Optimisation Non linèaire

Par

Professeur Abdellatif El Afia

Chapitre 7

Dualité lagrangienne

Professor Abdellatif El Afia

Dualité lagrangienne

$$Primal \begin{cases} Min & f(x) \\ s.t & f_i(x) \leq 0 \; i = 1, \dots, m \\ & x \in X \subseteq \mathbb{R}^n \end{cases}$$

- $D_R = \{x \in X \subseteq \mathbb{R}^n : f_i(x) \leq \; 0 \;\; i = 1, \dots, m\}$

Définition

Le problème dual lagrangien associé au problème primal relativement au D_R, s'écrit :

$$Dual \begin{cases} Max & g(\lambda) = \inf_{x \in \mathbb{R}^n} \left\{ f(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right\} \\ s.t & \lambda \in \Lambda \end{cases}$$

- $g(\lambda) = \inf_{x \in \boldsymbol{X}} \{f(x) + \sum_{i=1}^{m} \lambda_i f_i(x)\}$ est la fonction objective du dual
- $\Lambda = \{\lambda \in \mathbb{R}^m : \lambda \geq 0\}$est l'ensemble des contraintes du dual problème

Dualité lagrangienne: Exemple

Soit le problème quadratique convexe suivant :

$$\begin{cases} Min & \frac{1}{2}x^T Q x + c^T x \\ s.t & Ax \geq b \\ & x \in X \subseteq \mathbb{R}^n \end{cases}$$

Où Q est symétrique et définie positive.

Pour construire le problème dual, on commence par réécrire les contraintes d'inégalité comme suit $b - Ax \leq 0$, en suite nous construisons le Lagrangien :

$$L(x,\lambda) = \frac{1}{2}x^T Q x + c^T x + \lambda^T(b - Ax) = \lambda^T b + (c^T - A^T\lambda)^T x + \frac{1}{2}x^T Q x$$

La fonction duale est :

$$g(\lambda) = \min_{x\in\mathbb{R}^n} L(x,\lambda) = \min_{x\in\mathbb{R}^n} \lambda^T b + (c^T - A^T\lambda)^T x + \frac{1}{2}x^T Q x$$

$$= \lambda^T b + \min_{x\in\mathbb{R}^n} (c^T - A^T\lambda)^T x + \frac{1}{2}x^T Q x$$

Dualité lagrangienne: Exemple

Pour une valeur de λ quelconque, soit x^* la valeur optimale de $\min_{x \in \mathbb{R}^n} (c^T - A^T\lambda)^T x + \frac{1}{2} x^T Q x$

Comme cette expression est une fonction quadratique convexe le gradient du lagrangien en x^* est nul :

$$\nabla_x L(x^*, \lambda) = (c - A^T\lambda) + Qx^* = 0$$

L'expression de x^* est alors donnée par : $x^* = -Q^{-1}(c - A^T\lambda)$

La fonction duale devient alors : $g(\lambda) = L(x^*, \lambda) = \lambda^T b - \frac{1}{2}(c - A^T\lambda)^T Q^{-1}(c - A^T\lambda)$

Comme le problème dual est définit par : $\max_{\lambda \geq 0} g(\lambda)$

Le dual du problème quadratique convexe est donc donné par :

$$\begin{cases} Max \quad \lambda^T b - \dfrac{1}{2}(c - A^T\lambda)^T Q^{-1}(c - A^T\lambda) \\ s.t \qquad \lambda \geq 0 \\ \qquad\quad \lambda \in \mathbb{R}^m \end{cases}$$

Dualité lagrangienne

Théorème

La fonction objective du dual $g(\lambda)$ est une fonction concave sur l'ensemble Λ pour toutes fonctions f et f_i

Preuve

Soient $\lambda_1 \in \Lambda$, $\lambda_2 \in \Lambda$ alors $\lambda_1 \geq 0$ et $\lambda_2 \geq 0$, et soit $\lambda = \theta\lambda_1 + (1-\theta)\lambda_2 \in \Lambda$ avec $\theta \in [0,1]$ alors

