Support Vector machine

Professor Abdellatif El Afia

Support Vector machine

Duality Theory in the Convex Programming Problem

Professor Abdellatif El Afia

Plan

1. Convex programming
2. Derivation of the Dual problem
3. Weak Duality theorem
4. Strong Duality Theorem
5. Optimality Conditions

Convex Programming Problem

Definition: **(Convex programming problem)**

A Convex Programming Problem is an optimization problem in the form

$$(P)\begin{cases} Min & f_0(x) \\ s.t & f_i(x) \leq 0 \quad i \in I_1 \\ & h_i(x) = a_i^T x + b_i = 0 \quad i \in I_2 \\ & x \in \mathbb{R}^n \end{cases}$$

Where $f_0(x), \{f_i(x)\}_{i \in I_1}$ are continuous convex functions on $\mathbb{R}^d$ and $\{h_i(x)\}_{i \in I_2}$ are linear functions

Theorem:

- The feasible domain, D_F, is convex $D_F = \{x \in \mathbb{R}^d \,|\, f_i(x) \leq 0 \quad i \in I_1, h_i(x) = 0 \quad i \in I_2\}$
- The solution set of (P) are convex closed set
- If x^* is the local solution of (P) then x^* is also the global solution
- If the objective function $f_0(x)$ is strictly convex then the solution of (P) is unique

Convex Programming Problem

Definition: **(Quadratic Programming (QP) problem)**

$$(QP)\begin{cases} Min & \frac{1}{2}x^TQx + c^Tx \\ s.t & \bar{A}x - \bar{b} \leq 0 \\ & Ax - b = 0 \\ & x \in \mathbb{R}^n \end{cases}$$

Where $Q \in \mathbb{R}^{n\times n}$ c $\in \mathbb{R}^n$, $\bar{A} \in \mathbb{R}^{m\times n}$, $A \in \mathbb{R}^{p\times n}$, $\bar{b} \in \mathbb{R}^m$, b $\in \mathbb{R}^p$

If Q is possitive semidefinite, then the QP-Problem is a convex programming

Theorem:

- The set $D_F = \left\{x \in \mathbb{R}^d \middle| \bar{A}x - \bar{b} \leq 0, Ax - b = 0\right\}$ is convex
- The solution set of (QP) is convex closed set
- If $x^*$is the local solution of (QP) then $x^*$is also the global solution
- If $\boldsymbol{Q}$ is possitive definite then The solution of (QP) is unique

Convex Programming Problem

Definition:

Consider the Convex programming problem (P) with variable x being partitioned into the form $x = (x_1, x_2) \in \mathbb{R}^n$.

$x_1^* \in \mathbb{R}^{m_1}$ is called its solution with respect to (w.r.t) x_1 if there exists a $x_2^* \in \mathbb{R}^{n-m_1}$ such that $x^* = (x_1^*, x_2^*)$ is its solution. The set of all solutions w.r.t x_1 are called the solution set w.r.t x_1

Theorem:

- If the Convex programming problem (P) with variable $x = (x_1, x_2) \in \mathbb{R}^n$, then
  - its solution set w.r.t x_1 is convex closed set
  - If $f_0(x) = F_1(x_1) + F_2(x_2)$ where is strictly convex of variable x_1 then the solution to the (P) w.r.t x_1 is unique when it has a solution

Derivation of the Dual problem

Consider the Convex programming problem

$$Primal\ \ (P)\begin{cases} Min & f_0(x) \\ s.t & f_i(x) \le 0 \quad i \in I_1 \\ & h_i(x) = a_i^T x + b_i = 0 \quad i \in I_2 \\ & x \in \mathbb{R}^n \end{cases}$$

Where $f_0(x) \in \mathcal{C}^2, \forall i \in I_1\ f_i(x) \in \mathcal{C}^2$ and are convex in $\mathbb{R}^d$,

We start from estimating its optimal value

$$p^* = inf\{f_0(x)|x \in D_F\} = \inf_{x \in D_F} f_0(x)$$

where $D_F = \{x \in \mathbb{R}^d \big| f_i(x) \le 0 \quad i \in I_1, h_i(x) = 0 \quad i \in I_2\}$

Introduce the Lagrangian function

$$L(x, \lambda, \mu) = f_0(x) + \sum_{i \in I_1} \lambda_i f_i(x) + \sum_{i \in I_2} \mu_i h_i(x)$$

where $\lambda = (\lambda_1, \dots, \lambda_m)^T$ and $\mu = \left(\mu_1, \dots, \mu_p\right)^T$ are lagrangian mutiplers.

