Mathematical optimization

Mathematical **optimization problem**:

$$\begin{aligned} &\text{minimize} && f_0(\mathbf{x}) \\ &\text{subject to} && f_i(\mathbf{x}) \le \mathbf{b}_i, \quad i = 1, \cdots, m \end{aligned}$$

where

- $\mathbf{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n$: optimization variables
- $f_0 : \mathbb{R}^n \to \mathbb{R}$: objective function
- $f_i : \mathbb{R}^n \to \mathbb{R}$: constraint function

Optimal solution $\mathbf{x}^*$ has smallest value of f_0 among all vectors that satisfy the constraints.

Examples

- transportation - product transportation plan
- finance - portfolio management
- machine learning - support vector machines, graphical model structure learning

Transportation problem

We have a product that can be produced in amounts a_i at location i with $i = 1, \cdots, m$. The product must be shipped to n destinations, in quantities b_j to destination j with $j = 1, \cdots, n$. The amount shipped from origin i to destination j is x_{ij}, at a cost of c_{ij} per unit.

Transportation problem

We have a product that can be produced in amounts a_i at location i with $i = 1, \cdots, m$. The product must be shipped to n destinations, in quantities b_j to destination j with $j = 1, \cdots, n$. The amount shipped from origin i to destination j is x_{ij}, at a cost of c_{ij} per unit.

To find the transportation plan that minimizes the total cost, we solve an LP:

$$\begin{aligned}
\min \quad & \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} c_{ij} \\
\text{s. t.} \quad & \sum_{j=1}^{n} x_{ij} = a_i \quad i = 1, \cdots, m \\
& \sum_{i=1}^{m} x_{ij} = b_j \quad j = 1, \cdots, n \\
& x_{ij} \geq 0
\end{aligned}$$

Markowitz portfolio optimization

Consider a simple portfolio selection problem with n stocks held over a period of time:

- $\mathbf{x} = (x_1, \cdots, x_n)$: the optimization variable with x_i denoting the amount to invest in stock i
- $\mathbf{p} = (p_1, \cdots, p_n)$: a random vector with p_i denoting the reward from stock i. Suppose its mean μ and covariance matrix Σ are known.
- $r = \mathbf{p}^T\mathbf{x}$: the overall return on the portfolio. r is a random variable with mean $\mu^T\mathbf{x}$ and variance $x^T\Sigma x$.

Markowitz portfolio optimization

The Markowitz portfolio optimization problem is the QP

$$\begin{aligned} \min \quad & \mathbf{x}^T \Sigma \mathbf{x} \\ \text{s.t.} \quad & \mu^T \mathbf{x} \geq r_{\min} \\ & 1^T x = B \\ & x_i \geq 0, \quad i = 1, \cdots, n \end{aligned}$$

which find the portfolio that minimizes the return variance subject to three constraints:

- achieving a minimum acceptable mean return $r_{\min}$
- satisfying the total budget B
- no short positions ($x_i \geq 0$)

Support vector machines (SVMs)

Input: a set of training data,

$$\mathcal{D} = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbb{R}^p,\ y_i \in \{-1, 1\}, i = 1, \cdots, n\}$$

where y_i is either 1 or -1, indicating the class to which $\mathbf{x}_i$ belongs.

Problem: find the **optimal separating hyperplane** that separates the two classes and maximizes the distance to the closet point from either class.

Support vector machines (SVMs) 2

Define a hyperplane by $w^T x - b = 0$. Suppose the training data are linearly separably. So we can find w and b such that $w^T x_i - b \geq 1$ for all x_i from class 1 and $w^T x_i - b \leq -1$ for all x_i from class -1.

The distance between the two parallel hyperplans, $w^T x_i - b = 1$ and $w^T x_i - b = -1$, is $\frac{2}{\|w\|}$, called **margin**.

To find the optimal separating hyperplane, we choose w and b that maximize the margin:

$$\begin{aligned} \min \quad & \|w\|^2 \\ \text{s.t.} \quad & y_i(w^T x_i - b) \geq 1, \quad i = 1, \cdots, n \end{aligned}$$

Undirected graphical models

Input: a set of training data,

$$\mathcal{D} = \{(\mathbf{x}_i) \mid \mathbf{x}_i \in \mathbb{R}^p \; i = 1, \cdots, n\}$$

Assume the data were sampled from a Gaussian graphical model with mean $\mu \in \mathbb{R}^p$ and covariance matrix $\Sigma \in \mathbb{R}^{p \times p}$. The inverse covariance matrix, Σ^{-1}, encodes the structure of the graphical model in the sense that the variables i and j are connected only if the (i, j)-entry of Σ^{-1} is nonzero.

