EECS 545: Machine Learning

Supplementary Materials: Brief Intro to Convex Optimization

Honglak Lee, Jongwook Choi

02/13/2024

Last Update: 02/12/2023 (v2)

* Many slides are based on Stephen Boyd's course: Convex Optimization (website: http://www.stanford.edu/class/ee364a/)

UNIVERSITY OF MICHIGAN

Basics of convex optimization

- General optimization problem
  - very difficult to solve
  - methods involve some compromise, e.g., very long computation time, or not always finding the exact optimal solution

$$
\begin{aligned}
\min_{\mathbf{x}} \quad & f(\mathbf{x}) \\
\text{subject to} \quad & g_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, m \\
& h_i(\mathbf{x}) = 0,\ i = 1, \ldots, p
\end{aligned}
$$

- Exceptions: certain problem classes can be solved efficiently and reliably.
  - least-squares problems
  - linear programming
  - **convex** optimization problems

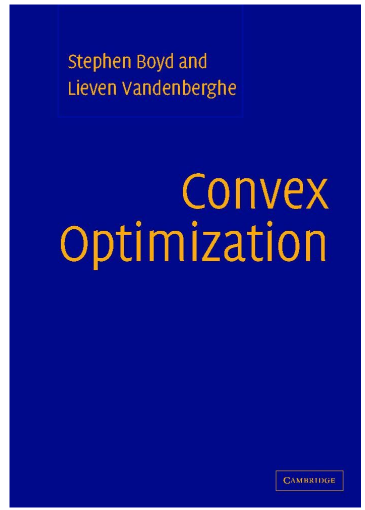
http://www.stanford.edu/class/ee364a/

Slide credit: Stephen Boyd

Contents

- Review: Convex Set, Convex Function
- Linear Programming, Quadratic Programming
- Constrained Optimization
- Lagrangian and Duality
- KKT Conditions for Strong Duality

  ← the goal of today!
- Exercise: One-class SVM



https://web.stanford.edu/class/ee364a/
https://stanford.edu/~boyd/cvxbook/

Convex Sets

line segment between x_1 and x_2: all points

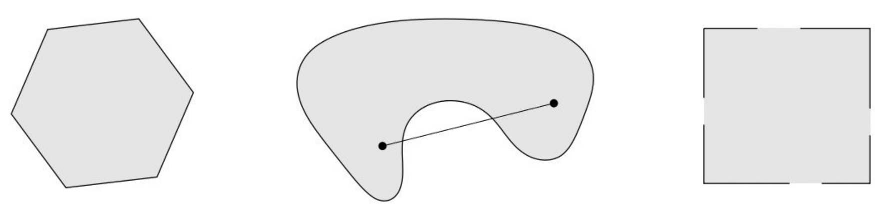
$$x = \theta x_1 + (1 - \theta) x_2$$

with $0 \leq \theta \leq 1$

convex set: contains line segment between any two points in the set

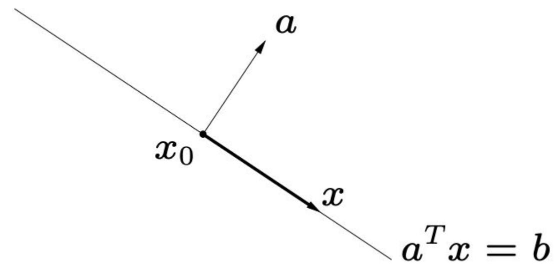
$$x_1, x_2 \in C, \quad 0 \leq \theta \leq 1 \quad \Longrightarrow \quad \theta x_1 + (1 - \theta) x_2 \in C$$

examples (one convex, two nonconvex sets)

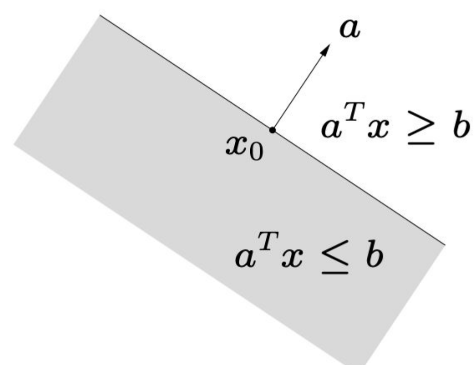


Example: Hyper-planes and half-spaces

hyperplane: set of the form $\{x \mid a^T x = b\}$ $(a \neq 0)$



halfspace: set of the form $\{x \mid a^T x \leq b\}$ $(a \neq 0)$



- a is the normal vector
- hyperplanes are affine and convex; halfspaces are convex

Slide credit: Stephen Boyd

Example: Euclidean balls

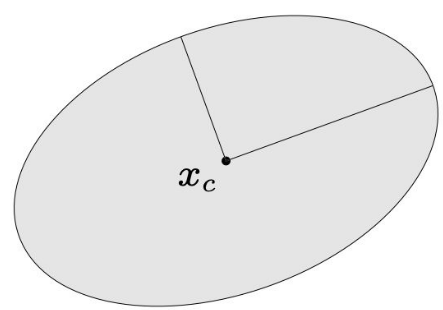
(Euclidean) ball with center x_c and radius r:

$$B(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\} = \{x_c + ru \mid \|u\|_2 \leq 1\}$$

ellipsoid: set of the form

$$\{x \mid (x - x_c)^T P^{-1} (x - x_c) \leq 1\}$$

with $P \in \mathbf{S}^n_{++}$ ($i.e.$, P symmetric positive definite)

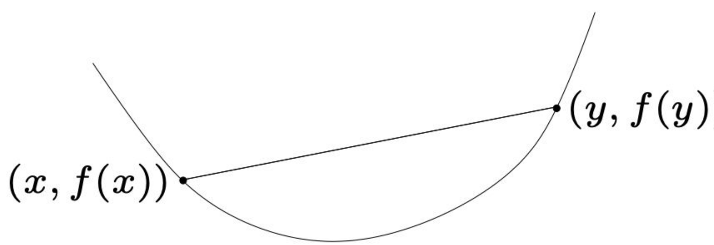


Convex Functions

$f : \mathbf{R}^n \to \mathbf{R}$ is convex if $\mathbf{dom}\, f$ is a convex set and

