DUALITY

Lagrangian

standard form problem (not necessarily convex)

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

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

The basic idea : take the constraints into account by augmenting the objective function with a weighted sum of the constraint functions.

Lagrangian

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

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

Lagrangian

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

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

- 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 $h_i(x) = 0$

Lagrange dual function

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

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

Lagrange dual function

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

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

concave

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

Lagrange dual function

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

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

concave

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

proof: 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 $p^\star \geq g(\lambda, \nu)$

Example

Least-squares solution of linear equations

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

- Lagrangian is $L(x, \nu) = x^T x + \nu^T (Ax - b)$

Example

Least-squares solution of linear equations

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

Convex quadratic function

optimality condition

- Lagrangian is $L(x, \nu) = x^T x + \nu^T (Ax - b)$

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0 \quad \Rightarrow \quad x = -(1/2) A^T \nu$$

Example

Least-squares solution of linear equations

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

Convex quadratic function

optimality condition

- Lagrangian is $L(x, \nu) = x^T x + \nu^T (Ax - b)$

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0 \quad \Rightarrow \quad x = -(1/2) A^T \nu$$

$$g(\nu) = L((-1/2)A^T \nu, \nu) = -\frac{1}{4} \nu^T A A^T \nu - b^T \nu$$

concave

lower bound property: $p^\star \geq -(1/4)\nu^T A A^T \nu - b^T \nu$ for all ν

Example

Standard form LP

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

$$L(x, \lambda, \nu) = c^T x - \sum_{i=1}^{n} \lambda_i x_i + \nu^T (Ax - b) = -b^T \nu + (c + A^T \nu - \lambda)^T x.$$

Example

Standard form LP

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

$$L(x, \lambda, \nu) = c^T x - \sum_{i=1}^{n} \lambda_i x_i + \nu^T (Ax - b) = -b^T \nu + (c + A^T \nu - \lambda)^T x.$$

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = -b^T \nu + \inf_x (c + A^T \nu - \lambda)^T x,$$

Example

Standard form LP

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

$$L(x, \lambda, \nu) = c^T x - \sum_{i=1}^{n} \lambda_i x_i + \nu^T (Ax - b) = -b^T \nu + (c + A^T \nu - \lambda)^T x.$$

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = -b^T \nu + \inf_x (c + A^T \nu - \lambda)^T x,$$

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

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

Dual problem

For each pair (λ, ν) with $\lambda \succeq 0$, the Lagrange dual function gives us a lower bound on the optimal value $p^\star$ of the optimization problem

What is the best lower bound that can be obtained from the Lagrange dual function?

Dual problem

What is the best lower bound that can be obtained from the Lagrange dual function?

For each pair (λ, ν) with $\lambda \succeq 0$, the Lagrange dual function gives us a lower bound on the optimal value $p^\star$ of the optimization problem

Lagrange dual problem

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

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

Dual problem

What is the best lower bound that can be obtained from the Lagrange dual function?

For each pair (λ, ν) with $\lambda \succeq 0$, the Lagrange dual function gives us a lower bound on the optimal value $p^\star$ of the optimization problem

Lagrange dual problem

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

$$\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \le 0, & i = 1, \ldots, m \\ & h_i(x) = 0, & i = 1, \ldots, p, \end{array}$$

- λ, ν are dual feasible if $\lambda \succeq 0$, $(\lambda, \nu) \in \mathbf{dom}\, g$
- finds best lower bound on $p^\star$, obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted $d^\star$

Example

Lagrange dual of standard form LP

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

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

Example

Lagrange dual of standard form LP

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

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

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) = \begin{cases} -b^T\nu & A^T\nu - \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases} \\ \text{subject to} & \lambda \succeq 0. \end{array}$$

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

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

Example

Lagrange dual of inequality form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b. \end{array} \quad \rightarrow \quad L(x,\lambda) = c^T x + \lambda^T (Ax - b) = -b^T \lambda + (A^T \lambda + c)^T x,$$

