Duality Theory

Jitkomut Songsiri

Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University

CUEE

March 14, 2024

Outline

Lagrangian and dual function

General setting

(mathematical) optimization problem

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

- $x = (x_1, \dots, x_n)$: optimization variable
- $f_0 : \mathbf{R}^n \to \mathbf{R}$: objective function (generally, nonlinear)
- $f_i : \mathbf{R}^n \to \mathbf{R}, i = 1, \dots, m$: inequality constraint functions
- $h_i : \mathbf{R}^n \to \mathbf{R}, i = 1, \dots, p$: equality constraint functions

domain of the problem: $\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\, h_i$

Lagrangian

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

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

- L is a weigthed sum of objective and constraint functions
- $\lambda \in \mathbf{R}^m_+$ is the Lagrange multiplier corresponding to inequality constraints
- $\nu \in \mathbf{R}^p$ is the Lagrange multiplier corresponding to equality constraints

Lagrange dual function

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

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \\ &= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right) \end{aligned}$$

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

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

- if $\tilde{x}$ is feasible and $\lambda \succeq 0$ then

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

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

Least-norm solution of linear equations

problem: minimize $(1/2)x^T x$ subject to $Ax = b$

dual function

- Lagrangian is $L(x, \nu) = (1/2)x^T x + \nu^T (Ax - b)$
- to minimize L over x, set gradient equal to zero:

$$\nabla_x L(x, \nu) = x + A^T \nu = 0 \quad \Rightarrow \quad x = -A^T \nu$$

- substitute x in L to obtain g

$$g(\nu) = L(-A^T \nu, \nu) = -(1/2)\nu^T A A^T \nu - b^T \nu$$

  which is concave in ν

lower bound property: $p^\star \geq -(1/2)\nu^T A A^T \nu - b^T \nu$ for all ν

Standard form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \quad x \succeq 0 \end{array}$$

- Lagrangian is

$$\begin{aligned} L(x, \lambda, \nu) &= c^T x + \nu^T (Ax - b) - \lambda^T x \\ &= -b^T \nu + (c + A^T \nu - \lambda)^T x \end{aligned}$$

- since L is affine in x

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \begin{cases} -b^T \nu, & \text{if } A^T \nu - \lambda + c = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

g is linear on affine domain $\{(\lambda, \nu) \mid A^T \nu - \lambda + c = 0\}$, hence concave

lower bound property: $p^\star \geq -b^T \nu$ if $A^T \nu + c \succeq 0$

Dual problem

The dual problem

Lagrange dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

- we find the best lower bound on $p^\star$ obtained from Lagrange dual function
- a convex problem (even if the primal is non-convex); optimal value denoted $d^\star$
- λ, ν are dual feasible if $\lambda \succeq 0$ for $(\lambda, \nu) \in \mathbf{dom}\, g$
- often simplified by making implicit constraint $(\lambda, \nu) \in \mathbf{dom}\, g$ explicit

example: standard form LP and its dual

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array} \qquad\qquad \begin{array}{ll} \text{maximize} & -b^T \nu \\ \text{subject to} & A^T v + c \succeq 0 \end{array}$$

(dual of LP is an LP)

Weak and strong duality

weak duality: $d^\star \leq p^\star$ (always holds for convex and non-convex problems)

- can be used to find non-trivial lower bounds for difficult problems
- if the primal in unbounded below ($p^\star = -\infty$), then $d^\star = -\infty$ (the dual is infeasible)
- if the dual is unbounded above ($d^\star = \infty$), we have $p^\star = \infty$ (the primal is infeasible)
- $p^\star - d^\star$ is called the **duality gap** and always non-negative

strong duality: $d^\star = p^\star$

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

Slater's condition

Slater's constraint qualification

strong duality holds for a **convex** problem

$$\begin{array}{ll} \text{minmize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, 2, \ldots, m \\ & Ax = b \end{array}$$

if it is strictly feasible, *i.e.*,

$$\exists x \in \textbf{int}\, \mathcal{D} : \quad f_i(x) < 0, \quad i = 1, 2, \ldots, m, \quad Ax = b$$

- strong duality also guarantees that the dual optimum is attained (if $p^\star > -\infty$)

$$\exists \text{ a dual feasible } (\lambda^\star, \nu^\star) \text{ with } g(\lambda^\star, \nu^\star) = d^\star = p^\star$$

