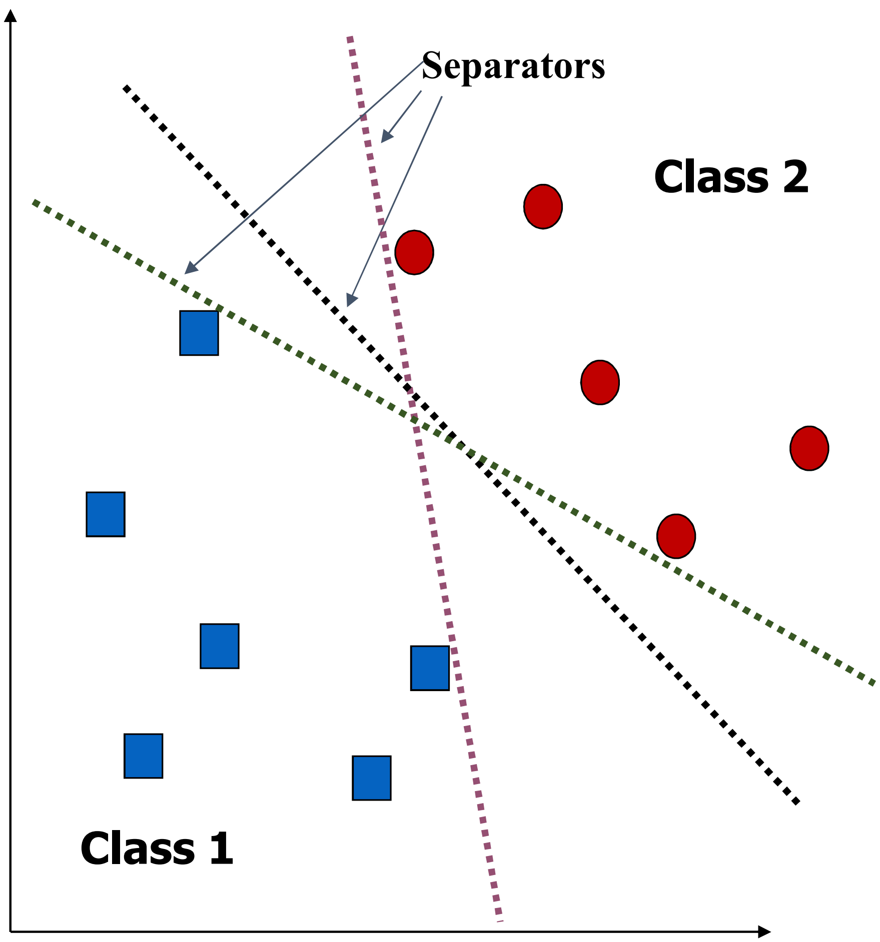
Support Vector machine
Linear Classification

Professor Abdellatif El Afia

Plan

- Support Vector Classification For linearly separable (Hard margin SVC)
- Linear C- Support Vector Classification(Soft margin SVC) ($C - SVC$)
- Bounded $C -$ Support Vector Classification($BC - SVC$)
- Least Squares $C -$ Support Vector Classification($\mathrm{L}SC - SVC$)
- Proximal $C -$ Support Vector Classification($\mathrm{P}C - SVC$)
- v-Soft margin SVC($v - SVC$)