- $g(\lambda) = \inf\limits_{x\in\mathbb{R}^n}\{f(x) + \sum_{i=1}^{m} \lambda_i f_i(x)\}$
- $g(\lambda) = \inf\limits_{x\in\mathbb{R}^n}\{\theta\big(f(x) + \sum_{i=1}^{m} \lambda_{1_i} f_i(x)\big) + (1-\theta)\big(f(x) + \sum_{i=1}^{m} \lambda_{2_i} f_i(x)\big)\}$
- $\Longrightarrow g(\lambda) \geq \theta \inf\limits_{x\in\mathbb{R}^n}\{f(x) + \sum_{i=1}^{m} \lambda_{1_i} f_i(x)\} + (1-\theta) \inf\limits_{x\in\mathbb{R}^n}\{f(x) + \sum_{i=1}^{m} \lambda_{2_i} f_i(x)\}$
- $\Longrightarrow g(\lambda) \geq \theta g(\lambda_1) + (1-\theta) g(\lambda_2)$

D'où $g(\lambda)$ est une fonction concave sur l'ensemble Λ pour toutes fonctions f et f_i

Dualité lagrangienne

Théorème:

Soient $A \subset \mathbb{R}^n$ et $B \subset \mathbb{R}^m$.

Si

- A est convexe
- Pour toute valeur $b \in B, h(a, b): A \times B \to \mathbb{R}$ est convexe sur A,

alors

- l'ensemble $A^* = \{a \in A: l(a) = Sup_{b \in B}\{h(a, b)\} < \infty\}$ est convexe
- l est convexe sur A^*

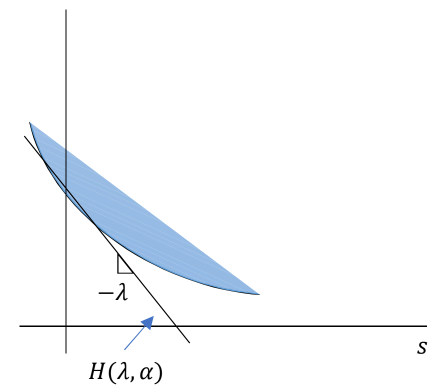
Définition

- La **valeur optimale** du problème **primal** est le infinimum de la fonction $f(x)$ sur D_R
- La **valeur optimale** du problème **dual** est le supremum de la fonction $g(\lambda)$ sur Λ
- Par convention, le infinimum (supremum) d'une fonction sur un ensemble vide est égal à ∞ $(-\infty)$.

Interprétation géométrique du problème dual

Soit

- $x \in \mathbb{R}^n$, Posons $(s, z) = (s_1, \dots, s_m, z) = (f_1(x), \dots, f_m(x), z)$
- $I = \{(s, z) \in \mathbb{R}^{m+1} : \exists x \in \mathbb{R}^n où\ z \geq f(x)\ et\ s \geq \big(f_i(x)\big), i = 1, \dots, m\}$



Cet ensemble est illustré par la figure

Proposition:

Si X est convexe et f et $f_i, i = 1, \dots, m$ sont convexes sur X alors I est aussi convexe.

Maintenant, on définit l'hyperplan

$$H_{\lambda,\alpha} \coloneqq \{(s, z) \in \mathbb{R}^{m+1}: \lambda^T s + z = \alpha\}$$

On dit que $H_{\lambda,\alpha}$ est *support inférieur* de I si

$$\forall (s, z) \in I\ \lambda^T s + z \geq \alpha$$

On définit aussi la droite

Théorèmes de dualité

Pour le premier théorème de dualité faible, nous n'avons pas besoin d'hypothèses particulières sur X, ni sur les fonctions f et $f_i, i = 1, \dots, m$.

Théorème de dualité faible

Si $\bar{x}$ est une solution réalisable du primal et $\bar{\lambda}$ une solution réalisable du dual, alors $f(\bar{x}) \geq g(\bar{\lambda})$.

Preuve

Considérons la quantité $f(\bar{x}) + \sum_{i=1}^{m} \bar{\lambda}_i\, f_i(\bar{x}) \leq f(\bar{x})$, Puisque $f_i(\bar{x}) \leq 0$ et $\bar{\lambda}_i \geq 0, i = 1, \dots, m$, D'autre part,

$$g(\bar{\lambda}) = \inf_{x \in X} \left\{ f(x) + \sum_{i=1}^{m} \bar{\lambda}_i f_i(x) \right\} \leq f(\bar{x}) + \sum_{i=1}^{m} \bar{\lambda}_i\, f(\bar{x})$$

Puisque $\bar{x} \in X$

Donc $g(\bar{\lambda}) = \inf_{x \in X} \{f(x) + \sum_{i=1}^{m} \bar{\lambda}_i\, f_i(x)\} \leq f(\bar{x}) + \sum_{i=1}^{m} \bar{\lambda}_i f_i(\bar{x}) \leq f(\bar{x})$.

