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

Lagrangian

standard form problem

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

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 $\mathrm{dom}\, L = \mathcal{D} \times \mathcal{R}^m \times \mathcal{R}^p$

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

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

Lagrange dual function

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

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

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

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

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

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

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

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

The dual problem

Lagrange dual problem

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

- 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

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: $d^* = p^*$

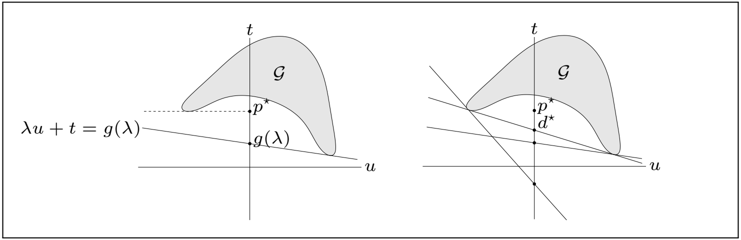
- does not hold in general
- holds for convex problems
- conditions that guarantee strong duality in convex problems are called constraint qualifications

Geometric interpretation

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

interpretation of dual function

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



- $\lambda u + t = g(\lambda)$ 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. complementary slackness: $\lambda_i f_i(x) = 0, i = 1, ..., m$
4. gradient of Lagrangian with respect to x vanishes:

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

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

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