- weak form of Slater's condition: strong duality holds when some of f_i's are affine

$$f_i(x) \leq 0, \quad i = 1, 2, \ldots, k, \quad f_i(x) < 0, \quad i = k+1, \ldots, m, \quad Ax = b$$

Inequality form LP

primal problem (P)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \end{array}$$

dual function

$$g(\lambda) = \inf_x [(c + A^T\lambda)^T x - b^T\lambda] = \begin{cases} -b^T\lambda, & \text{if } A^T\lambda + c = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

dual problem (D)

$$\begin{array}{ll} \text{maximize} & -b^T\lambda \\ \text{subject to} & A^T\lambda + c = 0, \quad \lambda \succeq 0 \end{array}$$

- from Slater's condition: $p^\star = d^\star$ if $A\tilde{x} \prec b$ for some $\tilde{x}$ (primal is feasible)
- in fact, $p^\star = d^\star$ except when primal and dual are infeasible
- we can verify that the Lagrange dual of problem D is equivalent to the primal P

Quadratic program

primal problem (assume $P \in \mathbf{S}^n_{++}$)

$$\begin{array}{ll} \text{minimize} & x^T P x \\ \text{subject to} & Ax \preceq b \end{array}$$

dual function

$$g(\lambda) = \inf_x \left(x^T P x + \lambda^T (Ax - b)\right) = -\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

dual problem

$$\begin{array}{ll} \text{maximize} & -(1/4)\lambda^T A P^{-1} A^T \lambda - b^T \lambda \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

- from Slater's condition: $p^\star = d^\star$ if $A\tilde{x} \prec b$ for some $\tilde{x}$

Complementary slackness

assume strong duality holds, $x^\star$ is primal optimal, $(\lambda^\star, \nu^\star)$ is dual optimal

$$
\begin{aligned}
f_0(x^\star) = g(\lambda^\star, \nu^\star) &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right) \\
&\leq f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star) \\
&\leq f_0(x^\star) \quad (\text{because } h_i(x) = 0 \text{ and } \lambda_i f_i(x^\star) \leq 0)
\end{aligned}
$$

hence, the two inequalities hold with equality and we must have

- $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$
- $\lambda_i^\star f_i(x^\star) = 0$ for $i = 1, 2, \ldots, m$ (known as complementary slackness)

$$
\lambda_i^\star > 0 \Longrightarrow f_i(x^\star) = 0, \quad f_i(x^\star) < 0 \Longrightarrow \lambda_i^\star = 0
$$

Karush-Kuhn-Tucker (KKT) conditions

Karush-Kuhn-Tucker (KKT) conditions

for a problem with differentiable f_i, h_i, the four conditions are called **KKT**

1. **primal feasibility:** $f_i(x) \leq 0$, $i = 1, \ldots, m$, $h_i = 0$, $i = 1, \ldots, p$
2. **dual feasiblity:** $\lambda \succeq 0$
3. **complementary slackness:** $\lambda_i f_i(x) = 0$, $i = 1, 2, \ldots, m$
4. **zero gradient of Lagrangian** with respect to x

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

KKT as necessary conditions: if strong duality holds and $(x^\star, \lambda^\star, \nu^\star)$ are optimal, then they must satisfy the KKT conditions (follow from page 16)

KKT conditions for convex problems

if $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ satisfy KKT for a convex problem, then they are optimal:

- from the 1st KKT: $\tilde{x}$ is primal feasible
- from the 2nd KKT ($\lambda_i \geq 0$) and convexity: $L(x, \tilde{\lambda}, \tilde{\nu})$ is convex in x
- from the 4th KKT: $\tilde{x}$ minimizes $L(x, \tilde{\lambda}, \tilde{\nu})$ over $x \Rightarrow g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$
- from the 3rd KKT (complementary slackness) and $h_i(\tilde{x}) = 0$

$$g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) = f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\tilde{x}) = f_0(\tilde{x})$$

conclusion: $\tilde{x}$ and $(\tilde{\lambda}, \tilde{\nu})$ have zero duality gap and are primal and dual optimal

for **convex** problems, KKT conditions are **sufficient** for optimality if **Slater's condition** is satisfied for **convex** problems