Problem: Find the maximum likelihood estimation of Σ^{-1} with a sparsity constraint, $\|\Sigma^{-1}\|_1 \leq \lambda$.

Undirected graphical models 2

Let S be the empirical covariance matrix:

$$S := \frac{1}{n}\sum_{k=1}^{n}(\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^T.$$

Denote $\Theta = \Sigma^{-1}$.

The convex optimization problem:

$$\begin{aligned} \min \quad & -\log\det\Theta + \operatorname{tr}(S\Theta) \\ \text{s.t.} \quad & \|\Theta\|_1 \leq \lambda \\ & \Theta \succ 0 \end{aligned}$$

Solving optimization problems

The optimization problem is in general difficult to solve: taking very long long time, or not always finding the solution

Exceptions: certain classes of problems can be solved efficiently:

- least-square problems
- linear programming problems
- convex optimization problems

Least-squares

$$\text{minimize} \quad \|Ax - b\|_2^2$$

where $x \in \mathbb{R}^n$, $b \in \mathbb{R}^k$ and $A \in \mathbb{R}^{k \times n}$.

- analytical solution: $x^* = (A^T A)^{-1} A^T b$ (assuming $k > n$ and **rank** $A = n$)
- reliable and efficient algorithms available
- computational time proportional to $n^2 k$, and can be further reduced if A has some special structure

Linear programming

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & a_i^T x \leq b_i, \quad i = 1, \cdots, m \end{aligned}$$

where the optimization variable $x \in \mathbb{R}^n$, and $c, a_i, b_i \in \mathbb{R}^n$ are parameters.

- no analytical formula for solution
- reliable and efficient algorithms available (e.g., Dantzig's simplex method, interior-point method)
- computational time proportional to $n^2 m$ if $m \leq n$ (interior-point method); less with structure

Linear programming: example

The Chebyshev approximation problem:

$$\text{minimize} \quad \|Ax - b\|_\infty$$

with $x \in \mathbb{R}^n$, $b \in \mathbb{R}^k$ and $A \in \mathbb{R}^{k \times n}$. The problem is similar to the least-square problem, but with the ℓ_∞-norm replacing the ℓ_2-norm:

$$\|Ax - b\|_\infty = \max_{i=1,\cdots,k} |a_i^T x - b_i|$$

where $a_i \in R^n$ is the ith column of A^T.

Linear programming: example

The Chebyshev approximation problem:

$$\text{minimize} \quad \|Ax - b\|_\infty$$

with $x \in \mathbb{R}^n$, $b \in \mathbb{R}^k$ and $A \in \mathbb{R}^{k\times n}$. The problem is similar to the least-square problem, but with the ℓ_∞-norm replacing the ℓ_2-norm:

$$\|Ax - b\|_\infty = \max_{i=1,\cdots,k} |a_i^T x - b_i|$$

where $a_i \in R^n$ is the ith column of A^T.

An equivalent linear programming:

$$\begin{aligned} \min \quad & t \\ \text{s.t.} \quad & a_i^T x - t \leq b_i, \quad i = 1, \cdots, k \\ & -a_i^T x - t \leq -b_i, \quad i = 1, \cdots, k \end{aligned}$$

Convex optimization problems

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \le \mathbf{b}_i, \quad i = 1, \cdots, m \end{aligned}$$

where $x \in \mathbb{R}^n$.

- both objective and constraint functions are convex

  $$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y)$$

  for any $0 \le \theta \le 1$, and any x and y in the domain of f_0 and f_i for all i.
- includes least-square and linear programming problems as special cases.
- no analytical formula for solution
- reliable and efficient algorithms available

Topics to be covered

- Convex sets and convex functions
- Duality
- Unconstrained optimization
- Equality constrained optimization
- Interior-point methods
- Semidefinite programming

Brief history of optimization

- 1700s: theory for unconstrained optimization (Fermat, Newton, Euler)
- 1797: theory for equality constrained optimization (Lagrange)
- 1947: simplex method for linear programming (Dantzig)
- 1960s: early interior-point methods (Fiacco, McCormick, Dikin, etc)
- 1970s: ellipsoid method and other subgradient methods
- 1980s: polynomial-time interior-point methods for linear programming (Karmarkar)
- 1990s: polynomial-time interior-point methods for nonlinear convex optimization (Nesterorv & Nemirovski)
- 1990-now: many new applications in engineering (control, signal processing, communications, etc); new problem classes (semidefinite and second-order cone programming, robust optimization, convex relaxation, etc)