Example

Lagrange dual of inequality form LP

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

$$L(x, \lambda) = c^T x + \lambda^T (Ax - b) = -b^T \lambda + (A^T \lambda + c)^T x,$$

$$g(\lambda) = \inf_x L(x, \lambda) = -b^T \lambda + \inf_x (A^T \lambda + c)^T x.$$

Example

Lagrange dual of inequality form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b. \end{array} \rightarrow L(x,\lambda) = c^T x + \lambda^T (Ax - b) = -b^T \lambda + (A^T \lambda + c)^T x,$$

$$g(\lambda) = \inf_x L(x,\lambda) = -b^T \lambda + \inf_x (A^T \lambda + c)^T x.$$

$$g(\lambda) = \begin{cases} -b^T \lambda & A^T \lambda + c = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

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

Weak Duality

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

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

weak duality: $d^\star \leq p^\star$

- always holds (for convex and nonconvex problems)

$$p^\star - d^\star \longrightarrow \text{optimal duality gap}$$

- can be used to find nontrivial lower bounds for difficult problems

Strong Duality

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

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

Slater's constraint qualification

Convex problem

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

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

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

strong duality holds

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

Slater’s constraint qualification

Slater's condition can be refined when some of the inequality constraint functions are **affine**.

If the first k constraint functions $f_1, \ldots, f_k$ are affine,

$$\exists x \in \textbf{int}\, \mathcal{D} : \; f_i(x) \leq 0, \quad i = 1, \ldots, k, \qquad f_i(x) < 0, \quad i = k+1, \ldots, m, \qquad Ax = b.$$

refined Slater condition reduces to feasibility when the constraints are all linear equalities and inequalities

Example

Inequality form LP

primal problem

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

Example

Inequality form LP

primal problem

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

dual function

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

dual problem

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

Example

Quadratic program

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

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

Example

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

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

- in fact, $p^\star = d^\star$ always

Complementary slackness

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

$$f_0(x^\star) = g(\lambda^\star, \nu^\star) = \inf_x \left(f_0(x) + \sum_{i=1}^{m} \lambda_i^\star f_i(x) + \sum_{i=1}^{p} \nu_i^\star h_i(x) \right)$$

Complementary slackness

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

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

Complementary slackness

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

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

Complementary slackness

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

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

hence, the two inequalities hold with equality

- $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$

Complementary slackness

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

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

hence, the two inequalities hold with equality

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

$$\boxed{\lambda_i^\star > 0 \implies f_i(x^\star) = 0, \qquad f_i(x^\star) < 0 \implies \lambda_i^\star = 0}$$

Karush-Kuhn-Tucker(KKT) conditions

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

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

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

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

KKT conditions for convex problems

When the primal problem is convex, the KKT conditions are also sufficient for the points to be primal and dual optimal.

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

- from 4th condition (and convexity): $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$

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

- from complementary slackness: $f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$ hence, $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$

for any convex optimization problem with differentiable objective and constraint functions, any points that satisfy the KKT conditions are primal and dual optimal, and have zero duality gap.

Example

Equality constrained convex quadratic minimization

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

$P \in \mathbf{S}_+^n$

KKT conditions for this problem

Example

Equality constrained convex quadratic minimization

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

$P \in \mathbf{S}_+^n$

KKT conditions for this problem

$$Ax^\star = b$$

$$Px^\star + q + A^T \nu^\star = 0,$$

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ \nu^\star \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$$

Duality and problem reformulations

- ✓equivalent formulations of a problem can lead to very different duals
- ✓reformulating the primal problem can be useful when the dual is difficult to derive, or uninteresting