- from page 13, it implies duality gap is zero and the dual optimum is attained
- so, x is optimal if and only if there are (λ, ν), together with x, satisfy the KKT conditions

Projection onto probability simplex

Dual of projection onto the probability simplex

consider the problem of projecting a onto the probability simplex:

$$\underset{x}{\text{minimize}} \quad (1/2)\|x-a\|_2^2 \quad \text{subject to} \quad x \succeq 0, \quad \mathbf{1}^T x = 1$$

- Lagrangian: $L(x,\lambda,\nu) = (1/2)\|x-a\|_2^2 - (\lambda - \nu\mathbf{1})^T x - \nu$
- use the fact that $(1/2)\|x-a\|_2^2 - y^T x$ is minimized over x when $x = y + a$ and the minimum is $-(1/2)\|y\|_2^2 - y^T a$
- the dual problem is QCQP

$$\underset{\lambda}{\text{maximize}} \quad g(\lambda,\nu) := -(1/2)\|\lambda - \nu\mathbf{1}\|_2^2 - (\lambda-\nu\mathbf{1})^T a - \nu \quad \text{subject to} \quad \lambda \succeq 0$$

- KKT conditions:

$$\text{primal feasibility: } x^\star \succeq 0, \quad \mathbf{1}^T x^\star = 1, \quad \text{dual feasibility: } \lambda^\star \succeq 0,$$
$$\text{zero-gradient: } x^\star = \lambda^\star - \nu^\star \mathbf{1} + a, \quad \text{complimentary slackness: } \lambda_i^\star x_i = 0, \quad \forall i$$

the dual probelm can be further simplified

$$-g(\lambda,\nu) = (1/2)\|\lambda - (\nu\mathbf{1} - a)\|_2^2 + \nu - (1/2)\|a\|_2^2 \ \triangleq \ \tilde{g}(\lambda,\nu)$$

(completing square in λ) – which can be minimized over λ first

$$\lambda^\star = \begin{cases} \nu\mathbf{1} - a, & \nu\mathbf{1} - a \geq 0, \\ 0, & \text{otherwise} \end{cases} \ \triangleq \ \max(0, \nu\mathbf{1} - a) \ \triangleq \ (\nu\mathbf{1} - a)^+$$

the dual problem becomes the minimization of $\tilde{g}(\lambda^\star,\nu)$ given by

$$\begin{aligned} \tilde{g}(\lambda^\star,\nu) &= (1/2)\|(\nu\mathbf{1} - a)^+ - (\nu\mathbf{1} - a)\|_2^2 + \nu - (1/2)\|a\|_2^2 \\ &= (1/2)\|(a - \nu\mathbf{1})^+\|_2^2 + \nu - (1/2)\|a\|_2^2 \end{aligned}$$

(we have used $z = z^+ - z^-$ and $z^- = -\min(0,z) = \max(0,-z) = (-z)^+$)

there is an efficient way to find $\nu^\star$; one of them is to find the subgradient

$$\partial \tilde{g} = [(a - \nu \mathbf{1})^+]^T g + 1 =$$

where $g = (g_1, g_2, \ldots, g_n)$ and $g_k = -1$ if $a_k - \nu > 0$ and $g_k = 0$ otherwise

then zero is one of the subgradients (optimality condition) – find ν such that

$$\partial \tilde{g} = 1 - \mathsf{sum}(a - \nu \mathbf{1})^+ = 0$$

once we obtain $\nu^\star$, we solve $x^\star$ from KKT

$$x^\star = \lambda^\star - \nu^\star \mathbf{1} + a = (\nu^\star - a)^+ - (\nu^\star \mathbf{1} - a) = (a - \nu^\star \mathbf{1})^+$$

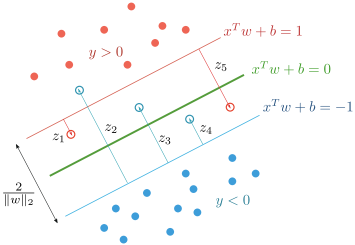
Soft-margin SVM

Soft-margin SVM

problem parameters: $x_i \in \mathbf{R}^n$ and $y_i \in \{1, -1\}$ for $i = 1, \ldots, N, C > 0$

optimization variables: $w \in \mathbf{R}^n, b \in \mathbf{R}, z \in \mathbf{R}^N$

$$
\begin{array}{ll}
\text{minimize} & (1/2)\|w\|_2^2 + C\mathbf{1}^T z \\
\text{subject to} & y_i(x_i^T w + b) \geq 1 - z_i, \quad i = 1, \ldots, N \\
& z \succeq 0
\end{array}
$$