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$$

for all $x, y \in \mathbf{dom}\, f$, $0 \leq \theta \leq 1$



- f is concave if $-f$ is convex
- f is strictly convex if $\mathbf{dom}\, f$ is convex and

$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta) f(y)$$

for $x, y \in \mathbf{dom}\, f$, $x \neq y$, $0 < \theta < 1$

Examples of convex functions

convex:

- affine: $ax + b$ on **R**, for any $a, b \in$ **R**
- exponential: e^{ax}, for any $a \in$ **R**
- powers: x^α on $\mathbf{R}_{++}$, for $\alpha \geq 1$ or $\alpha \leq 0$
- powers of absolute value: $|x|^p$ on **R**, for $p \geq 1$
- negative entropy: $x \log x$ on $\mathbf{R}_{++}$

concave:

- affine: $ax + b$ on **R**, for any $a, b \in$ **R**
- powers: x^α on $\mathbf{R}_{++}$, for $0 \leq \alpha \leq 1$
- logarithm: $\log x$ on $\mathbf{R}_{++}$

Examples of convex functions

affine functions are convex and concave; all norms are convex

examples on $\mathbf{R}^n$

- affine function $f(x) = a^T x + b$
- norms: $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ for $p \geq 1$; $\|x\|_\infty = \max_k |x_k|$

examples on $\mathbf{R}^{m \times n}$ ($m \times n$ matrices)

- affine function

$$f(X) = \mathbf{tr}(A^T X) + b = \sum_{i=1}^m \sum_{j=1}^n A_{ij} X_{ij} + b$$

First-order condition for convexity

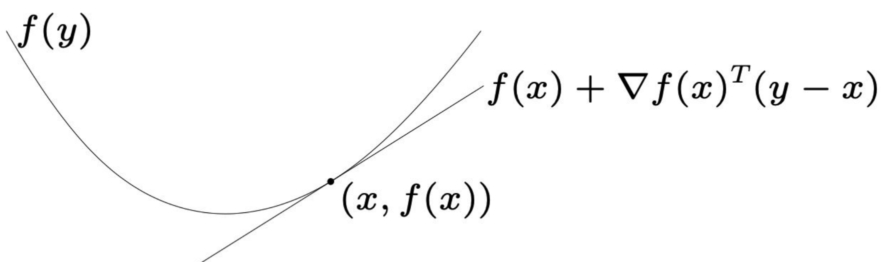
f is **differentiable** if $\mathbf{dom}\, f$ is open and the gradient

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n}\right)$$

exists at each $x \in \mathbf{dom}\, f$

1st-order condition: differentiable f with convex domain is convex iff

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \text{for all } x, y \in \mathbf{dom}\, f$$



first-order approximation of f is global underestimator

Second-order condition for convexity

f is **twice differentiable** if $\mathbf{dom}\, f$ is open and the Hessian $\nabla^2 f(x) \in \mathbf{S}^n$,

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i,j = 1, \ldots, n,$$

exists at each $x \in \mathbf{dom}\, f$

2nd-order conditions: for twice differentiable f with convex domain

- f is convex if and only if

$$\nabla^2 f(x) \succeq 0 \quad \text{for all } x \in \mathbf{dom}\, f$$

(i.e., Hessian matrix at x is positive semi-definite for all x.)

- if $\nabla^2 f(x) \succ 0$ for all $x \in \mathbf{dom}\, f$, then f is strictly convex

(positive definite)

Slide credit: Stephen Boyd

Examples

quadratic function: $f(x) = (1/2)x^T Px + q^T x + r$ (with $P \in \mathbf{S}^n$)

$$\nabla f(x) = Px + q, \qquad \nabla^2 f(x) = P$$

convex if $P \succeq 0$ $\leftarrow$ Definition: P is positive semi-definite.

least-squares objective: $f(x) = \|Ax - b\|_2^2$

$$\nabla f(x) = 2A^T(Ax - b), \qquad \nabla^2 f(x) = 2A^T A$$

convex (for any A)

Jensen's inequality

basic inequality: if f is convex, then for $0 \leq \theta \leq 1$,

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$$

extension: if f is convex, then

$$f(\mathbf{E}\, z) \leq \mathbf{E}\, f(z)$$

for any random variable z

basic inequality is special case with discrete distribution

$$\mathbf{prob}(z = x) = \theta, \qquad \mathbf{prob}(z = y) = 1 - \theta$$

Operations that preserve convexity

practical methods for establishing convexity of a function

1. verify definition (often simplified by restricting to a line)
   $$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y)$$
2. for twice differentiable functions, show $\nabla^2 f(x) \succeq 0$
   (i.e. Hessian is positive semi-definite)
3. show that f is obtained from simple convex functions by operations that preserve convexity
   - nonnegative weighted sum
   - composition with affine function
   - pointwise maximum and supremum
   - composition
   - minimization
   - perspective

Operations that preserve convexity

if $f_1, \ldots, f_m$ are convex, then $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is convex

examples

- piecewise-linear function: $f(x) = \max_{i=1,\ldots,m}(a_i^T x + b_i)$ is convex

composition of $g : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R} \to \mathbf{R}$:

$$f(x) = h(g(x))$$

f is convex if
- g convex, h convex, $\tilde{h}$ nondecreasing
- g concave, h convex, $\tilde{h}$ nonincreasing

General (Constrained) Optimization

- General optimization problem:

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) && \text{objective (cost) function} \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0, i = 1, ..., m && \text{inequality constraint functions} \\ & h_i(\mathbf{x}) = 0, i = 1, ..., p && \text{equality constraint functions} \end{aligned}$$

- If **x** satisfies all the constraints, **x** is called <u>feasible</u> (a feasible solution).
- In general, this is a nontrivial problem to solve, so we use techniques for convex optimization.

Convex Optimization

- Standard form of **convex problem** has the form:

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0, i = 1, ..., m \\ & h_i(\mathbf{x}) = 0, i = 1, ..., p \end{aligned}$$

(where *f*, g_i are convex, *and* h_i are affine)

- Convex optimization problems are a family of optimization problems that are *easier* to solve.