The conjugate function

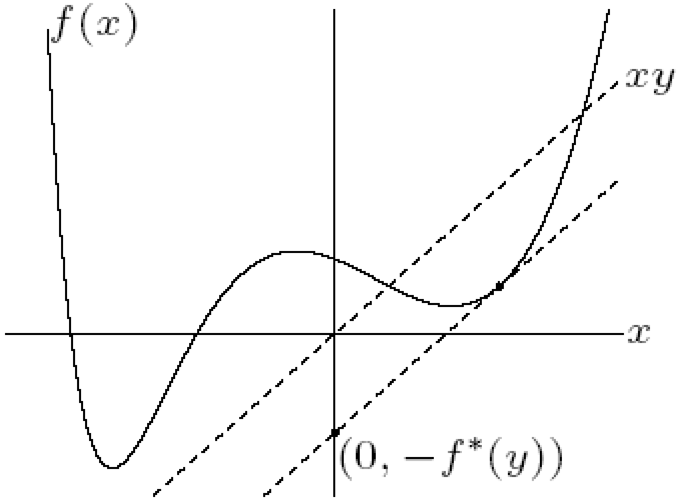
the conjugate of a function f is

$$f^*(y) = \sup_{x \in \mathbf{dom}\, f} \left(y^T x - f(x) \right)$$

The conjugate function

the conjugate of a function f is

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

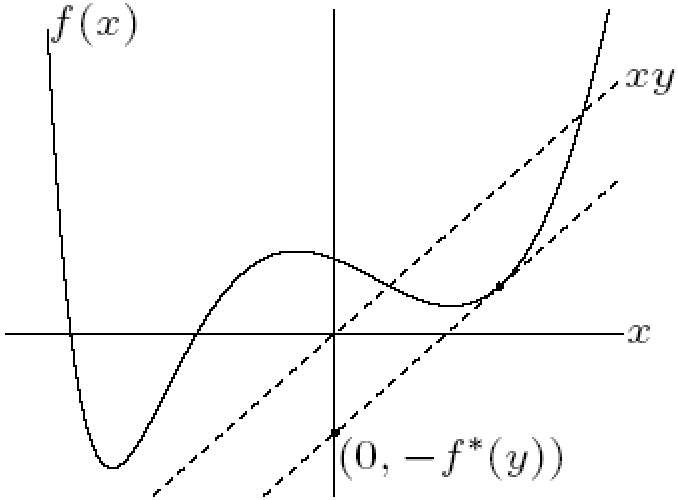


- f^* is convex (even if f is not)

The conjugate function

the conjugate of a function f is

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



- f^* is convex (even if f is not)

f is convex and differentiable ⇒ maximum gap occurs in $y = \nabla f(x^*)$

Introducing new variables and equality constraints

$$\boxed{\text{minimize} \quad f_0(Ax + b)}$$

Introducing new variables and equality constraints

$$\boxed{\text{minimize} \quad f_0(Ax+b)}$$

- dual function is constant: $g = \inf_x L(x) = \inf_x f_0(Ax+b) = p^\star$
- we have strong duality, but dual is quite useless

Introducing new variables and equality constraints

$$\text{minimize} \quad f_0(Ax+b)$$

- dual function is constant: $g = \inf_x L(x) = \inf_x f_0(Ax+b) \; = p^\star$
- we have strong duality, but dual is quite useless

reformulated problem and its dual

$$\begin{array}{ll} \text{minimize} & f_0(y) \\ \text{subject to} & Ax+b-y=0 \end{array}$$

dual function follows from

$$g(\nu) \;=\; \inf_{x,y}(f_0(y) - \nu^T y + \nu^T Ax + b^T \nu)$$

Introducing new variables and equality constraints

$$\text{minimize} \quad f_0(Ax+b)$$

- dual function is constant: $g = \inf_x L(x) = \inf_x f_0(Ax+b) \ = p^\star$
- we have strong duality, but dual is quite useless

reformulated problem and its dual

$$\begin{array}{ll} \text{minimize} & f_0(y) \\ \text{subject to} & Ax+b-y=0 \end{array}$$

dual function follows from

$$g(\nu) \quad = \quad \inf_{x,y}(f_0(y) - \nu^T y + \nu^T Ax + b^T\nu) \quad = \quad \begin{cases} -f_0^*(\nu) + b^T\nu & A^T\nu = 0 \\ -\infty & \text{otherwise} \end{cases}$$

$$b^T\nu + \inf_y (f_0(y) - \nu^T y) = b^T\nu - f_0^*(\nu),$$

$$\begin{array}{ll} \text{maximize} & b^T\nu - f_0^*(\nu) \\ \text{subject to} & A^T\nu = 0. \end{array}$$