Derivation of the Dual problem

Obviously, when $x \in D_F$, $\lambda \geq 0$, we have $L(x,\lambda,\mu) \leq f_0(x)$ thus

$$\inf_{x\in\mathbb{R}^n} L(x,\lambda,\mu) \leq \inf_{x\in D_F} L(x,\lambda,\mu) \leq \inf_{x\in D_F} f_0(x)$$

Therefore, introducing the langrangian dual function $g(\lambda,\mu) = \inf_{x\in\mathbb{R}^n} L(x,\lambda,\mu)$ yields

$$g(\lambda,\mu) = \inf_{x\in\mathbb{R}^n} L(x,\lambda,\mu) \leq \inf_{x\in D_F} f_0(x) = p^* \Longrightarrow g(\lambda,\mu) \leq p^*$$

The above inequality indicates that, for any $\lambda \geq 0$, $g(\lambda,\mu)$ is a lower bound of p^*.
Among these lower bounds, finding the best one lead to the optimization problem called the dual problem of the primal problem (P)

$$\boldsymbol{Dual}: (\boldsymbol{D}) \begin{cases} \boldsymbol{Max} & \boldsymbol{g(\lambda,\mu) = \inf_{x\in\mathbb{R}^n} L(x,\lambda,\mu)} \\ \boldsymbol{s.t} & \boldsymbol{\lambda \geq 0} \end{cases}$$

The optimal value of the dual Problem (D), which we denote $d^* = sup\{g(\lambda,\mu)|\lambda \geq 0\}$

Theorem: Dual problem is convex programming problem

Derivation of the Dual problem $x \in D_F, \lambda \geq 0$

- $L(x,\lambda,\mu) = f_0(x) + \sum_{i\in I_1} \lambda_i f_i(x) + \sum_{i\in I_2} \mu_i h_i(x)$
- $L(x,\lambda,\mu) \leq f_0(x) \rightarrow \inf\limits_{x\in D_F} L(x,\lambda,\mu) \leq \inf\limits_{x\in D_F} f_0(x)$
- $\inf\limits_{x\in \mathbb{R}^n} L(x,\lambda,\mu) \leq \inf\limits_{x\in D_F} L(x,\lambda,\mu)$

- $\inf\limits_{x\in \mathbb{R}^n} L(x,\lambda,\mu) \leq \inf\limits_{x\in D_F} L(x,\lambda,\mu) \leq \inf\limits_{x\in D_F} f_0(x)$
- $g(\lambda,\mu) = \inf\limits_{x\in \mathbb{R}^n} L(x,\lambda,\mu) \leq \inf\limits_{x\in D_F} L(x,\lambda,\mu) \leq \inf\limits_{x\in D_F} f_0(x)$
- $d^* = sup\{g(\lambda,\mu) | \lambda \geq 0\}$

Derivation of the Dual problem

- $Primal \ \ (P) \begin{cases} Min & f_0(x) \\ s.t & f_i(x) \leq 0 \ \ i \in I_1 \\ & h_i(x) = a_i^T x + b_i = 0 \ \ i \in I_2 \\ & x \in \mathbb{R}^n \end{cases}$
- $p^* = inf\{f_0(x) | x \in D_F\} = \inf_{x \in D_F} f_0(x)$
- $L(x, \lambda, \mu) = f_0(x) + \sum_{i \in I_1} \lambda_i f_i(x) + \sum_{i \in I_2} \mu_i h_i(x)$
- $Dual: (D) \begin{cases} Max & g(\lambda, \mu) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu) \\ s.t & \lambda \geq 0 \end{cases}$
- $d^* = sup\{g(\lambda, \mu) | \lambda \geq 0\}$