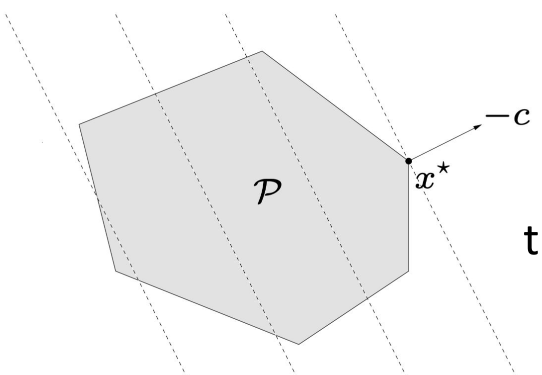
Linear Programming

We say a convex optimization problem is **a linear program (LP)** if both the objective function f and inequality constraints g_i are **affine**.

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

← element-wise inequality
(g_i is the *i*-th row of G)

where $x \in \mathbb{R}^n, c \in \mathbb{R}^n, d \in \mathbb{R}, G \in \mathbb{R}^{m \times n}, h \in \mathbb{R}^m, A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p$



the feasible set is a polyhedron

Linear Programming: applications

diet problem: choose quantities $x_1, \ldots, x_n$ of n foods

- one unit of food j costs c_j, contains amount a_{ij} of nutrient i
- healthy diet requires nutrient i in quantity at least b_i

to find cheapest healthy diet,

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

piecewise-linear minimization

$$\text{minimize} \quad \max_{i=1,\ldots,m}(a_i^T x + b_i)$$

equivalent to an LP

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & a_i^T x + b_i \leq t, \quad i = 1, \ldots, m \end{array}$$

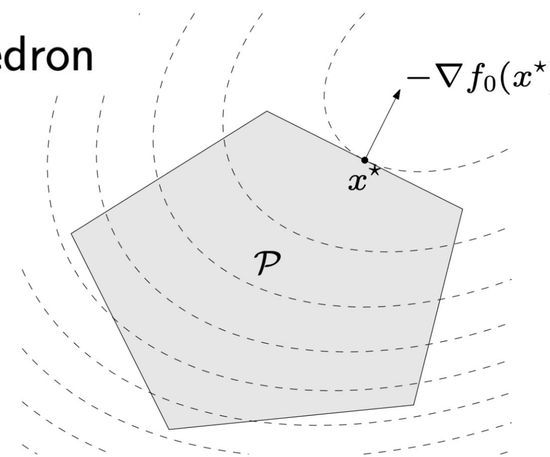
Quadratic Programming

We say a convex optimization problem is **a quadratic program (QP)** if f is a convex quadratic function, and g_i are **affine**.

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{array}$$

← element-wise inequality
(g_i is the *i*-th row of G)

- $P \in \mathbf{S}^n_+$, so objective is convex quadratic
- minimize a convex quadratic function over a polyhedron



Quadratic Programming: applications

least-squares

$$\text{minimize} \quad \|Ax - b\|_2^2$$

- analytical solution $x^\star = A^\dagger b$ ($A^\dagger$ is pseudo-inverse)
- can add linear constraints, *e.g.*, $l \preceq x \preceq u$

Quadratic Programming: applications

linear program with random cost

$$\begin{array}{ll}\text{minimize} & \bar{c}^T x + \gamma x^T \Sigma x = \mathbf{E}\, c^T x + \gamma\, \mathbf{var}(c^T x) \\ \text{subject to} & Gx \preceq h, \quad Ax = b\end{array}$$

- c is random vector with mean $\bar{c}$ and covariance Σ
- hence, $c^T x$ is random variable with mean $\bar{c}^T x$ and variance $x^T \Sigma x$
- $\gamma > 0$ is risk aversion parameter; controls the trade-off between expected cost and variance (risk)

Solving Constrained Optimization: General Overview and Recipe

General (Constrained) Optimization

- General optimization problem:

$$\min_{\mathbf{x}} \quad f(\mathbf{x})$$ objective (cost) function

$$\text{subject to} \quad g_i(\mathbf{x}) \leq 0, i = 1, ..., m$$ inequality constraint functions

$$h_i(\mathbf{x}) = 0, i = 1, ..., p$$ equality constraint functions

A Big Picture

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0,\ i = 1, ..., m \\ & h_i(\mathbf{x}) = 0,\ i = 1, ..., p \end{aligned}$$

Constrained Optimization Problem

$$\mathcal{L}(\mathbf{x}, \lambda, \nu) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x})$$

Lagrangian

Primal Optimization Problem (min-max)

$$\min_{\mathbf{x}} \max_{\nu, \lambda : \lambda_i \geq 0, \forall i} \mathcal{L}(\mathbf{x}, \lambda, \nu)$$

e.g. convex optimizations, KKT conditions

strong duality (if some conditions are met)

$$p^* = d^*$$

weak duality

$$p^* \geq d^*$$

Dual Optimization Problem (max-min)

$$\max_{\nu, \lambda : \lambda_i \geq 0, \forall i} \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda, \nu)$$

Lagrangian Formulation

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0,\ i = 1, ..., m \\ & h_i(\mathbf{x}) = 0,\ i = 1, ..., p \end{aligned}$$

- The **Lagrangian function** is:

$$\mathcal{L}(\mathbf{x}, \lambda, \nu) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x})$$

  – Here, $\lambda = [\lambda_1, \ldots, \lambda_m]$ $(\lambda_i \geq 0 \ \ \forall i)$ and $\nu = [\nu_1, \ldots, \nu_p]$ are called **Lagrange multipliers** (or **dual variables**)

    - 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 $h_i(x) = 0$

Example: Lagrangian Derivation

Consider the following problem:

$$\begin{aligned} \text{minimize}_{x \in \mathbb{R}^2} \quad & (2x_1 - 1)^2 + (x_2 - 2)^2 \\ \text{subject to} \quad & 3x_1 + 2x_2 \leq 4 \\ & x_2 \geq x_1 \end{aligned}$$