Théorèmes de dualité

Définition: Point selle

(x^*, λ^*) est dit point selle du Lagrangien $L(x, \lambda)$ si

- $x^* \in X \; et \; \lambda^* \geq 0$
- $L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*) \quad \forall x \in X \,, \forall \lambda \geq 0$

Théorème 1:

(x^*, λ^*) est un point selle du Lagrangien si et seulement si les trois conditions suivantes sont vérifiées :

$$\begin{cases} \boldsymbol{C_1}: & L(x^*, \lambda^*) = g(\lambda^*) \\ \boldsymbol{C_2}: & F(x^*) \leq 0 \\ \boldsymbol{C_3}: & {\lambda^*}^T F(x^*) = 0 \end{cases}$$

En plus, (x^*, λ^*) est point selle pour $L(x, \lambda)$ si et seulement si les deux conditions suivantes sont vérifiées :

i. (x^*, λ^*) est respectivement, solution optimale du problème primal et dual

ii. Il n'y a pas de saut de dualité, c'est-à-dire que $f(x^*) = g(\lambda^*)$

Théorèmes de dualité

Théorème 2 :

Supposons $X = \mathbb{R}^n$ et $f(x), f_1(x), \dots, f_m(x))$ sont convexes et différentiables, et (x^*, λ^*) satisfait les conditions de **K.K.T** alors $x^*\ et\ \lambda^*$ sont solution du problème primal et dual sans saut de dualité.

(x^*, λ^*) satisfait les conditions de **K.K.T** :

$$\begin{cases} \nabla f(x^*) + {\lambda^*}^T \nabla F(x^*) = 0 \\ F(x^*) \leq 0 \\ \lambda^* \geq 0 \\ {\lambda^*}^T F(x^*) = 0 \end{cases}$$

Théorème 3 :

Si (x^*, λ^*) est un point selle de L alors

$$g^* = \sup_{\lambda \geq 0} \inf_{x \in X} L(x, \lambda) = L(x^*, \lambda^*) = \inf_{x \in \mathrm{X}} \sup_{\lambda \geq 0} L(x, \lambda) = f^*$$

Problème de programmation convexe

Définition:

Un problème de programmation convexe est un problème d'optimisation sous la forme

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

Où $f_0(x), \{f_i(x)\}_{i \in I_1}$ sont des fonctions convexes continues sur $\mathbb{R}^d$et $\{h_i(x)\}_{i \in I_2}$ sont des fonctions linéaires

Théorème:

- Le domaine réalisable, D_F, est convexe $D_F = \{x \in \mathbb{R}^d \middle| f_i(x) \leq 0 \quad i \in I_1, h_i(x) = 0 \;\; i \in I_2\}$
- L'ensemble solution de (P) est un ensemble fermé convexe
- Si x^* est la solution locale de (P) alors x^* est aussi la solution globale
- Si la fonction objectif $f_0(x)$ est strictement convexe alors la solution de (P) est unique

Problème de programmation convexe

Definition: **(Problème de programmation quadratique (QP))**

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

Où $Q \in \mathbb{R}^{n\times n}$ 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 $\in \mathbb{R}^p$

Si $\boldsymbol{Q}$ est semi-défini positif, alors le problème $\boldsymbol{QP}$ est une programmation convexe

Théorème:

- L'ensemble $D_F = \left\{x \in \mathbb{R}^d \middle| \bar{A}x - \bar{b} \leq 0, Ax - b = 0\right\}$ est convexe
- L'ensemble solution de (QP) est un ensemble fermé convexe
- Si x^* est la solution locale de(QP)alors x^* est aussi la solution globale
- Si $\boldsymbol{Q}$ est définie positive alors La solution de (QP) est unique

Problème de programmation convexe

Définition:

Considérons le problème de programmation convexe (P) avec la variable x étant partitionné sous la forme $x = (x_1, x_2) \in \mathbb{R}^n$.
$x_1^* \in \mathbb{R}^{m_1}$ est appelé sa solution par rapport à (w.r.t) x_1 s'il existe un $x_2^* \in \mathbb{R}^{n-m_1}$ tel que $x^* = (x_1^*, x_2^*)$ est sa solution. L'ensemble de toutes les solutions par rapport à x_1 est appelé l'ensemble de solutions par rapport à x_1

Théorème:

Si le problème de programmation convexe (P) avec la variable $x = (x_1, x_2) \in \mathbb{R}^n$, alors

- son ensemble solution par rapport à x_1 est un ensemble fermé convexe
- Si $f_0(x) = F_1(x_1) + F_2(x_2)$ où est strictement convexe de la variable x_1 alors la solution de (P) par rapport à x_1 est unique lorsqu'elle a une solution

Problème de programmation convexe

Considérons le problème de programmation convexe

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

où $f_0(x) \in \mathcal{C}^2, \forall i \in I_1\ f_i(x) \in \mathcal{C}^2$et convexe dans $\mathbb{R}^d$,
On part de l'estimation de sa valeur optimale

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

où $D_F = \{x \in \mathbb{R}^d \big| f_i(x) \leq 0 \quad i \in I_1, h_i(x) = 0 \ \ i \in I_2\}$
Introduisons la fonction lagrangienne

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

où $\lambda = (\lambda_1, \dots, \lambda_m)^T$et $\mu = \left(\mu_1, \dots, \mu_p\right)^T$ sont des multiplers lagrangiens.

Problème de programmation convexe: : Conditions d'optimalité

Définition**(Conditions de Karush-Kuhn-Tuker (KKT))**

Considérons le problème principal de programmation convexe (P). x^* est dit satisfaire les conditions KKT s'il existe les multiplicateurs $\lambda^* = (\lambda_1^*, \dots, \lambda_m^*)^T$ et $\mu^* = \left(\mu_1^*, \dots, \mu_p^*\right)^T$ correspondant respectivement aux contraintes du problème primal (P) , telles que la fonction lagrangienne (P)

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

Satisfait

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

Problème de programmation convexe: Conditions d'optimalité

Définition (**Condition de Slater**)
On dit que le problème primal de programmation convexe (P) satisfait la condition de Slater s'il existe une solution réalisable x telle que:

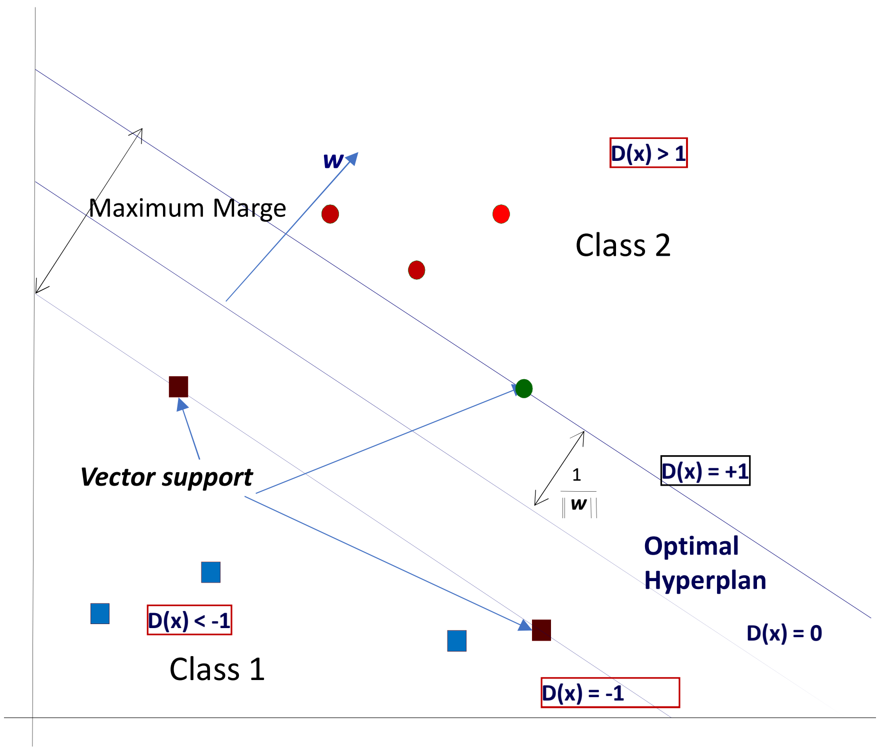
$$\begin{cases} f_i(x) < 0 \quad i \in I_1 \\ h_i(x) = a_i^T x + b_i = 0 \quad i \in I_2 \end{cases}$$