Support Vector Classification For linearly separable

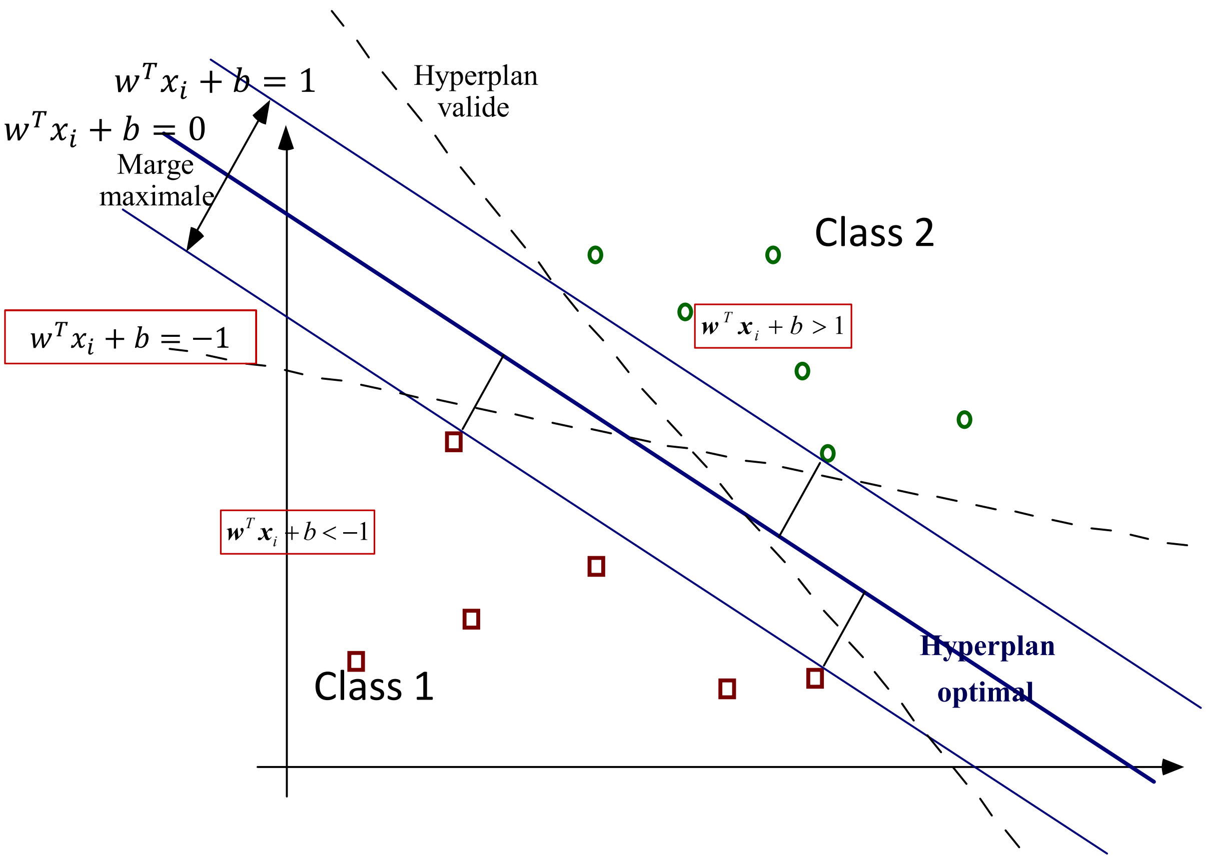


Problem with two linearly separable classes

- Several separators exist to separate classes; which one to choose?

- To minimize sensitivity to noise, the Separator should be as far away from the data the relatives of each class