Lagrangian:

$$\begin{aligned} \mathcal{L}(x, \lambda) &= (2x_1 - 1)^2 + (x_2 - 2)^2 + \lambda_1(3x_1 + 2x_2 - 4) + \lambda_2(x_1 - x_2) \\ &= \left(2x_1 - \frac{4 - 3\lambda_1 - \lambda_2}{4}\right)^2 + \left(x_2 - \frac{4 - 2\lambda_1 + \lambda_2}{2}\right)^2 \\ &\quad - \frac{1}{16}\left[25\lambda_1^2 + 5\lambda_2^2 - 10\lambda_1\lambda_2 - 24\lambda_1 + 24\lambda_2\right] \end{aligned}$$

(see p.31)

Lagrangian Formulation

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0,\ i = 1, \ldots, p \end{aligned}$$

- The **Lagrangian function** is:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x})$$

- This leads to a primal optimization problem (see the next slide):

$$\min_{\mathbf{x}} \max_{\boldsymbol{\nu}, \boldsymbol{\lambda}:\, \lambda_i \geq 0\, \forall i} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$$

- Difficult to solve directly!

Primal and Feasibility

- Primal optimization problem:

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0,\ i = 1, \ldots, p \end{aligned}$$

$$p^* = \min_{\mathbf{x}} \max_{\boldsymbol{\nu}, \boldsymbol{\lambda} :\, \lambda_i \geq 0 \, \forall i} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$$

where $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x})$

- Notice that:

$$\mathcal{L}_p(\mathbf{x}) = \max_{\nu, \lambda : \lambda_i \geq 0, \forall i} \mathcal{L}(\mathbf{x}, \lambda, \nu) = \begin{cases} f(\mathbf{x}) & \text{if } \mathbf{x} \text{ is feasible} \\ \infty & \text{otherwise} \end{cases}$$

This eliminates the constraints on $\mathbf{x}$, yielding an equivalent optimization problem.

Lagrange Dual

primal vs dual: switching the order of min / max

Note: these are different problems!

- Dual optimization problem:

$$d^* = \max_{\boldsymbol{\nu}, \boldsymbol{\lambda}:\, \lambda_i \geq 0\, \forall i} \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$$

cf) primal optimization problem

$$p^* = \min_{\mathbf{x}} \max_{\boldsymbol{\nu}, \boldsymbol{\lambda}: \lambda_i \geq 0, \forall i} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$$

- We can also write as:

$$\begin{aligned} \max_{\boldsymbol{\lambda}, \boldsymbol{\nu}} \min_{\mathbf{x}} \quad & \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{subject to} \quad & \lambda_i \geq 0, \forall i \end{aligned}$$

$$\begin{aligned} \max_{\boldsymbol{\lambda}, \boldsymbol{\nu}} \quad & \tilde{\mathcal{L}}(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{subject to} \quad & \lambda_i \geq 0, \forall i \\ \text{where} \quad & \tilde{\mathcal{L}}(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \end{aligned}$$

Lagrange Dual function

Example: Dual Problem Derivation

Consider the following problem:

$$\begin{aligned} \text{minimize}_{x\in\mathbb{R}^2} \quad & (2x_1-1)^2+(x_2-2)^2 \\ \text{subject to} \quad & 3x_1+2x_2 \le 4 \\ & x_2 \ge x_1 \end{aligned}$$

Lagrangian:

$$\begin{aligned} \mathcal{L}(x,\lambda) &= (2x_1-1)^2+(x_2-2)^2+\lambda_1(3x_1+2x_2-4)+\lambda_2(x_1-x_2) \\ &= \left(2x_1-\frac{4-3\lambda_1-\lambda_2}{4}\right)^2+\left(x_2-\frac{4-2\lambda_1+\lambda_2}{2}\right)^2 \\ &\quad -\frac{1}{16}\left[25\lambda_1^2+5\lambda_2^2-10\lambda_1\lambda_2-24\lambda_1+24\lambda_2\right] \end{aligned}$$

Dual Objective:

$$g(\lambda)=\min_x \mathcal{L}(x,\lambda)=-\frac{1}{16}\left[25\lambda_1^2+5\lambda_2^2-10\lambda_1\lambda_2-24\lambda_1+24\lambda_2\right]$$

Example: Quadratic Programming

Primal problem:

$$\begin{array}{ll} \text{minimize} & x^T P x \qquad (P \succ 0) \\ \text{subject to} & Ax \preceq b \end{array}$$

Lagrangian: $\mathcal{L} = x^T P x + \lambda^T (Ax - b)$

Primal: $\min_x \max_{\lambda \geq 0} \left\{ x^T P x + \lambda^T (Ax - b) \right\}$

Dual: $\max_{\lambda \geq 0} \min_x \left\{ x^T P x + \lambda^T (Ax - b) \right\}$

Dual Function: $\tilde{\mathcal{L}}(\lambda) = \min_x (x^T P x + \lambda^T (Ax - b)) = -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda$

Dual:

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

Weak Duality

strong duality (if some conditions are met)

$p^* = d^*$

Primal Optimization Problem (min-max)

$$\min_{\mathbf{x}} \max_{\boldsymbol{\nu},\boldsymbol{\lambda}:\lambda_i \geq 0, \forall i} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$$

weak duality

$p^* \geq d^*$

Dual Optimization Problem (max-min)

$$\max_{\boldsymbol{\nu},\boldsymbol{\lambda}:\lambda_i \geq 0, \forall i} \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$$

- Claim:

$$\begin{aligned} d^* &= \max_{\boldsymbol{\nu},\boldsymbol{\lambda}:\lambda_i \geq 0} \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \\ &\leq \min_{\mathbf{x}} \max_{\boldsymbol{\nu},\boldsymbol{\lambda}:\lambda_i \geq 0} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \\ &= p^* \end{aligned}$$

- Difference between p^* and d^* is called the **duality gap**.
- In other words, the dual maximization problem (usually easier) gives a "**lower bound**" for the primal minimization problem (usually more difficult).