Ou lorsque les premières contraintes d'inégalité de k sont des contraintes linéaires, il existe une solution réalisable x telle que :

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

Théorème:
Considérons le problème primal de programmation convexe (P) satisfaisant la condition de Slater. Si x^* est une solution alors x^* satisfait aux conditions KKT.
Théorème:
Considérons le problème primal de programmation convexe (P) satisfaisant la condition de Slater. Alors pour sa solution x^*, c'est une condition néssecaire et suffissante que x^*satisfait aux conditions KKT

Hard margin SVC



- Distance from one point to the Separator:

$$D(x) = \frac{|w^T x + b|}{\|w\|}$$

Maximum margin before reaching the boundaries of both classes ($\boldsymbol{|w^T x + b| = 1}$) :

$$m = \frac{1}{\|w\|}$$

- To maximize m is to minimize $\|w\|$ while preserving the classification power:

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

Hard margin SVC

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

Theorem:

For a linearly seprable problem, there exists a solution unique (w^*, b^*) to optimization problem SVC and the solution satisfies:

- $w^* \neq 0$
- $\exists\, j \in \{i \in \{1, \dots, n\} | y_i = 1\}$ such that $(w^*)^T x_j + b^* = 1$
- $\exists\, k \in \{i \in \{1, \dots, n\} | y_i = -1\}$ such that $(w^*)^T x_k + b^* = -1$

Hard margin SVC: Dual form

The SVC approach uses Lagrange multipliers for a simpler solution

$$L_H(w,b,\lambda) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \lambda_i (y_i(w^T x_i + b) - 1)$$

According to chapter 1, the dual problem should have a form of

$$Dual\ SVC \begin{cases} Max \quad g(\lambda) = \inf\limits_{w\in\mathbb{R}^d, b\in\mathbb{R}} L_H(w,b,\lambda) \\ s.t \qquad\qquad \lambda_i \geq 0 \ \ i = 1,..,n \end{cases}$$

As $L_H(w,b,\lambda)$is strictly convex quadratic function of w, its minimal value is achieved at w satisfying $\nabla_{w,b} L_H(w,b,\lambda) = 0$, then

- $\nabla_w L_H(w,b,\lambda) = w - \sum_{i=1}^{n} \lambda_i y_i x_i = 0 \Longrightarrow w = \sum_{i=1}^{n} \lambda_i y_i x_i$
- $\nabla_b L_H(w,b,\lambda) = \sum_{i=1}^{n} \lambda_i y_i = 0$

Hard margin SVC: Dual form

- $L_H(w,b,\lambda) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n}\lambda_i(y_i(w^T x_i + b) - 1)$

- $w = \sum_{i=1}^{n}\lambda_i y_i x_i \Longrightarrow$

$$\boldsymbol{L_H(w,b,\lambda) = \sum_{i=1}^{n}\lambda_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_j\lambda_i y_j y_i\left(x_j{}^T x_i\right) - b\sum_{i=1}^{n}\lambda_i y_i}$$

One has by substitution in $L_H(w,b,\lambda)$:

$$\inf_{w\in\mathbb{R}^d, b\in\mathbb{R}} L_H(w,b,\lambda) = \begin{cases} \sum_{i=1}^{n}\lambda_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_j\lambda_i y_j y_i\left(x_j{}^T x_i\right) \; if \sum_{i=1}^{n}\lambda_i y_i = 0 \\ -\infty \;\; if \sum_{i=1}^{n}\lambda_i y_i \neq 0 \end{cases}$$

Hard margin SVC: Dual form

$$D-SVC:\begin{cases} Max \quad g(\lambda)=\sum_{i=1}^{n}\lambda_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_j\lambda_i y_j y_i\left({x_j}^T x_i\right) \\ s.t \quad \sum_{i=1}^{n}\lambda_i y_i=0 \\ \lambda_i\geq 0 \; i=1,\dots,n \end{cases}$$