Dual Norm

Let $\|\cdot\|$ be a norm on $\mathbf{R}^n$. The associated *dual norm*, denoted $\|\cdot\|_*$, is defined as

$$\|z\|_* = \sup\{z^T x \mid \|x\| \leq 1\}$$

Dual Norm

Let $||\cdot||$ be a norm on $\mathbf{R}^n$. The associated *dual norm*, denoted $||\cdot||_*$, is defined as

$$||z||_* = \sup\{z^T x \mid ||x|| \leq 1\}$$

The dual of the Euclidean norm is the Euclidean norm

The dual of the ℓ_1-norm is the ℓ_∞-norm

dual of the ℓ_∞-norm is the ℓ_1-norm

the dual of the ℓ_p-norm is the ℓ_q-norm, where q satisfies $1/p + 1/q = 1$

Example

Norm

$$f(x) = \|x\|$$

$$f^*(y) = \begin{cases} 0 & \|y\|_* \leq 1 \\ \infty & \text{otherwise,} \end{cases}$$

$$\|z\|_* = \sup\{z^T x \mid \|x\| \leq 1\}$$

the dual of the ℓ_p-norm is the ℓ_q-norm, where q satisfies $1/p + 1/q = 1$.

$\|y\|_* \leq 1 \longrightarrow$ for all x, $y^T x - \|x\| \leq 0 \longrightarrow$ maximum value 0

$\|y\|_* > 1 \longrightarrow$ $z \in \mathbf{R}^n$ with $\|z\| \leq 1$ and $y^T z > 1$. Taking $x = tz$ and letting $t \to \infty$, $\longrightarrow$

$$y^T x - \|x\| = t(y^T z - \|z\|) \to \infty,$$

Implicit constraints

LP with box constraints:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & -\mathbf{1} \preceq x \preceq \mathbf{1} \end{array}$$

$$\begin{array}{ll} \text{maximize} & -b^T \nu - \mathbf{1}^T \lambda_1 - \mathbf{1}^T \lambda_2 \\ \text{subject to} & c + A^T \nu + \lambda_1 - \lambda_2 = 0 \\ & \lambda_1 \succeq 0, \quad \lambda_2 \succeq 0 \end{array}$$

Implicit constraints

LP with box constraints:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & -\mathbf{1} \preceq x \preceq \mathbf{1} \end{array}$$

$$\begin{array}{ll} \text{maximize} & -b^T \nu - \mathbf{1}^T \lambda_1 - \mathbf{1}^T \lambda_2 \\ \text{subject to} & c + A^T \nu + \lambda_1 - \lambda_2 = 0 \\ & \lambda_1 \succeq 0, \quad \lambda_2 \succeq 0 \end{array}$$

reformulation with box constraints made implicit

$$\begin{array}{ll} \text{minimize} & f_0(x) = \begin{cases} c^T x & -\mathbf{1} \preceq x \preceq \mathbf{1} \\ \infty & \text{otherwise} \end{cases} \\ \text{subject to} & Ax = b \end{array}$$

Implicit constraints

LP with box constraints:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & -\mathbf{1} \preceq x \preceq \mathbf{1} \end{array}$$

$$\begin{array}{ll} \text{maximize} & -b^T \nu - \mathbf{1}^T \lambda_1 - \mathbf{1}^T \lambda_2 \\ \text{subject to} & c + A^T \nu + \lambda_1 - \lambda_2 = 0 \\ & \lambda_1 \succeq 0, \quad \lambda_2 \succeq 0 \end{array}$$