Weak Duality Theorem(WDT)

Theorem(**WDT**):

Let p^*be optimal value of the primal problem(P) and d^*be optimal value of the dual Problem (D) then

$$(WDT)\ \ p^* = inf\{f_0(x)|x \in D_F\} \geq d^* = sup\{g(\lambda,\mu)|\lambda \geq 0\}$$

Note that the inequality still holds when $p^*$and $d^*$are inifinite. For example:

- If the primal problem is unbounded below, so that $p^* = -\infty$, we must have $d^* = -\infty$: the dual problem is infeasible
- If the dual problem is unbounded above, so that $d^* = +\infty$, we must have $p^* = +\infty$: the primal problem is infeasible

Corollary:

Let x be the feasible solution of the primal problem (P) and (λ,μ) be the feasible of the dual problem (D). If $f_0(x) = g(\lambda,\mu)$ then x and $(\lambda,\mu)$are their solutions respectively.

Strong Duality Theorem(SDT)

Strong Duality Theorem concerns the case where the inequality in (WDT) holds with equality.

Definition (**Slater's Condition**)
Convex Programming Primal Problem$(P)$is said to satisfy Slater's Condition if there exists a feasible solution x such that:

$$\begin{cases} f_i(x) < 0 \quad i \in I_1 \\ h_i(x) = a_i^T x + b_i = 0 \quad i \in I_2 \end{cases}$$

Or when the first k inequality constraints are linear constraints, there exists a feasible solution x such that:

$$\begin{cases} f_i(x) = a_i^T x + b_i \leq 0 \quad i \in I_1^k \\ f_i(x) < 0 \quad i \in I_1^{m-k} \\ h_i(x) = a_i^T x + b_i = 0 \quad i \in I_2 \end{cases}$$

Strong Duality Theorem (SDT)

Strong Duality Theorem concerns the case where the inequality in (WDT) holds with equality.

Theorem (SDT)
Consider the Convex Programming Primal Problem(P) satisfing Slater's Condition. Let p^*be optimal value of the primal problem(P) and d^*be optimal value of the dual Problem (D). Then

$$(SDT)\ \ p^* = inf\{f_0(x)|x \in D_F\} = d^* = sup\{g(\lambda,\mu)|\lambda \geq 0\}$$

Furthermore,
if $p^*$is attained, that means there exists a solution $x^*$to the primal problem (P), then $d^*$is attained, that means there exists a solution$(\lambda^*,\mu^*)$to the Dual problem (D)such that

$$(SDT)\ \ \ p^* = f_0(x^*) = d^* = g(\lambda^*,\mu^*) < \infty$$

Optimality Conditions

Definition(Karush-Kuhn-Tuker (KKT) Conditions)
Consider the convex programming Primal Problem (P). $x^*$is said to satisfy the KKT conditions if there exist the multipliers $\lambda^* = (\lambda_1^*, \dots, \lambda_m^*)^T$ and $\mu^* = \left(\mu_1^*, \dots, \mu_p^*\right)^T$ coresponding to constraints of Primal Problem (P) respectively, such that the Lagrangian function

$$L(x, \lambda, \mu) = f_0(x) + \sum_{i \in I_1} \lambda_i f_i(x) + \sum_{i \in I_2} \mu_i h_i(x) \Longrightarrow \nabla_x L(x, \lambda, \mu) = \nabla f_0(x) + \sum_{i \in I_1} \lambda_i \nabla f_i(x) + \sum_{i \in I_2} \mu_i \nabla h_i(x)$$

