Support Vector Machine ← convex opt.

Preclass 03: Convex Optimization

대표적 최적화

[SCS4049] Machine Learning and Data Science

Seongsik Park (s.park@dgu.edu)

School of AI Convergence & Department of Artificial Intelligence, Dongguk University

Optimization problem in standard form

$$
\begin{array}{lllr}
\boxed{\text{minimize}} & f_0(x) & & (1) \\
\text{subject to} & f_i(x) \leq 0, & i = 1, 2, ..., m & (2) \\
 & h_i(x) = 0, & i = 1, 2, ..., p & (3)
\end{array}
$$

constraint
구속조건

inequality constraint

equality constraint.

- $x \in \mathcal{R}^n$ is the optimization variable
- $f_0 : \mathcal{R}^n \to \mathcal{R}$ is the objective or cost function ← 최적화 대상
- $f_i : \mathcal{R}^n \to \mathcal{R}, i = 1, 2, ..., m$ are the inequality constraint functions
- $h_i : \mathcal{R}^n \to \mathcal{R}$ are the equality constraint functions

ML ← 최적화.

SVM

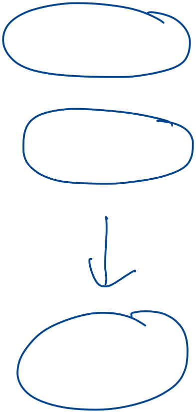
const. given $(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}) \dots$

object $\min_{\theta} \sum_{n=1}^{N} (y^{(n)} - \theta^T x^{(n)})^2$

↓

solution · normal eqn.

· G.D.



경로찾기

solution

서울 → 부산

최적경로 : 시간을 최소화. + 비용도 최소화

구속조건 : 대중교통 / 자가용 + 고속도로 지방도.

비행기
기차
버스
⋮

Convex optimization problem

Standard form convex optimization problem → *global Solution!

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) \xrightarrow{} \text{convex function} && \quad (4)\\
\text{subject to} \quad & f_i(x) \le 0, \quad i = 1, 2, ..., m && \quad (5)\\
& a_i^T x = b_i, \quad i = 1, 2, ..., p && \quad (6)
\end{aligned}
$$

linear.

S.V.M
↓
convex problem.

- $f_0, f_1, ..., f_m$ are convex
- equality constraints are ~~affine~~ linear

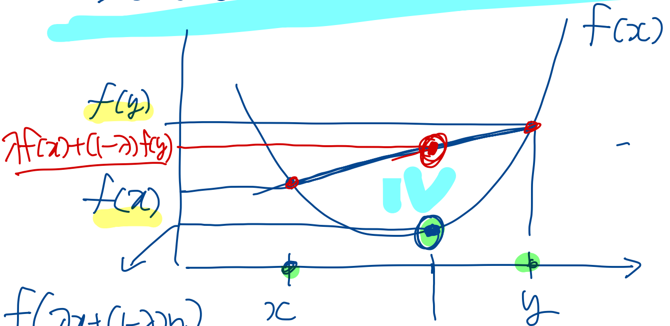
Often written as

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) && \quad (7)\\
\text{subject to} \quad & f_i(x) \le 0, \quad i = 1, 2, ..., m && \quad (8)\\
& Ax = b && \quad (9)
\end{aligned}
$$

Important property: feasible set of a convex optimization problem is convex

Convex function if

$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y)$ $\quad 0 \le \lambda \le 1$