Weak Duality

$$d^* = \max_{\boldsymbol{\nu}, \boldsymbol{\lambda}: \lambda_i \geq 0} \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq \min_{\mathbf{x}} \max_{\boldsymbol{\nu}, \boldsymbol{\lambda}: \lambda_i \geq 0} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = p^*$$

- Proof: Let $\tilde{\mathbf{x}}$ be feasible. Then for any $\boldsymbol{\lambda}, \boldsymbol{\nu}$ with $\lambda_i \geq 0$,

$$\mathcal{L}(\tilde{\mathbf{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \lambda_i g_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{\mathbf{x}}) \leq f(\tilde{\mathbf{x}})$$

$$\boxed{\begin{array}{c}\text{Thus,} \quad \tilde{\mathcal{L}}(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq \mathcal{L}(\tilde{\mathbf{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f(\tilde{\mathbf{x}}) \\ \text{for any } \boldsymbol{\lambda}, \boldsymbol{\nu} \text{ with } \lambda_i \geq 0, \text{ any feasible } \tilde{\mathbf{x}}\end{array}}$$

Then,

$$d^* = \max_{\boldsymbol{\nu}, \boldsymbol{\lambda}: \lambda_i \geq 0} \tilde{\mathcal{L}}(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f(\tilde{\mathbf{x}}) \text{ for any feasible } \tilde{\mathbf{x}}$$

Finally,

$$d^* = \max_{\boldsymbol{\nu}, \boldsymbol{\lambda}: \lambda_i \geq 0} \tilde{\mathcal{L}}(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq \min_{\tilde{\mathbf{x}}:\text{feasible}} f(\tilde{\mathbf{x}}) = p^*$$

Strong Duality

- If $p^* = d^*$, we say **strong duality** holds.
- What are the conditions for strong duality?
  - does not hold in general
  - usually holds for convex problems (under mild conditions)
  - conditions that guarantee strong duality in convex problems are called constraint qualification.
- Two well-known conditions (in convex problems)
  - Slater's constraint qualification
  - Karush-Kuhn-Tucker (KKT) condition (main focus)

(Sufficient) Conditions for strong duality: **Slater's constraint qualification**

- Strong duality holds for a **convex** problem

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0, i = 1, ..., m \\ & h_i(\mathbf{x}) = 0, i = 1, ..., p \end{aligned}$$

(where f, g_i are **convex**, *and* h_i are **affine**)

if the constraint is <u>strictly</u> feasible (by any solution), i.e.,

$$\begin{aligned} \exists \mathbf{x} : \quad & g_i(\mathbf{x}) < 0, \forall i = 1, ..., m \\ & h_i(\mathbf{x}) = 0, \forall i = 1, ..., p \end{aligned}$$

(Not necessarily an optimal solution)

Slater's condition is a sufficient condition for strong duality to hold for a convex problem

(Sufficient) Conditions for strong duality: **Slater's constraint qualification (A special case: affine constraints)**

- Strong duality holds for a **convex** problem with **affine** constraints:

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0, i = 1, ..., m \\ & h_i(\mathbf{x}) = 0, i = 1, ..., p \end{aligned}$$

(where f is **convex**, *and* g_i, h_i are all **affine**)

if the problem is feasible.

(Note) this version of the Slater's constraint does not require strict feasibility.

Example: Quadratic Programming

Primal:

$$\begin{array}{ll} \text{minimize} & x^T P x \qquad\qquad (P \succ 0) \\ \text{subject to} & Ax \preceq b \end{array}$$

Dual function:

$$\tilde{\mathcal{L}}(\lambda) = \min_x (x^T P x + \lambda^T (Ax - b)) = -\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

Dual:

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

From the Slater's condition, $p^* = d^*$ if $A\tilde{x} \prec b$ for some $\tilde{x}$

For QP, this is always satisfied, i.e. strong duality always holds

Example: SVM (Support Vector Machine)

Primal:

(See Lecture 10 for details)

$$\min_{\mathbf{w},\boldsymbol{\xi}} \quad C\sum_{n=1}^{N}\xi^{(n)} + \frac{1}{2}\|\mathbf{w}\|^2$$

$$\text{subject to} \quad y^{(n)}\left(\mathbf{w}^\top\phi(\mathbf{x}^{(n)}) + b\right) \geq 1 - \xi^{(n)}, \quad n = 1, ..., N$$

$$\xi^{(n)} \geq 0, \quad n = 1, ..., N$$

Dual:

$$\max_{\boldsymbol{\alpha}} \quad \sum_{n=1}^{N}\alpha^{(n)} - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{M}\alpha^{(n)}\alpha^{(m)}y^{(n)}y^{(m)}k\left(\mathbf{x}^{(n)}, \mathbf{x}^{(m)}\right)$$

$$\text{subject to} \quad 0 \leq \alpha^{(n)} \leq C \quad \sum_{n=1}^{N}\alpha^{(n)}y^{(n)} = 0$$

The primal problem is a convex problem with affine constraints.

By the Slater's constraint qualification, **strong duality** holds for SVM.

Karush-Kuhn-Tucker (KKT) condition

Let $\mathbf{x}^*$ be a primal optimal and $\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*$ be a dual optimal solution.
If the strong duality holds, then we have the following:

$$\nabla_{\mathbf{x}} f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla_{\mathbf{x}} g_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* \nabla_{\mathbf{x}} h_i(\mathbf{x}^*) = 0,$$ Stationarity (1)

$$g_i(\mathbf{x}^*) \leq 0, \quad i = 1, \dots, m,$$ Primal feasibility (2)

$$h_i(\mathbf{x}^*) = 0, \quad i = 1, \dots, p,$$ Primal feasibility (3)

$$\lambda_i^* \geq 0, \quad i = 1, \dots, m,$$ Dual feasibility (4)

$$\boxed{\lambda_i^* g_i(\mathbf{x}^*) = 0, \quad i = 1, \dots, m}$$ Complementary slackness (5)

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0, \; i = 1, \dots, m \\ & h_i(\mathbf{x}) = 0, \; i = 1, \dots, p \end{aligned}$$

$$\begin{aligned} \max_{\lambda,\nu} \min_{\mathbf{x}} \quad & \mathcal{L}(\mathbf{x}, \lambda, \nu) \\ \text{subject to} \quad & \lambda_i \geq 0, \; \forall i \end{aligned}$$

Dual problem

Note: we do **not** assume the optimization problem is necessarily convex for describing KKT condition. However, when the problem is convex (and differentiable), KKT condition ensures strong duality.