- z_i is called a *slack variable*, allowing some of the hard constraints to be relaxed
- if $z_i^\star > 0$, the ith data point is relaxed to lie on the wrong side of its margin
- $\sum_i z^\star$ is the total distance of points on the wrong side of their margin (called **margin errors**)
- the *penalty parameter* C controls the trade-off between maximizing the margin and the margin errors

Dual of soft-margin SVM

dual problem of soft-margin SVM: with variable $\alpha \in \mathbf{R}^N$

$$\begin{array}{ll} \text{maximize}_\alpha & \mathbf{1}^T\alpha - (1/2)\sum_{i=1}^N \sum_{j=1}^N \alpha_i\alpha_j y_i y_j x_i^T x_j \\ \text{subject to} & \sum_{i=1}^N \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \ldots, N \end{array}$$

let α and λ be Lagrange multipliers (w.r.t. 1st and 2nd inequalities on page 26)

$$L(w, b, z, \alpha, \lambda) = \frac{1}{2}\|w\|_2^2 - \sum_{i=1}^N \alpha_i y_i x_i^T w - b\sum_{i=1}^N \alpha_i y_i + (C\mathbf{1} - \alpha - \lambda)^T z + \mathbf{1}^T\alpha$$

note that L is quadratic in w: $\frac{1}{2}\|w\|_2^2 - d^T w$ and L is linear in b and z

- $\inf_w L$ occurs when $w = d = \sum_i \alpha_i y_i x_i$ and the infimum is

$$-(1/2)\|d\|_2^2 = -(1/2)d^T d = -(1/2)\sum_i \sum_j \alpha_i\alpha_j y_i y_j x_i^T x_j$$

- since L is linear in z, b, $\inf_z L$ and $\inf_b L$ exist (and are zero) only when

$$\sum_i \alpha_i y_i = 0, \quad C\mathbf{1} - \alpha - \lambda = 0$$

- dual function: $g(\alpha) = -(1/2) \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j + \mathbf{1}^T \alpha$
- KKT conditions of SVM primal problem are

primal feasiblity:	$y_i(x_i^T w + b) \geq 1 - z_i, \quad i = 1, 2, \ldots, N,$
	$z \succeq 0$
dual feasiblity:	$\sum_{i=1}^N \alpha_i y_i = 0,$
	$0 \leq \alpha_i \leq C, \quad i = 1, 2, \ldots, N$
	or equivalently, $\lambda \succeq 0, \quad \alpha = C\mathbf{1} - \lambda$
zero-gradient of L:	$w = \sum_{i=1}^N \alpha_i y_i x_i$
complementary slackness:	$\alpha_i [y_i(x_i^T w + b) - (1 - z_i)] = 0$
	$\lambda_i z_i = 0, \quad i = 1, 2, \ldots, N$

Implications of SVM's KKT

dual feasibility and complementary slackness characterize three groups of points

$$\alpha_i = C - \lambda_i, \;\; \lambda_i z_i = 0, \;\; \alpha_i[y_i(x_i^T w + b) - (1 - z_i)] = 0$$

correct side of the margin

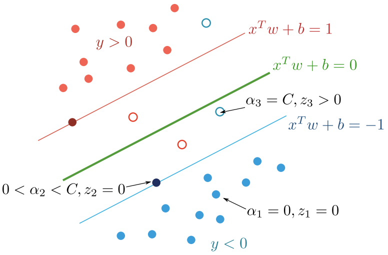
$$\alpha_i = 0, \;\; \lambda_i = C, \;\; z_i = 0, \;\; y_i(x_i^T w + b) \geq 1$$

edge of the margin

$$0 < \alpha_i < C, \;\; \lambda_i > 0, \;\; z_i = 0, \;\; y_i(x_i^T w + b) = 1$$

wrong side of the margin

$$\alpha_i = C, \;\; \lambda_i = 0, \;\; y_i(x_i^T w + b) = 1 - z_i, \;\; z_i > 0$$