Satisfies

$$\textbf{KKT Conditions:} \begin{cases} \nabla_x L(x^*, \lambda^*, \mu^*) = \nabla f_0(x^*) + \sum_{i \in I_1} \lambda_i^* \nabla f_i(x^*) + \sum_{i \in I_2} \mu_i^* \nabla h_i(x^*) = 0 \\ f_i(x^*) \leq 0 \quad i \in I_1 : |I_1| = m \\ h_i(x^*) = a_i^T x^* + b_i = 0 \quad i \in I_2 : \ |I_2| = p \\ \lambda_i^* f_i(x^*) = 0 \qquad i \in I_1 \\ \lambda_i^* \geq 0 \qquad i \in I_1 \end{cases}$$

Optimality Conditions

Theorem:
Consider the convex programming Primal Problem (P) satisfing Slater's Condition. If $x^*$is its solutions then x^*satisfies the KKT conditions

Theorem:
Consider the convex programming Primal Problem (P) satisfing Slater's Condition. Then for its solution x^*, it is necessary and sufficient that x^*satisfies the KKT conditions

Exercises

- $L(x,\lambda,\mu) = f_0(x) + \sum_{i\in I_1} \lambda_i f_i(x) + \sum_{i\in I_2} \mu_i h_i(x)$
- $\nabla_x L(x,\lambda,\mu) = 0$
- $g(\lambda,\mu) = \inf\limits_{x\in\mathbb{R}^n} L(x,\lambda,\mu) \Longrightarrow Dual: (D) \begin{cases} Max & g(\lambda,\mu) = \inf\limits_{x\in\mathbb{R}^n} L(x,\lambda,\mu) \\ s.t & \lambda \geq 0 \end{cases}$

Find the dual of the Following problems

$$(P_1)\begin{cases} Min & c^T x \\ S.t & Ax = b \\ & x \in \mathbb{R}^n \end{cases} \quad (P_2)\begin{cases} Min & c^T x \\ S.t & Ax \leq b \\ & x \in \mathbb{R}^n \end{cases} \quad (P_3)\begin{cases} Min & c^T x \\ S.t & Ax \leq b \\ & x \geq 0 \end{cases}$$

$$(P_4)\begin{cases} Max & c^T x \\ S.t & Ax = b \\ & x \in \mathbb{R}^n \end{cases} \quad (P_5)\begin{cases} Max & c^T x \\ S.t & Ax \leq b \\ & x \in \mathbb{R}^n \end{cases} \quad (P_6)\begin{cases} Max & c^T x \\ S.t & Ax \leq b \\ & x \geq 0 \end{cases}$$

Exercises

Find the dual of the Following problems

- $SVC: \begin{cases} Min & \frac{1}{2}\|w\|^2 \\ s.t & y_i(w^T x_i + b) \geq 1 \;\; i = 1,..,n \\ & w \in \mathbb{R}^d, b \in \mathbb{R} \end{cases}$ $\quad C - SVC \begin{cases} Min & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i \\ s.t & y_i(w^T x_i + b) \geq 1 - \xi_i, i = 1,..,n \\ & \xi_i \geq 0, w \in \mathbb{R}^d, b \in \mathbb{R} \end{cases}$

- $\text{L}\varepsilon SVR: \begin{cases} & min \quad \frac{1}{2}w^T w + C\sum_{i=1}^{n}(\xi_i + \eta_i) \\ & y_i - (w^T x_i - b) \leq \varepsilon + \xi_i, \quad i = 1, \dots, n \\ s.à & w^T x_i + b - y_i \quad \leq \varepsilon + \eta_i, \quad i = 1, \dots, n \\ & \xi_i \geq 0, \eta_i \geq 0, (w, b) \in \mathbb{R}^d \times \mathbb{R} \end{cases}$ $\quad \text{L}\varepsilon SVR: \begin{cases} & min \quad \frac{1}{2}w^T w \\ & y_i - (w^T x_i + b) \leq \varepsilon, \quad i = 1, \dots, n \\ s.à & w^T x_i + b - y_i \quad \leq \varepsilon, \quad i = 1, \dots, n \\ & (w, b) \in \mathbb{R}^d \times \mathbb{R} \end{cases}$