Example: KKT Conditions

$$\begin{aligned} \text{minimize}_{x \in \mathbb{R}^2} \quad & (2x_1 - 1)^2 + (x_2 - 2)^2 \\ \text{subject to} \quad & 3x_1 + 2x_2 \leq 4 \\ & x_2 \geq x_1 \end{aligned}$$

KKT Conditions:

$$\begin{bmatrix} 4(2x_1^* - 1) + 3\lambda_1^* + \lambda_2^* \\ 2(x_2^* - 2) + 2\lambda_1^* - \lambda_2^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad (1)$$

$$3x_1^* + 2x_2^* - 4 \leq 0, x_1^* - x_2^* \leq 0 \qquad (2)$$

$$\lambda_1^* \geq 0, \lambda_2^* \geq 0 \qquad (4)$$

$$\begin{aligned} \lambda_1^*(3x_1^* + 2x_2^* - 4) = 0 \\ \lambda_2^*(x_1^* - x_2^*) = 0 \end{aligned} \qquad (5)$$

(1) $\nabla_{\mathbf{x}} f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla_{\mathbf{x}} g_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* \nabla_{\mathbf{x}} h_i(\mathbf{x}^*) = 0,$

(2) $g_i(\mathbf{x}^*) \leq 0, \quad i = 1, \ldots, m,$

(3) $h_i(\mathbf{x}^*) = 0, \quad i = 1, \ldots, p,$

(4) $\lambda_i^* \geq 0, \quad i = 1, \ldots, m,$

(5) $\lambda_i^* g_i(\mathbf{x}^*) = 0, \quad i = 1, \ldots, m$

Strong duality ⇒ Complementary slackness

Let $\mathbf{x}^*$ be a primal optimal and λ^*, ν^* be a dual optimal solution.
Claim: If strong duality holds, $\boxed{\lambda_i^* g_i(\mathbf{x}^*) = 0 \text{ for all } i = 1, \ldots, m.}$

- Provided: Strong duality holds, i.e., $p^* = f(\mathbf{x}^*) = \tilde{\mathcal{L}}(\lambda^*, \nu^*) = d^*$
- $$\begin{aligned} f(\mathbf{x}^*) &= \tilde{\mathcal{L}}(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\ &= \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\ &= \min_{\mathbf{x}} \left(f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* g_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}) \right) \\ &\leq f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* g_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}^*) \\ &\leq f(\mathbf{x}^*) \end{aligned}$$

Primal:
$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0, i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0, i = 1, \ldots, p \end{aligned}$$

Dual:
$$\begin{aligned} \max_{\boldsymbol{\lambda}, \boldsymbol{\nu}} \quad & \tilde{\mathcal{L}}(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{subject to} \quad & \lambda_i \geq 0, \forall i \\ \text{where} \quad & \tilde{\mathcal{L}}(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \end{aligned}$$

- Therefore, $\sum_{i=1}^{m} \lambda_i^* g_i(\mathbf{x}^*) = 0$ and thus $\lambda_i^* g_i(\mathbf{x}^*) = 0$ $(i = 1, \ldots, m)$.

(Sufficient) Conditions for strong duality: KKT Conditions

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \le 0, i = 1, ..., m \\ & h_i(\mathbf{x}) = 0, i = 1, ..., p \end{aligned}$$

- Assume f, g_i, h_i are differentiable
- If the original problem is **convex** (where f, g_i are **convex** and h_i are affine), and $\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*$ satisfy the KKT conditions, then:
  - $\mathbf{x}^*$ is primal optimal
  - $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ is dual optimal, and
  - the duality gap is zero (i.e., strong duality holds)

For convex optimization problems (+ differentiable objectives/constraints)**, KKT is a sufficient condition for strong duality.**

Proof for sufficiency (KKT => Strong duality)

Claim: When KKT (1)-(5) holds, the strong duality holds.

- From (2) and (3), $\mathbf{x}^*$ is primal feasible.
- From (4), $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ is dual feasible.
- $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ is a convex differentiable function.
  Thus, from (1), $\mathbf{x}^*$ is a minimizer of $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$.
- Then,

$$
\begin{aligned}
\tilde{\mathcal{L}}(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) &= \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\
&= \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\
&= f(\mathbf{x}^*) + \underbrace{\sum_i \lambda_i^* g_i(\mathbf{x}^*)}_{=0} + \sum_i \nu_i^* h_i(\mathbf{x}^*) \\
&= f(\mathbf{x}^*)
\end{aligned}
$$

(See also: derivation of complementary slackness)

$= 0 \quad \because$ (5) complementary slackness

- But, $\tilde{\mathcal{L}}(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \leq \underbrace{\max_{\boldsymbol{\lambda}, \boldsymbol{\nu}: \lambda_i \geq 0} \tilde{\mathcal{L}}(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq \min_{\mathbf{x}:\mathbf{x} \text{ is feasible}} f(\mathbf{x})}_{\text{weak duality}} \leq f(\mathbf{x}^*) = \tilde{\mathcal{L}}(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$

- Then, $\max_{\boldsymbol{\lambda}, \boldsymbol{\nu}: \lambda_i \geq 0} \tilde{\mathcal{L}}(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \min_{\mathbf{x}:\mathbf{x} \text{ is feasible}} f(\mathbf{x})$

which proves that the strong duality holds (i.e., duality gap is zero).

KKT conditions: Conclusion

- If a constrained optimization is differentiable and has convex objective function and constraint sets, then the KKT conditions are **(necessary and) sufficient conditions** for **strong duality** (zero duality gap).

- Thus, the KKT conditions can be used to solve such convex optimization problems.