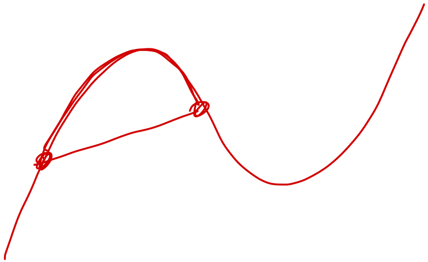


$f(\lambda x+(1-\lambda)y)$

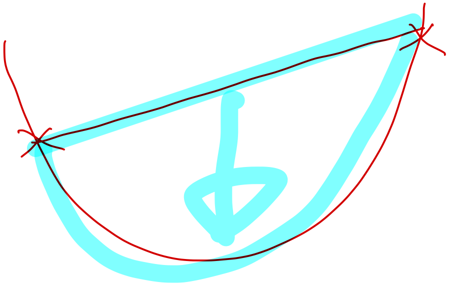
$\lambda \in 1$

$\lambda x+(1-\lambda)y$

$0 \le \lambda \le 1,$



convexx



complementary slackness

Dual problem and KKT conditions

SVM 써먹음.

primal problem

$\min f_0(x)$

s.t. $f_i(x) \le 0$

$h_i(x) = 0$

dual problem

p^* = d^*

convex problem.

standard form problem ~~convex~~ 일반.

$$\begin{array}{lll} \text{minimize}_{x} & f_0(x) & \quad (10) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, 2, ..., m & \quad (11) \\ & h_i(x) = 0, \quad i = 1, 2, ..., p & \quad (12) \end{array}$$

variable $x \in \mathcal{R}^n$, domain $\mathcal{D}$, optimal value p^*

1. Lagrangian: $L : \mathcal{R}^n \times \mathcal{R}^m \times \mathcal{R}^p \to \mathcal{R}$ with $\mathrm{dom}\, L = \mathcal{D} \times \mathcal{R}^m \times \mathcal{R}^p$

함수 정의

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \tag{13}$$

- weighted sum of objective and constraint functions
- λ_i is Lagrange multiplier associated with $f_i(x) \le 0$
- ν_i is Lagrange multiplier associated with $h_i(x) = 0$

$$\min_{x,y} \; x+y \quad \text{s.t.} \quad x^2+y^2=1$$

$$\min_{x,y} \; x+y \quad \text{s.t.} \quad x^2+y^2-1=0$$

$$\mathcal{L}(x,y,\nu) = (x+y) + \nu\,(x^2+y^2-1)$$

Lagrange dual function

2

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

$$g(\lambda, \nu) = \min_{x \in \mathcal{D}} L(x, \lambda, \nu) \tag{14}$$

$$= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right) \tag{15}$$

g is concave, can be $-\infty$ for some λ, ν

lower bound property: if $\lambda \geq 0$, then $g(\lambda, \nu) \leq p^*$

proof: if $\tilde{x}$ is feasible and $\lambda \geq 0$, then

$$f_0(\tilde{x}) \geq L(\tilde{x}, \lambda, \nu) \geq \min_{x \in \mathcal{D}} L(x, \lambda, \nu) = g(\lambda, \nu) \tag{16}$$

minimizing over all feasible $\tilde{x}$ gives $p^* \geq g(\lambda, \nu)$

The dual problem

③ Lagrange dual problem

$$\begin{aligned} \underset{\lambda,\nu}{\text{maximize}} \quad & g(\lambda, \nu) && (17) \\ \text{subject to} \quad & \lambda \geq 0 && (18) \end{aligned}$$

inequality, $\mathcal{L}$ multiplier.

- finds best lower bound on p^*, obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted d^*.
- λ, ν are dual feasible if $\lambda \geq 0$, $(\lambda, \nu) \in \operatorname{dom} g$
- often simplified by making implicit constraint $(\lambda, \nu) \in \operatorname{dom} g$ explicit

primal problem

$$\min f_0(x)$$

$$\text{s.t.} \begin{cases} f_1(x) \le 0 \leftarrow \\ f_2(x) \le 0 \leftarrow \end{cases}$$

$$\begin{cases} h_1(x) = 0 \\ h_2(x) = 0 \\ h_3(x) = 0. \end{cases}$$

① $\mathcal{L}(x, \lambda_1, \lambda_2, \nu_1, \nu_2, \nu_3)$

$$= f_0(x) + \lambda_1 f_1(x) + \lambda_2 f_2(x) + \nu_1 h_1(x) + \nu_2 h_2(x) + \nu_3 h_3(x)$$