- the observations i for which $\alpha_i > 0$ are called **support vectors** because w is a linear combination of only those terms: $w = \sum_{i=1}^N \alpha_i y_i x_i$
- margin points: $y_i(x_i^T w + b) = 1 \Leftrightarrow b = -x_i^T w + y_i$ (averaging all solutions)

a compact form of SVM dual

$$\begin{array}{ll} \text{minimize} & (1/2)\alpha^T G\alpha - \mathbf{1}^T\alpha \\ \text{subject to} & \alpha^T y = 0, \;\; 0 \preceq \alpha \preceq C\mathbf{1} \end{array}$$

where $G \in \mathbf{R}^{N\times N}$, $G_{ij} = \langle y_i x_i, y_j x_j \rangle$ (called a **Gram** matrix); clearly, $G \succeq 0$

- it is a QP with a linear constraint and a box constraint
- this formulation is called C-SVC (C-support vector classification)
- available algorithms:
  - quadratic programming solvers (active-set, interior-point) on the dual
  - sequential minimal optimization (SMO) on the dual (used in `fitcsvm` by `MATLAB` and `libsvm` library, which supports nonlinear classifiers)
  - coordinate descent on the dual (large-scale linear SVM, used in `liblinear`)

Conjugate function

Conjugate function and Lagrange dual

conjugate function: $f^*(y) = \sup_{x \in \mathbf{dom}\, f}(y^T x - f(x))$

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & Ax \preceq b, \quad Cx = d
\end{array}
$$

dual function

$$
\begin{aligned}
g(\lambda, \nu) &= \inf_{x \in \mathbf{dom}\, f_0} \left[f_0(x) + (A^T\lambda + C^T\nu)^T x \right] - b^T\lambda - d^T\nu \\
&= -f_0^*(-A^T\lambda - C^T\nu) - b^T\lambda - d^T\nu
\end{aligned}
$$

if conjugate of f_0 is known, it can simplify the derivation of dual

examples:

- entropy: $f_0(x) = \sum_{i=1}^n x_i \log x_i, \quad f_0^*(y) = \sum_{i=1}^n e^{y_i - 1}$
- quadratic: $f_0(x) = (1/2)\|x - a\|_2^2, \quad f_0^*(y) = (1/2)\|y\|_2^2 + y^T a$

Importance of KKT conditions

Importance of KKT conditions

many important roles of KKT conditions

- it is possible to solve KKT analytically in some problems

  $$\text{minimize: } (1/2)x^T Px + q^T x + r \quad \text{subject to } Ax = b \quad (\text{where } P \in \mathbf{S}^n_+)$$

  KKT conditions are system of linear equations: $Ax^\star = b$ and $Px^\star + q + A^T\nu^\star = 0$
- many algorithms for convex optimization can be interpreted as methods for solving KKT conditions
- the dual problem can be easier to solve than the primal – once $(\lambda^\star, \nu^\star)$ is obtained, it is possible to compute a primal optimal from a dual optimal solution
- $(\lambda^\star, \nu^\star)$ provide information for perturbation and sensitivity analysis – how the primal objective changes under a problem parameter perturbation

Solving the primal solution via the dual

suppose we have strong duality and a dual optimal $(\lambda^\star, \nu^\star)$ is known

- any primal optimal point is also a minimizer of $L(x, \lambda^\star, \nu^\star)$
- suppose that the solution of

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

  is *unique* (for example, when $L(x, \lambda^\star, \nu^\star)$ is strictly convex in x)
- if the solution of (1) is primal feasible, it must be primal optimal
- if the solution of (1) is not primal feasible, then no primal optimal point can exist – that is, the primal optimum is not attained

Entropy maximization

$$\begin{array}{ll} \text{minimize} & f_0(x) := \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & Ax \preceq b \\ & \mathbf{1}^T x = 1 \end{array}$$

dual problem:

$$\begin{array}{ll} \text{maximize}_{\lambda,\nu} & -b^T\lambda - \nu - e^{-\nu-1} \sum_{i=1}^n e^{-a_i^T \lambda} \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

