

Support Vector Machine ← convex opt.

Preclass 04: Convex Optimization

대표작
초/중/고

[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


$$\begin{array}{ll} \boxed{\text{minimize}} & f_0(x) \quad (1) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, 2, \dots, m \quad \text{inequality constraint} \quad (2) \\ & h_i(x) = 0, \quad i = 1, 2, \dots, p \quad \text{equality constraint} \quad (3) \end{array}$$

constraint
구속조건

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

ML $\leftarrow$ 최적화.

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

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



solution · Normal eqn.
· G.D.

SVM



경로 찾기

서로 solution → 복사

→ 최적 명문: ⁶ 시간을 최소화. + ⁷ 비용도 최소화

구속조건: $\frac{\text{대중교통}}{\text{비행기}}$ / $\frac{\text{지나봄}}{\text{기차}}$ + $\frac{\text{고속도로}}{\text{지방도}}$

비행기

기차

버스

⋮

Convex optimization problem

Standard form convex optimization problem



**global solution!*

$$\left[\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, 2, \dots, m \\ & a_i^T x = b_i, \quad i = 1, 2, \dots, p \end{array} \right] \quad \begin{array}{l} \text{convex function} \\ (4) \\ (5) \\ (6) \end{array}$$

linear.

- $f_0, f_1, \dots, f_m$ are convex
- equality constraints are affine *linear*

S.V.M



Convex problem.

Often written as

$$\text{minimize} \quad f_0(x) \quad (7)$$

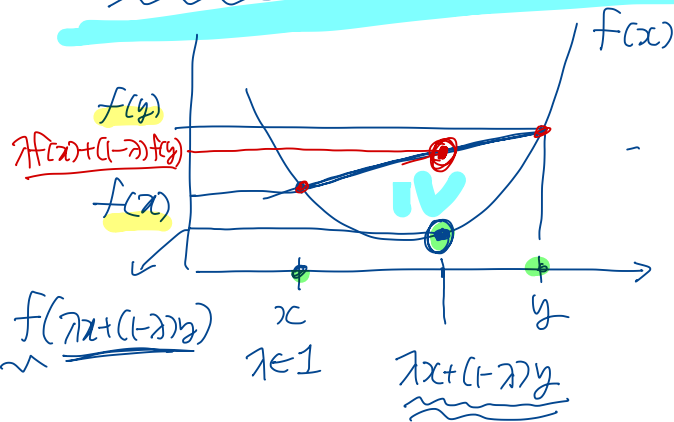
$$\text{subject to} \quad f_i(x) \leq 0, \quad i = 1, 2, \dots, m \quad (8)$$

$$Ax = b \quad (9)$$

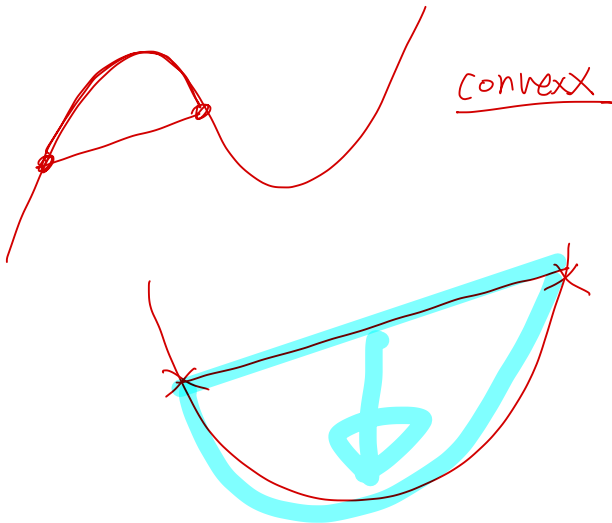
Important property: feasible set of a convex optimization problem is convex

convex function: f

$$\underline{f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)} \quad 0 \leq \lambda \leq 1$$



$$0 \leq \lambda \leq 1$$



Dual problem and KKT conditions

complementary
Slackness

SVM 시작은.

primal problem

$$\begin{aligned} \min f_0(x) \\ \text{s.t. } f_i(x) \leq 0 \\ h_i(x) = 0 \end{aligned}$$

p^*

≡

convex
problem.

dual problem

d^*

d^*

Lagrangian

standard form problem

~~convex~~ 일반.