② $g(\lambda_1, \lambda_2, \nu_1, \nu_2, \nu_3)$

$$= \min_x \mathcal{L}(x, \lambda_1, \lambda_2, \nu_1, \nu_2, \nu_3)$$

dual problem

③ $\max \; g(\lambda_1, \lambda_2, \nu_1, \nu_2, \nu_3)$

$$\text{s.t.} \quad \lambda_1 \ge 0$$
$$\lambda_2 \ge 0$$

Weak and strong duality

weak duality: $d^* \leq p^*$

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems

strong duality: $d^* = p^*$

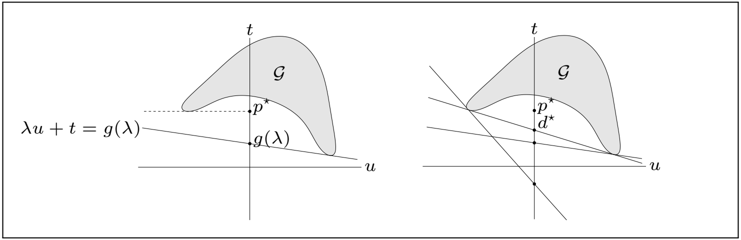
- does not hold in general
- holds for convex problems
- conditions that guarantee strong duality in convex problems are called constraint qualifications

Geometric interpretation

for simplicity, consider problem with one constraint $f_1(x) \leq 0$

interpretation of dual function

$$g(\lambda) = \min_{(u,t)\in\mathcal{G}} (t + \lambda u) \qquad \text{where } \mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in \mathcal{D}\} \tag{19}$$



- $\lambda u + t = g(\lambda)$ is supporting hyperplane to $\mathcal{G}$
- hyperplane intersects t-axis at $t = g(\lambda)$

Karush-Kuhn-Tucker (KKT) conditions

the following four conditions are called KKT conditions (for a problem with differentiable f_i, h_i)

1. primal constraints: $f_i(x) \leq 0$ i = 1, ..., m, $h_i(x) = 0$, i = 1, ..., p
2. dual constraints: $\lambda \geq 0$
3. $\boxed{\text{complementary slackness:}}$ $\lambda_i f_i(x) = 0, i = 1, \dots, m$
4. gradient of Lagrangian with respect to x vanishes:

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0 \tag{20}$$

convex problem

if strong duality holds and x, λ, ν are optimal, then they must satisfy the KKT conditions

Solution

primal constraint

$f_1(x) \leq 0$

$f_2(x) \leq 0$

$h_1(x) = 0$

$h_2(x) = 0$

$h_3(x) = 0$

dual constrain.

$\lambda_1 \geq 0$

$\lambda_2 \geq 0$

Complementary Slackness

$\overset{\geq 0}{\lambda_1} \overset{\leq 0}{f_1(x)} = 0 \Rightarrow$ 적어도 둘중 하나는 0이다.

$\lambda_2 f_2(x) = 0$

$f_1(x) > 0 \implies \lambda_1 = 0$

$\lambda_1 > 0 \implies f_1(x) = 0$

$f_i(x)$	λ_i
≥ 0	$= 0$
$= 0$	≥ 0

inequality > 0 $\quad \underline{\lambda_i = 0} \Longrightarrow$ 효과.

$= 0$ $\quad \lambda_i > 0 \Longrightarrow$ 효과.

(support vector)

Reference and further reading

- “Chap 7 | Sparse Kernel Machines” of C. Bishop, Pattern Recognition and Machine Learning
- “Chap 5 | Support Vector Machines” of A. Geron, Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow
- “Chap 4 | Convex Optimization Problems”, “Chap 5 | Duality” of S. Boyd, Convex Optimization