- assume (weak) Slater's condition holds; hence, strong duality holds
- suppose we have solved the dual and obtain $(\lambda^\star, \nu^\star)$ to form

$$L(x, \lambda^\star, \nu^\star) = \sum_{i=1}^n x_i \log x_i + \lambda^{\star T}(Ax - b) + \nu^\star(\mathbf{1}^T x - 1)$$

which is strictly convex on $\mathcal{D}$ and bounded below

Entropy maximization

- minimization of $L(x, \lambda^\star, \nu^\star)$ has a unique solution $x^\star$ given by

$$x^\star = 1/\exp\,(a_i^T \lambda^\star + \nu^\star + 1), \quad i = 1, 2, \ldots, n$$

  (a_i are the columns of A)
- if $x^\star$ is primal feasible, it must be the optimal solution of the primal problem
- if $x^\star$ is not primal feasible, then the primal optimum is not attained

Sensitivity analysis

a perturbed optimization problem:

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \leq u_i, \quad i = 1, 2, \dots, m \\
& h_i(x) = v_i, \quad i = 1, 2, \dots, p
\end{array}
$$

$$p^\star(u, v) = \inf \{ \ f_0(x) \mid \exists x \in \mathcal{D}, \ f_i(x) \leq u_i, i = 1, 2, \dots, m, \ h_i(x) = v_i, i = 1, 2, \dots, p \ \}$$

- when $u_i \geq 0$, we *relax* the ith inequality constraint
- when $v_i \neq 0$, we change the equality constraint
- $p^\star(u, v)$ is defined the optimal value of the perturbed problem
- we have $p^\star(0, 0) = p^\star$ (optimal value of unperturbed system)
- fact: when the original problem is convex, $p^\star$ is a convex function of u and v

Global inequality

for all u and v, it can be shown that

$$p^\star(u, v) \geq p^\star(0, 0) - \lambda^{\star T} u - \nu^{\star T} v$$

- if $\lambda_i^\star$ is **large** and $u_i < 0$ (tighten the ith inequality), then $p^\star(u, v)$ is guaranteed to increase greatly
- if $\lambda_i^\star$ is small and $u_i > 0$ (loosen the ith inequality), then $p^\star(u, v)$ will not decrease much
- if $\nu_i^\star$ is **large and positive** and $v_i < 0$), then $p^\star(u, v)$ is guaranteed to increase greatly
- if $\nu_i^\star$ is small and positive and $v_i > 0$, or if $\nu_i^\star$ is small and negative and $v_i < 0$, then $p^\star(u, v)$ will not decrease much

Local sensitivity analysis

suppose $p^\star(u, v)$ is differentiable at $u = 0, v = 0$

if strong duality holds, the optimal dual $\lambda^\star, \nu^\star$ are related to

$$\lambda_i^\star = -\frac{\partial p^\star(0,0)}{\partial u_i}, \quad \nu_i^\star = -\frac{\partial p^\star(0,0)}{\partial v_i}$$

- tightening the ith inequality ($u_i \leq 0$ and small) yields an *increase* in $p^\star$ of approximately $-\lambda_i^\star u_i$
- loosening the ith inequality ($u_i \geq 0$ and small) yields an *decrease* in $p^\star$ of approximately $\lambda_i^\star u_i$

Exercises

Exercises

derive the dual problem and KKT conditions; some of them has $x^\star$ in closed-form

1. minimize $(1/2)\|x - v\|_2^2$ subject to $x_1 = x_2 = \cdots = x_N$
2. minimize $(1/2)\|x - v\|_2^2$ subject to $a^T x \leq b$ (given that $a^T v \geq b$)
3. minimize $(1/2)\|Ax - b\|_2^2$ subject to $x \succeq 0$

References

duality theory

1. S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge Press, 2004
2. L. Vandenberghe, *Lecture notes of EE236B – Convex Optimization*, UCLA

algorithms for SVM

1. Chih-Chung Chang and Chih-Jen Lin, `libsvm`: a Library for support vector machines, https://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf
2. Rong-En Fan et.al., `liblinear`: a Library for large linear classifiers, https://www.csie.ntu.edu.tw/~cjlin/papers/liblinear.pdf
3. Cho-Jui Hsieh et.al, Dual coordinate descent method for large-scale linear SVM, https://www.csie.ntu.edu.tw/~cjlin/papers/cddual.pdf
4. `fitcsvm` by MATLAB, https://www.mathworks.com/help/stats/fitcsvm.html