$$\begin{cases} \text{minimize} & f_0(\mathbf{x}) \end{cases} \quad (10)$$

$$\begin{cases} \text{subject to} & f_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \dots, m \end{cases} \quad (11)$$

$$\begin{cases} & h_i(\mathbf{x}) = 0, \quad i = 1, 2, \dots, p \end{cases} \quad (12)$$

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

① Lagrangian: $L: \mathcal{R}^n \times \mathcal{R}^m \times \mathcal{R}^p \rightarrow \mathcal{R}$ with $\text{dom } L = \mathcal{D} \times \mathcal{R}^m \times \mathcal{R}^p$

함수 정의

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

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

$$\begin{array}{ll} \min & \underline{x+y} \\ \text{s.t.} & x^2+y^2=1 \end{array}$$



$$\begin{array}{ll} \min & \underline{x+y} \\ \text{s.t.} & \underline{x^2+y^2-1=0} \end{array}$$

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

Lagrange dual function

② Lagrange dual function: $g : \mathcal{R}^m \times \mathcal{R}^p \rightarrow \mathcal{R}$

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

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

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

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

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

The dual problem

③ Lagrange dual problem

$$\underset{\lambda, \nu}{\text{maximize}} \quad \underline{g(\lambda, \nu)} \quad (17)$$

$$\text{subject to} \quad \underline{\lambda \geq 0} \quad (18)$$

inequality, λ 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 $\lambda \geq 0, (\lambda, \nu) \in \text{dom } g$
- often simplified by making implicit constraint $(\lambda, \nu) \in \text{dom } g$ explicit

primal problem

$$\begin{aligned} \min & \ f_0(x) \\ \text{s.t.} & \begin{cases} f_1(x) \leq 0 \\ f_2(x) \leq 0 \end{cases} \end{aligned}$$

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

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

$$\begin{aligned} &= f_0(x) + \lambda_1 f_1(x) + \lambda_2 f_2(x) \\ &\quad + \nu_1 h_1(x) + \nu_2 h_2(x) + \nu_3 h_3(x) \end{aligned}$$

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

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

dual problem

$$\textcircled{3} \max g(\lambda_1, \lambda_2, \nu_1, \nu_2, \nu_3)$$

$$\begin{aligned} \text{s.t.} \quad & \lambda_1 \geq 0 \\ & \lambda_2 \geq 0 \end{aligned}$$

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: $d^* = p^*$

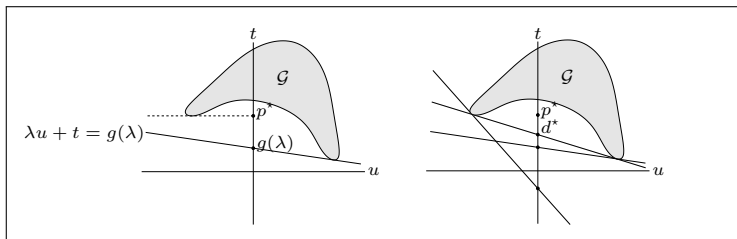
- does not hold in general
- holds for convex problems
- conditions that guarantee strong duality in convex problems are called constraint qualifications

Geometric interpretation

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

interpretation of dual function

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



- $\lambda u + t = g(\lambda)$ 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: $f_i(x) \leq 0, i = 1, \dots, m, h_i(x) = 0, i = 1, \dots, p$
2. dual constraints: $\lambda \geq 0$
3. complementary slackness: $\lambda_i f_i(x) = 0, i = 1, \dots, m$
4. gradient of Lagrangian with respect to x vanishes:

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

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
constraint.

$$\lambda_1 \geq 0$$

$$\lambda_2 \geq 0$$

Complementary

Slackness

$$\lambda_1^0 f_1(x^0) = 0 \Rightarrow$$

$$\lambda_2 f_2(x) = 0$$

λ_2 or f_2
 $\frac{\tau \lambda}{2} \delta \mu \frac{1}{2} 0$ 이다.

$$\underbrace{\forall f_i(x) > 0}_{\text{blue}} \implies \underbrace{\lambda_i = 0}_{\text{blue}}$$

$$\underbrace{\forall \lambda_i > 0}_{\text{red}} \implies \underbrace{f_i(x) = 0}_{\text{red}}$$

$f_i(x)$	λ_i
≥ 0	$= 0$
$= 0$	≥ 0

inequality > 0
 $= 0$

$\lambda_i = 0$ $\Rightarrow$ $\overline{\text{E.D.}}$
 $\lambda_i > 0$ $\Rightarrow$ $\underline{\text{E.D.}}$

(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