A Big Picture

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0,\ i = 1, \ldots, p \end{aligned}$$

Constrained Optimization Problem

$$\mathcal{L}(\mathbf{x}, \lambda, \nu) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x})$$

Lagrangian

Primal Optimization Problem (min-max)

$$\min_{\mathbf{x}} \max_{\nu, \lambda : \lambda_i \geq 0, \forall i} \mathcal{L}(\mathbf{x}, \lambda, \nu)$$

e.g. convex optimizations, KKT conditions

strong duality (if some conditions are met)

$$p^* = d^*$$

weak duality

$$p^* \geq d^*$$

Dual Optimization Problem (max-min)

$$\max_{\nu, \lambda : \lambda_i \geq 0, \forall i} \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda, \nu)$$

Recap: General Recipe

- Given an original optimization

$$
\begin{aligned}
\min_{\mathbf{x}} \quad & f(\mathbf{x}) \\
\text{subject to} \quad & g_i(\mathbf{x}) \leq 0,\ i = 1, ..., m \\
& h_i(\mathbf{x}) = 0,\ i = 1, ..., p
\end{aligned}
$$

- Solve dual optimization with <u>Lagrangian function</u>:

$$
\begin{aligned}
\max_{\lambda, \nu} \min_{\mathbf{x}} \quad & \mathcal{L}(\mathbf{x}, \lambda, \nu) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \\
\text{subject to} \quad & \lambda_i \geq 0,\ \forall i
\end{aligned}
$$

- Alternatively, solve the dual optimization with <u>Lagrange dual</u>:

$$
\begin{aligned}
\max_{\lambda, \nu} \quad & \tilde{\mathcal{L}}(\lambda, \nu) \qquad \text{where } \tilde{\mathcal{L}}(\lambda, \nu) = \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda, \nu) \\
\text{subject to} \quad & \lambda_i \geq 0,\ \forall i
\end{aligned}
$$

Recap: KKT Optimality condition

- Karush-Kuhn-Tucker (KKT) condition:

$$
\begin{aligned}
&\nabla_{\mathbf{x}} f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla_{\mathbf{x}} g_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* \nabla_{\mathbf{x}} h_i(\mathbf{x}^*) = 0, \\
&g_i(\mathbf{x}^*) \leq 0, \quad i = 1, \dots, m, \\
&h_i(\mathbf{x}^*) = 0, \quad i = 1, \dots, p, \\
&\lambda_i^* \geq 0, \quad i = 1, \dots, m, \\
&\lambda_i^* g_i(\mathbf{x}^*) = 0, \quad i = 1, \dots, m
\end{aligned}
$$

- The last condition is called complementary slackness, and guarantees the strong duality for convex optimization
- In Lecture 10, you'll learn how this condition is used to determine support vectors in SVM

Additional Resource

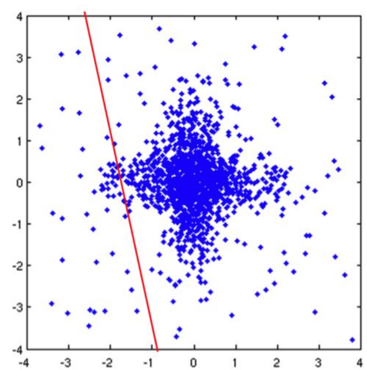
- Convex Optimization
  - http://www.stanford.edu/~boyd/cvxbook/
  - http://www.stanford.edu/class/ee364a/
  - For materials covered today, see Chapter 5 (and earlier chapters).

Exercise: One-class SVM

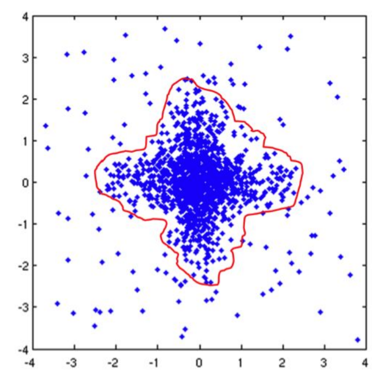
Given an unlabeled set of examples $\{\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(N)}\}$, the *one-class SVM* algorithm tries to find a direction $\mathbf{w}$ that maximally separates the data from the origin.[1]

More precisely, it solves the (primal) optimization problem:

$$
\begin{aligned}
\min_{\mathbf{w}} \quad & \frac{1}{2}\mathbf{w}^\top \mathbf{w} + \frac{C}{2}\sum_{i=1}^{N} \xi_i^2 \\
\text{s.t.} \quad & \mathbf{w}^\top \phi(\mathbf{x}^{(i)}) \geq 1 - \xi_i, \ i \in \{1, \cdots, N\} \\
& \xi_i \geq 0, \ i \in \{1, \cdots, N\}
\end{aligned}
$$



(a) using linear kernel



(b) using RBF kernel

Exercise: One-class SVM

Primal:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\mathbf{w}^\top\mathbf{w} + \frac{C}{2}\sum_{i=1}^{N}\xi_i^2$$

$$\text{s.t.} \quad \mathbf{w}^\top\phi(\mathbf{x}^{(i)}) \geq 1 - \xi_i,\ i \in \{1, \cdots, N\}$$

Lagrangian:

Exercise: One-class SVM

Primal:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\mathbf{w}^\top\mathbf{w} + \frac{C}{2}\sum_{i=1}^{N}\xi_i^2$$

$$\text{s.t.} \quad \mathbf{w}^\top\phi(\mathbf{x}^{(i)}) \geq 1 - \xi_i,\ i \in \{1, \cdots, N\}$$

Lagrangian:

$$\mathcal{L}(\mathbf{w}, \boldsymbol{\xi}, \boldsymbol{\alpha}) = \frac{1}{2}\mathbf{w}^\top\mathbf{w} + \frac{C}{2}\sum_{i=1}^{N}\xi_i^2 + \sum_{i=1}^{N}\alpha_i\left(1 - \xi_i - \mathbf{w}^\top\phi(x^{(i)})\right)$$

$$\alpha_i \geq 0, \quad i = 1, \ldots, N$$

KKT Conditions:

$$\nabla_{\mathbf{x}} f(\mathbf{x}^*) + \sum_{i=1}^{m}\lambda_i\nabla_{\mathbf{x}} g_i(\mathbf{x}^*) + \sum_{i=1}^{p}\nu_i^*\nabla_{\mathbf{x}} h_i(\mathbf{x}^*) = 0,$$

$$g_i(\mathbf{x}^*) \leq 0, \quad i = 1, \ldots, m,$$

$$h_i(\mathbf{x}^*) = 0, \quad i = 1, \ldots, p,$$

$$\lambda_i^* \geq 0, \quad i = 1, \ldots, m,$$

$$\lambda_i^* g_i(\mathbf{x}^*) = 0, \quad i = 1, \ldots, m$$

Exercise: One-class SVM

Primal:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\mathbf{w}^\top\mathbf{w} + \frac{C}{2}\sum_{i=1}^{N}\xi_i^2$$

$$\text{s.t.} \quad \mathbf{w}^\top\phi(\mathbf{x}^{(i)}) \geq 1 - \xi_i,\ i \in \{1,\cdots,N\}$$

