems

Convex Optimization Problem

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f_0(x) \\ \text{subject to} & f_i(x) \leq b_i, \ \text{i=1,...,m} \end{array}$$

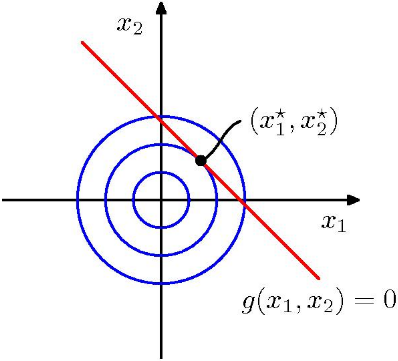
- objective and constraint functions are convex

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y)$$

  if $\alpha + \beta = 1$ and $\alpha \geq 0$, $\beta \geq 0$
- least-squares problems and linear programs are special cases
- Properties:
  1. every local minimum is a global minimum
  2. the optimal set is convex
  3. for strictly convex objective functions, at most one optimal point

Illustration of Convex Problem with Constraint

$$\begin{aligned} \text{maximize} \quad & 1 - x_1^2 - x_2^2 \\ \text{subject to} \quad & x_1 + x_2 - 1 = 0 \end{aligned}$$



Lagrangian

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \text{ i=1,...,m}, \\ & h_i(x) = 0, \text{ i=1,...,p} \end{aligned}$$

variable $x \in \mathbb{R}^n$, optimal value p^*

Lagrangian: $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- L: weighted sum of objective and constraint functions
- λ_i: Lagrange multipliers associated with f_i
- ν_i: Lagrange multipliers associated with h_i

Lagrange Dual Function

Lagrange dual function $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in D} L(x, \lambda, \nu) \\ &= \inf_{x \in D} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right) \end{aligned}$$

g is a concave function.

Lower bound property: if $\{\lambda_i \geq 0\}$ for all i, then $g(\lambda, \nu) \leq p^*$.

The Dual Problem

$$\begin{aligned} &\text{maximize} && g(\lambda, \nu) \\ &\text{subject to} && \lambda_i \geq 0, \ \text{i=1,...,m} \end{aligned}$$

- finds best lower bound on p, obtained from Lagrange dual function
- it is a convex optimization problem with optimal value d^*
- λ, ν are dual feasible if $\lambda \geq 0$, $(\lambda, \nu) \in$ **dom** g
- **weak duality**: $d^* \leq p^*$ (always holds)
- **strong duality**: $d^* = p^*$ (usually holds for convex problems only)

Example

Original Problem:

$$\begin{aligned} \text{maximize} \quad & 1 - x_1^2 - x_2^2 \\ \text{subject to} \quad & x_1 + x_2 - 1 = 0 \end{aligned}$$

Dual Problem:

$$\begin{aligned} \text{maximize} \quad & \lambda^2 - 2\lambda + 2 \\ \text{subject to} \quad & \lambda \geq 0 \end{aligned}$$

Karush-Kuhn-Tucker (KKT) conditions I

The KKT conditions (assuming differentiable f_i, h_i):

- primal constraints: $f_i(x) \leq 0, i = 1, ..., m$ and $h_i(x) = 0, i = 1, ..., p$
- dual constraints: $\lambda_i \geq 0$, $\forall i$
- complementary slackness: $\lambda_i f_i(x) = 0$, $\forall i$
- gradient of Lagrangian with respect to x vanishes:

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$$

Necessary Condition: If strong duality holds ($d^* = p^*$) and x, λ, ν are optimal, then they must satisfy the KKT conditions.

KKT conditions for convex problems: if $\hat{x}$, $\hat{\lambda}$, $\hat{\nu}$ satisfy KKT for a convex problem, then they are optimal.

Karush-Kuhn-Tucker (KKT) conditions II

Example 1:

$$\begin{aligned} &\text{maximize} && 1 - x_1^2 - x_2^2 \\ &\text{subject to} && x_1 + x_2 - 1 = 0 \end{aligned}$$

Example 2:

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^n \log(x_i + \alpha_i) \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1 \end{array}$$

x is optimal iff $x \succeq 0$, $\mathbf{1}^T x = 1$, and there exist $\lambda \in \mathbf{R}^n$, $\nu \in \mathbf{R}$ such that

$$\lambda \succeq 0, \qquad \lambda_i x_i = 0, \qquad \frac{1}{x_i + \alpha_i} + \lambda_i = \nu$$

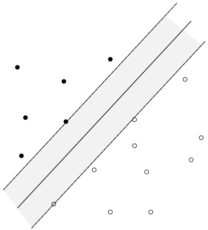
- if $\nu < 1/\alpha_i$: $\lambda_i = 0$ and $x_i = 1/\nu - \alpha_i$
- if $\nu \geq 1/\alpha_i$: $\lambda_i = \nu - 1/\alpha_i$ and $x_i = 0$
- determine ν from $\mathbf{1}^T x = \sum_{i=1}^n \max\{0, 1/\nu - \alpha_i\} = 1$

Source: Convex Optimization, Boyd and Vandenberghe, 2004

Robust Linear Classification and SVM I

Separating 2 sets of points, denoted by $\{x_1, \ldots, x_N\}$ and $\{y_1, \ldots, y_M\}$ via a hyperplane:

$$a^T x_i + b \geq 1, \quad i = 1, \ldots, N, \qquad a^T y_i + b \leq -1, \quad i = 1, \ldots, M,$$



Euclidean distance between the hyperplanes $\{z | a^T z + b = 1\}$ and $\{z | a^T z + b = -1\}$: $2/\|a\|_2$

Robust Linear Classification and SVM II

Robust linear classification

$$
\begin{aligned}
\underset{a,\, b}{\text{minimize}} \quad & (1/2)\|a\|_2 \\
\text{subject to} \quad & a^T x_i + b \geq 1, \quad i = 1, \ldots, N, \\
& a^T x_i + b \leq -1, \quad j = 1, \ldots, M
\end{aligned}
$$