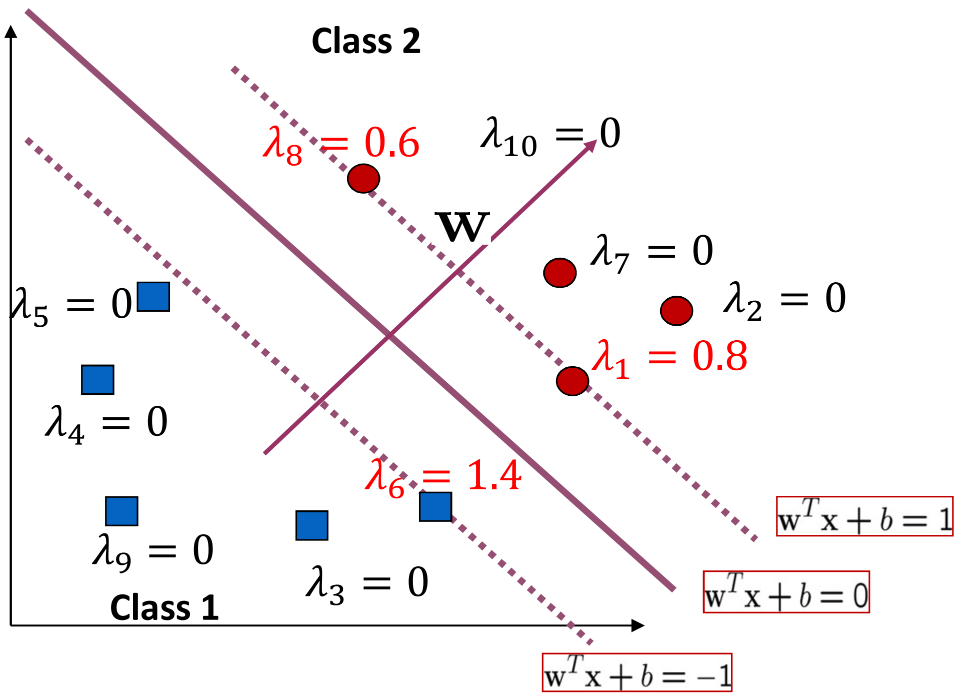
Theorem:

For separable problems,

- The $DC-SVC$ poblem is a Convex Quadratic Programming and has a solution λ^*
- For any solution $\lambda^*=(\lambda_1^*,\dots,\lambda_n^*)$, there must be a nonzero component $\lambda_j^*$and the unique solution to the primal SVC can be obtained in the following way

$$w^*=\sum_{i=1}^{n}\lambda_i^*\, y_i x_i \qquad and \;\; b^*=y_j-\sum_{i=1}^{n}\lambda_i^* y_i\left({x_j}^T x_i\right)$$

Hard margin SVC: Dual form



Geometric interpretation

- Only the points closest to the separation surface affect its definition

- There are theoretical limits for the misclassification of new data
  - The larger the margin, the smaller the limit
  - The smaller the number of SVC, the smaller the limit

Soft margin SVC($C - SVC$)

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

Theorem:

- There exists solutions to the $C - SVC$ problem w.r.t (w, b)
- The solution $w^*$of the $C - SVC$ problem (with respect to)w.r.t w is unique
- The solution set to the $C - SVC$ problem w.r.t b is a bounded close interval $[b_1, b_2]$ where $b_1 \leq b_2$.

Soft-margin $C - SVC$: Dual form

The C-SVC approach uses Lagrange multipliers for a simpler solution

$$L_{Soft}(w,\xi,b,\lambda,\mu) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\lambda_i(y_i(w^Tx_i+b)-1+\xi_i) - \sum_{i=1}^{n}\mu_i\xi_i$$

$$= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n}\lambda_i(y_i(w^Tx_i+b)-1)) - \sum_{i=1}^{n}(-C+\lambda_i+\mu_i)\xi_i$$

According to chapter 1, the dual problem should have a form of

$$Dual\ C - SVC\begin{cases} Max \quad g(\lambda,\mu) = \inf\limits_{w\in\mathbb{R}^d,\xi\in\mathbb{R}^n,b\in\mathbb{R}} L_{Soft}(w,\xi,b,\lambda,\mu) \\ s.t \qquad \lambda_i \geq 0, \mu_i \geq 0 \quad i = 1,..,n \end{cases}$$