reformulation with box constraints made implicit

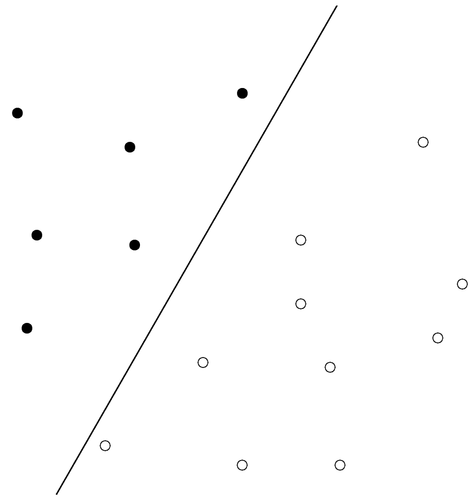
$$\begin{array}{ll} \text{minimize} & f_0(x) = \begin{cases} c^T x & -\mathbf{1} \preceq x \preceq \mathbf{1} \\ \infty & \text{otherwise} \end{cases} \\ \text{subject to} & Ax = b \end{array}$$

$$\begin{aligned} g(\nu) &= \inf_{-\mathbf{1} \preceq x \preceq \mathbf{1}} (c^T x + \nu^T (Ax - b)) \\ &= -b^T \nu - \|A^T \nu + c\|_1 \end{aligned}$$

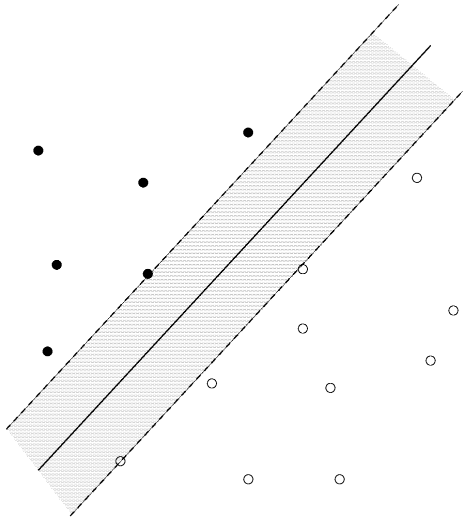
$$\text{maximize} \ -b^T \nu - \|A^T \nu + c\|_1$$

CLASSIFICATION

Support vector machines



$$g(x) = w^T x + w_0$$



$\{x_i, i = 1, \ldots, n\}$ two classes, ω_1 and ω_2.

Support vector machines

- OBJECTIVE: Maximize distance between two hyperplane $\frac{2}{||W||^2}$

$$\min \quad \frac{1}{2} w^T w$$

$$\begin{aligned} w^T x_i + w_0 &\geq +1 \\ w^T x_i + w_0 &\leq -1 \end{aligned} \qquad x \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

Support vector machines

- OBJECTIVE: Maximize distance between two hyperplane $\frac{2}{||W||^2}$

$$\min \quad \frac{1}{2} w^T w$$

$$\begin{aligned} w^T x_i + w_0 &\geq +1 \\ w^T x_i + w_0 &\leq -1 \end{aligned} \qquad x \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

ω_1 with corresponding numeric value, $y_i = +1$
ω_2 with corresponding numeric value, $y_i = -1$

$$\begin{aligned} w^T x_i + w_0 &\geq +1 \quad \text{for } y_i = +1 \\ w^T x_i + w_0 &\leq -1 \quad \text{for } y_i = -1 \end{aligned}$$

$$y_i(w^T x_i + w_0) \geq 1 \quad i = 1, \ldots, n$$

Support vector machines

$$\min \quad \frac{1}{2}w^T w$$

$$y_i(w^T x_i + w_0) \geq 1 \quad i = 1, \ldots, n$$

Lagrangian

$$L_p = \frac{1}{2}w^T w - \sum_{i=1}^{n} \alpha_i (y_i(w^T x_i + w_0) - 1)$$

Support vector machines

$$\min \quad \frac{1}{2} w^T w$$

$$y_i(w^T x_i + w_0) \geq 1 \quad i = 1, \ldots, n$$

Lagrangian

$$L_p = \frac{1}{2} w^T w - \sum_{i=1}^{n} \alpha_i (y_i (w^T x_i + w_0) - 1)$$

Differentiating L_p with respect to w_0 and ***w***

$$\sum_{i=1}^{n} \alpha_i y_i = 0$$

$$w = \sum_{i=1}^{n} \alpha_i y_i x_i$$

Support vector machines

Dual Function g

$$L_D = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

Dual problem

Max $$L_D = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$\alpha_i \geq 0 \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

