SVM

Preclass 08: Dual Problem and KKT conditions

[SCS4049] Machine Learning and Data Science

Seongsik Park (s.park@dgu.edu)

Department of Artificial Intelligence, Dongguk University

원래의 최적화 문제 primal problem.

↓ ↳ Lagrangian → Lagrange dual function.

「Lagrange function」 ↩

Lagrange multiplier.

↓

새로운 최적화 문제 dual problem.

↓

dual prob의 해를 구함 = primal prob의 해와 동일.

→ convex optimization 전제.

(strong duality)

Lagrangian

standard form problem 일반 최적화.

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

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

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

$$\boxed{L(x, \lambda, \nu)} = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \tag{4}$$

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

= 구속조건 앞에 곱해진 변수.

Lagrange dual function

고정 = 주어진 값.

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

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \boxed{L(x, \lambda, \nu)} = \min_{x} \mathcal{L}(x, \lambda, \nu) \tag{5}$$

$$= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right) \tag{6}$$

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

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

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

$$f_0(\tilde{x}) \ge L(\tilde{x}, \lambda, \nu) \ge \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = g(\lambda, \nu) \tag{7}$$

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

The dual problem

primal: f_0, f_i, h_i + λ_i, ν_i

$\mathcal{L}(x, \lambda, \nu)$

$g(\lambda, \nu)$

Lagrange $\boxed{\text{dual problem}}$

$$\begin{aligned} \text{maximize} \quad & g(\lambda, \nu) & (8) \\ \text{subject to} \quad & \boxed{\lambda \geq 0} & (9) \end{aligned}$$

ineq.

- 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

Weak and strong duality

p^*의 lower bound $= d^*$

$d^* =$ dual의 해 $= \max g(\lambda, \nu)$ s.t. $\lambda \geq 0$

① weak duality: $d^* \leq p^*$

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

$p^* =$ primal의 해 $= \min f_0(x)$ s.t. $f_i \leq 0$, $h_i = 0$

② 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

min $f_0(x) = t$

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

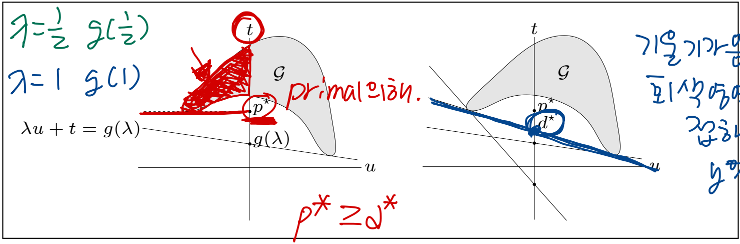
$f_1(x) = u$

interpretation of dual function

dual problem

$$g(\lambda) = \inf_{(u,t)\in\mathcal{G}} (t + \lambda u) \qquad \text{where } \mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in \mathcal{D}\} \qquad (10)$$

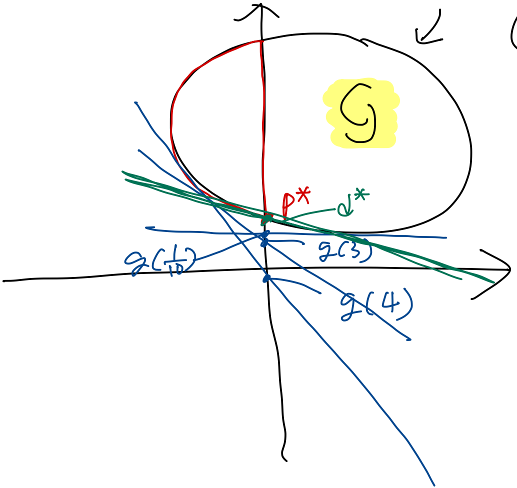
$\lambda \ge 0$ $\qquad\qquad$ (u, t)



기울기가 음수로 회색영역에 아래쪽을 접하는 선의 y^*절편.

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

max $g(\lambda) = d^*$



$$G = \{(f_1(x), f_0(x)) \mid x \in D\}$$

가로 세로.

convex problem → G : convex set

$p^* = d^*$

Karush-Kuhn-Tucker (KKT) conditions

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

원래문제의 구속조건.

1. primal constraints: $f_i(x) \leq 0$ i = 1, ..., m, $h_i(x) = 0$, i = 1, ..., p
2. dual constraints: $\lambda \geq 0$ dual문제의 constraint.

   max $g(\lambda, \nu)$
   s.t. $\lambda \geq 0$.
3. complementary slackness: $\lambda_i f_i(x) = 0$, $i = 1, \ldots, m$
4. gradient of Lagrangian with respect to x vanishes:

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0 \qquad (11)$$

Convex prob,

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

해인경우.

1. primal constraint.

2. dual constraint.

3. complementary slackness.

$f_i(x) \le 0$.

$\lambda_i \ge 0$

$\lambda_i f_i(x) = 0$

둘중에 적어도 하나는 0이다.

$\lambda_i > 0 \quad f_i(x) = 0$

$\lambda_i = 0 \quad f_i(x) < 0$

Appendix

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