As $L_{Soft}(w,\xi,b,\lambda,\mu)$is strictly convex quadratic function of w, its minimal value is achieved at w satisfying

$$\nabla_{w,\xi,b}L_{soft}(w,\xi,b,\lambda,\mu) = 0$$

Soft-margin $C - SVC$: Dual form

Then

- $\nabla_w L_{Soft}(w, \xi, b, \lambda, \mu) = w - \sum_{i=1}^{n} \lambda_i y_i x_i = 0 \Longrightarrow \boldsymbol{w = \sum_{i=1}^{n} \lambda_i y_i x_i}$
- $\nabla_\xi L_{Soft}(w, \xi, b, \lambda, \mu) = C I_{n\times n} - \lambda - \mu = 0$
- $\nabla_b L_{Soft}(w, \xi, b, \lambda, \mu) = \sum_{i=1}^{n} \lambda_i y_i = 0$

One has by substitution in $L_{Soft}(w, \xi, b, \lambda, \mu)$:

$$L_{Soft}(w, \xi, b, \lambda, \mu) = \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_j \lambda_i y_j y_i \left({x_j}^T x_i \right) - b \sum_{i=1}^{n} \lambda_i y_i - \sum_{i=1}^{n} (-C + \lambda_i + \mu_i) \xi_i$$

If $\sum_{i=1}^{n} \lambda_i y_i = 0$ and $C I_{n\times n} - \lambda - \mu = 0$ then

- $\inf_{w \in \mathbb{R}^d, \xi \in \mathbb{R}^n, b \in \mathbb{R}} L_{Soft}(w, \xi, b, \lambda, \mu) = \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_j \lambda_i y_j y_i \left({x_j}^T x_i \right)$

Else

- $\inf_{w \in \mathbb{R}^d, \xi \in \mathbb{R}^n, b \in \mathbb{R}} L_{Soft}(w, \xi, b, \lambda, \mu) = -\infty$

Soft-margin $C - SVC$: Dual form

$$Dual: C - SVC \begin{cases} Max \quad g(\lambda,\mu) = g(\lambda) = \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_j \lambda_i y_j y_i \left({x_j}^T x_i \right) \\ s.t \quad \sum_{i=1}^{n} \lambda_i y_i = 0 \\ C - \lambda_i - \mu_i = 0, i = 1, \dots, n \\ \mu_i \geq 0, \lambda_i \geq 0 \; i = 1, \dots, n \end{cases}$$

- $\forall i = 1, \dots, n : \; C - \lambda_i - \mu_i = 0 \Leftrightarrow C - \lambda_i = \mu_i$
- $\mu_i \geq 0 \Leftrightarrow C - \lambda_i \geq 0 \Leftrightarrow C \geq \lambda_i$ and $\lambda_i \geq 0 \Leftrightarrow C \geq \lambda_i \geq 0$

Theorem:

- Dual $\mathrm{C} - SVC$ poblem has a solution (λ^*, μ^*)
- Dual can simplified to a problem only for a single variable λ by eliminating the variable μ and then rewritten as a minimization problem Dual $(C - SVC)_\lambda$

Soft-margin $C - SVC$: Dual $(C - SVC)_\lambda$ Form

$$Dual\ (C-SVC)_\lambda : \begin{cases} Max \quad L_{Soft}(\lambda) = \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_j \lambda_i y_j y_i \left(x_j{}^T x_i\right) \\ s.t \quad \sum_{i=1}^{n} \lambda_i y_i = 0 \\ \quad C \geq \lambda_i \geq 0 \ i = 1, \dots, n \end{cases}$$

Theorem:

- The $Dual\ (C - SVC)_\lambda$ poblem is a Convex Quadratic Programming and has a solution λ^*
- For any solution $\lambda^* = (\lambda_1^*, \dots, \lambda_n^*)$, If there exists a component of λ^*, λ_j^*, such that $\lambda_j^* \in$ $(0, C)$ then a solution (w^*, b^*) to the primal problem $C - SVC$ w.r.t (w, b) can be obtained in by

$$w^* = \sum_{i=1}^{n} \lambda_i^* y_i x_i \quad and \ \ b^* = y_j - \sum_{i=1}^{n} \lambda_i^* y_i \left(x_j{}^T x_i\right)$$