Lagrangian:

$$\mathcal{L}(\mathbf{w},\boldsymbol{\xi},\boldsymbol{\alpha}) = \frac{1}{2}\mathbf{w}^\top\mathbf{w} + \frac{C}{2}\sum_{i=1}^{N}\xi_i^2 + \sum_{i=1}^{N}\alpha_i\left(1 - \xi_i - \mathbf{w}^\top\phi(x^{(i)})\right)$$

$$\alpha_i \geq 0, \quad i = 1,\ldots,N$$

KKT Conditions:

$$\frac{\partial\mathcal{L}}{\partial\mathbf{w}} = \mathbf{w} - \sum_{i=1}^{N}\alpha_i\phi(\mathbf{x}^{(i)}) = 0$$

$$\frac{\partial\mathcal{L}}{\partial\xi_i} = C\xi_i - \alpha_i = 0$$

$$\alpha_i(1 - \xi_i - \mathbf{w}^\top\phi(\mathbf{x}^{(i)})) = 0$$

$$\mathbf{w}^\top\phi(\mathbf{x}^{(i)}) \geq 1 - \xi_i$$

$$\alpha_i \geq 0$$

$$\boxed{\begin{aligned}\mathbf{w} &= \sum_{i=1}^{N}\alpha_i\phi\left(\mathbf{x}^{(i)}\right),\\ \xi_i &= \frac{\alpha_i}{C},\ i \in \{1,\cdots,N\}.\end{aligned}}$$

Exercise: One-class SVM

Primal (min-max): $\min_{\mathbf{w},\boldsymbol{\xi}} \max_{\boldsymbol{\alpha}\succeq 0} \mathcal{L}(\mathbf{w},\boldsymbol{\xi},\boldsymbol{\alpha})$

Dual (max-min): $\max_{\boldsymbol{\alpha}\succeq 0} \min_{\mathbf{w},\boldsymbol{\xi}} \mathcal{L}(\mathbf{w},\boldsymbol{\xi},\boldsymbol{\alpha})$

$$\begin{aligned} \mathbf{w} &= \sum_{i=1}^{N} \alpha_i \phi\left(\mathbf{x}^{(i)}\right), \\ \xi_i &= \frac{\alpha_i}{C},\ i \in \{1, \cdots, N\}. \end{aligned}$$

$$\mathcal{L}(\mathbf{w},\boldsymbol{\xi},\boldsymbol{\alpha}) = \frac{1}{2}\mathbf{w}^\top\mathbf{w} + \frac{C}{2}\sum_{i=1}^{N}\xi_i^2 + \sum_{i=1}^{N}\alpha_i\left(1 - \xi_i - \mathbf{w}^\top\phi(x^{(i)})\right)$$

Solve the Lagrange dual function $\min_{\mathbf{w},\boldsymbol{\xi}} \mathcal{L}(\mathbf{w},\boldsymbol{\xi},\boldsymbol{\alpha})$

Exercise: One-class SVM

$$\mathcal{L}(\mathbf{w}, \boldsymbol{\xi}, \boldsymbol{\alpha}) = \frac{1}{2}\mathbf{w}^\top \mathbf{w} + \frac{C}{2}\sum_{i=1}^{N} \xi_i^2 + \sum_{i=1}^{N} \alpha_i \left(1 - \xi_i - \mathbf{w}^\top \phi(x^{(i)})\right)$$

$$\boxed{\begin{aligned} \mathbf{w} &= \sum_{i=1}^{N} \alpha_i \phi\left(\mathbf{x}^{(i)}\right), \\ \xi_i &= \frac{\alpha_i}{C},\ i \in \{1, \cdots, N\}. \end{aligned}}$$

$$\min_{\mathbf{w}, \boldsymbol{\xi}} \mathcal{L}(\mathbf{w}, \boldsymbol{\xi}, \boldsymbol{\alpha}) =$$

Exercise: One-class SVM

$$\mathcal{L}(\mathbf{w}, \boldsymbol{\xi}, \boldsymbol{\alpha}) = \frac{1}{2}\mathbf{w}^\top \mathbf{w} + \frac{C}{2}\sum_{i=1}^{N} \xi_i^2 + \sum_{i=1}^{N} \alpha_i \left(1 - \xi_i - \mathbf{w}^\top \phi(x^{(i)})\right)$$

$$\boxed{\begin{aligned} \mathbf{w} &= \sum_{i=1}^{N} \alpha_i \phi\left(\mathbf{x}^{(i)}\right), \\ \xi_i &= \frac{\alpha_i}{C}, \; i \in \{1, \cdots, N\}. \end{aligned}}$$

$$\min_{\mathbf{w}, \boldsymbol{\xi}} \mathcal{L}(\mathbf{w}, \boldsymbol{\xi}, \boldsymbol{\alpha}) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \left(\phi(\mathbf{x}^{(i)})^\top \phi(\mathbf{x}^{(j)}) + \frac{1}{C}\delta_{i,j}\right) + \sum_{i=1}^{N} \alpha_i$$

Dual Problem:

$$\max_{\boldsymbol{\alpha}} \quad -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \left(\phi(\mathbf{x}^{(i)})^\top \phi(\mathbf{x}^{(j)}) + \frac{1}{C}\delta_{i,j}\right) \alpha_j + \sum_{i=1}^{N} \alpha_i$$

$$\text{subject to} \;\; \alpha_i \geq 0$$