Support vector machines

Dual Function g

$$L_D = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

Dual problem

Max

$$L_D = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$\alpha_i \geq 0 \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

inner products of patterns, ***xi***

nonlinear support vector machines

KKT conditions

4th condition

$$w = \sum_{i=1}^{n} \alpha_i y_i x_i$$

KKT conditions

4th condition

$$w = \sum_{i=1}^{n} \alpha_i y_i x_i$$

Complementary Slackness

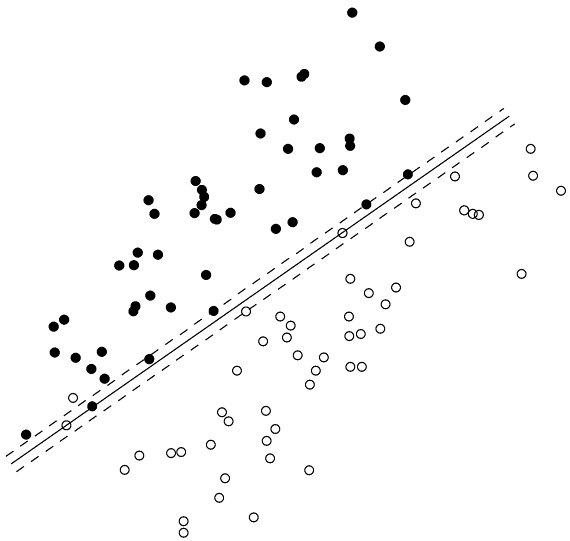
$$\alpha_i (y_i (x_i^T w + w_0) - 1) = 0$$

$\downarrow$ Nonzero dual optimal

$$y_i (x_i^T w + w_0) - 1) = 0$$

Support Vectors

Linearly nonseparable two-class data



- ✓two sets of points cannot be linearly separated
- ✓seek an affine function that approximately classifies the points
- ✓relax the constraints by introducing nonnegative variables

$$\begin{array}{ll} \boldsymbol{w}^T \boldsymbol{x}_i + w_0 \geq +1 - \xi_i & \text{for } y_i = +1 \\ \boldsymbol{w}^T \boldsymbol{x}_i + w_0 \leq -1 + \xi_i & \text{for } y_i = -1 \\ \xi_i \geq 0 & i = 1, \ldots, n \end{array}$$

Linearly nonseparable two-class data

'regularisation' parameter

$$\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_i \xi_i$$

$$\begin{aligned} &y_i(\boldsymbol{w}^T\boldsymbol{x}_i + w_0) \geq 1 - \xi_i && i = 1, \dots, n \\ &\xi_i \geq 0 && i = 1, \dots, n \end{aligned}$$

Linearly nonseparable two-class data

'regularisation' parameter

$$\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_i \xi_i$$

$$\begin{aligned} &y_i(\boldsymbol{w}^T\boldsymbol{x}_i + w_0) \geq 1 - \xi_i \quad && i = 1, \ldots, n \\ &\xi_i \geq 0 && i = 1, \ldots, n \end{aligned}$$

Lagrangian

$$L_p = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_i \xi_i - \sum_{i=1}^n \alpha_i(y_i(\boldsymbol{w}^T\boldsymbol{x}_i + w_0) - 1 + \xi_i) - \sum_{i=1}^n r_i\xi_i$$

Differentiating L_p with respect to w_0, *w and zeta*

$$\boldsymbol{w} = \sum_{i=1}^n \alpha_i y_i \boldsymbol{x}_i \qquad \sum_{i=1}^n \alpha_i y_i = 0 \qquad C - \alpha_i - r_i = 0$$

Linearly nonseparable two-class data

Dual Function g

$$L_D = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j$$

Dual problem

$$L_D = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \qquad 0 \leq \alpha_i \leq C$$