Hard margin SVC



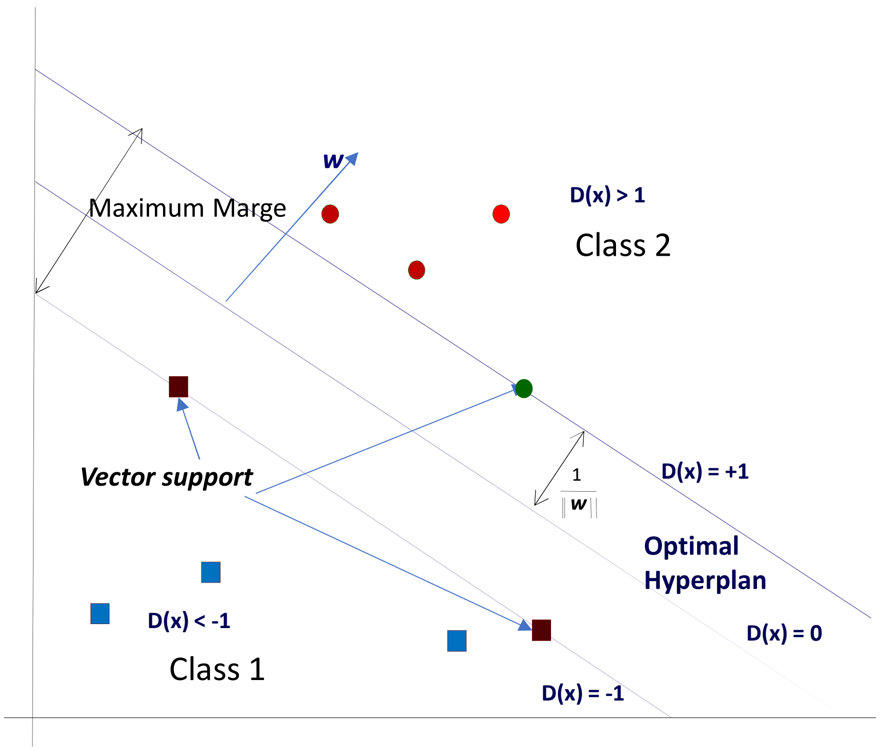
- Separator Equation :
$$y = w^T x + b$$
(Straight line in a two-dimensional space)

- If $\{(x_i, y_i\}_{i=1}^{n}$ is the data set and $y_i \in \{1, -1\}$ is the class of each, one should have:

$$y_i(w^T x_i + b) \geq 1 \quad \forall i$$

while having an optimal distance between x_i and the separator

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 ($|w^T x + b| = 1$) :

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

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

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

$$\text{Max m} \leftrightarrow Min\ \|w\| = \sqrt{w^T w}\ ;\ \text{m} = \frac{1}{\|w\|}$$

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

- $(SVC)_1$: $\begin{cases} Min & \|w\| \\ s.t & y_i(w^T x_i + b) \geq 1 \ \ i = 1, .., n \\ & w \in \mathbb{R}^d, b \in \mathbb{R} \end{cases}$

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: Algorithm

- Input: training set : $\{(x_i, y_i\}_{i=1}^{n}$ where $x_i \in \mathbb{R}^d, y_i \in \{1, -1\}$
- Contruct and solve the optimzation problem SVC obtaining (w^*, b^*)

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

- Contruit the separating hyperplane $(w^*)^T x + b^* = 0$ and the decision function is

$$h(x) = sign((w^*)^T x + b^*)$$

- Comput the Loss function

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

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

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

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

- $(\lambda \geq 0)$
- $g(\lambda) = \inf_{w \in \mathbb{R}^d, b \in \mathbb{R}} L_H(w, b, \lambda)$

Hard margin SVC: Dual form

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

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

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

Hard margin SVC: Dual form

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

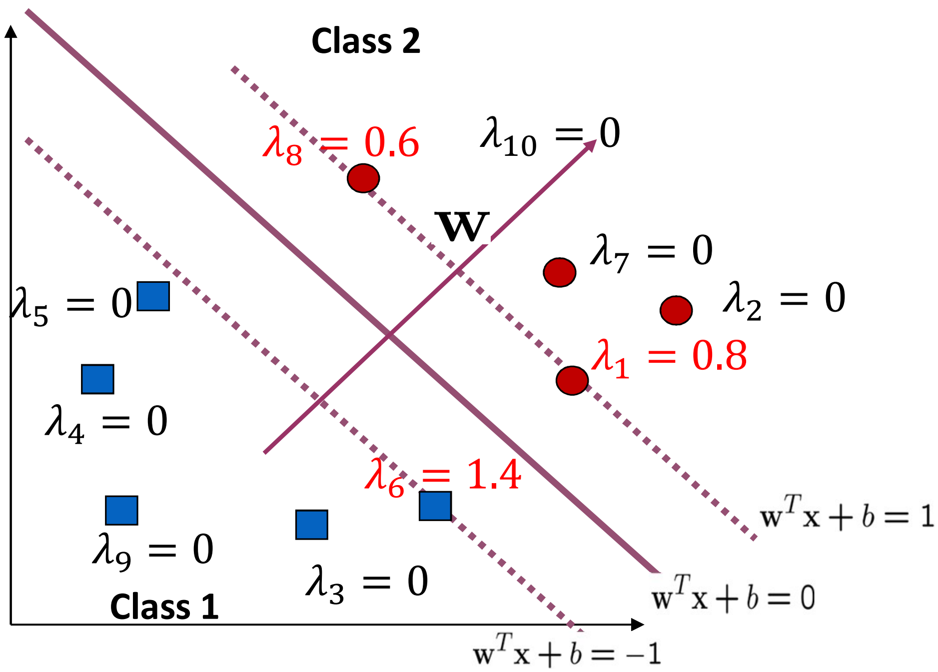
Theorem:

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

Hard margin SVC: Dual-Algorithm

- Input: training set : $S = \{(x_i, y_i\}_{i=1}^{n}$ where $x_i \in \mathbb{R}^d, y_i \in \{1, -1\}$
- Contruct and solve the optimzation problem $\text{Dual} - SVC$ obtaining $\lambda^* = (\lambda_1^*, \dots, \lambda_n^*)$

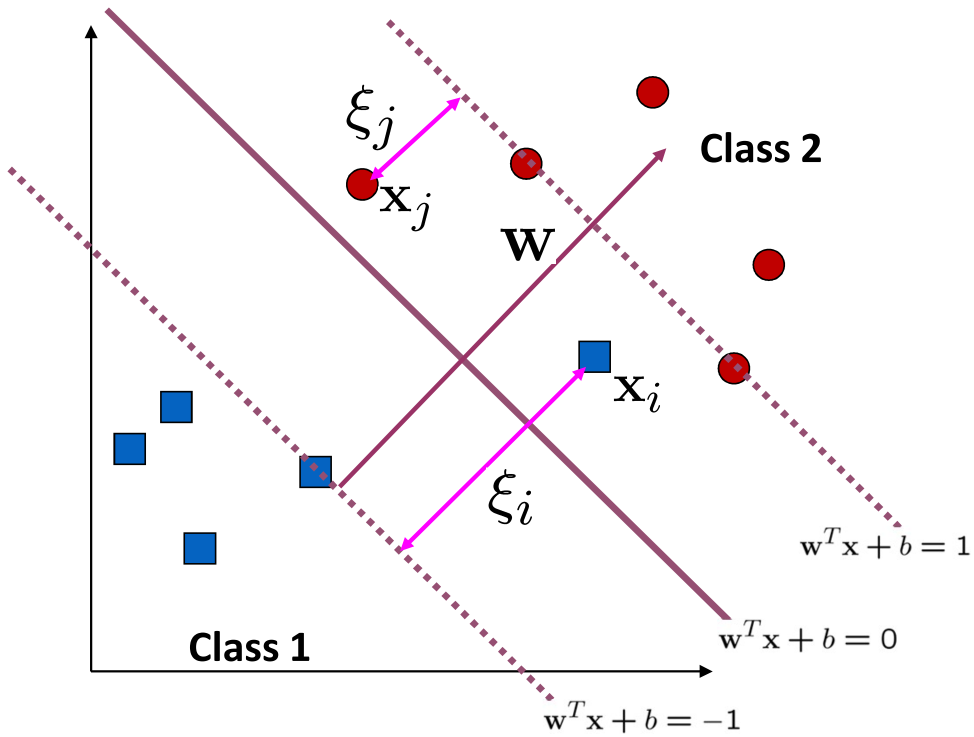
$$D - SVC: \begin{cases} Max \quad L_H(\lambda) = \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_j \lambda_i y_j y_i \left(x_j{}^T x_i \right) \\ s.t \quad \sum_{i=1}^{n} \lambda_i y_i = 0 \\ \qquad \lambda_i \geq 0 \; i = 1, \dots, n \end{cases}$$

- Choose a possitive component of λ^*, λ_j^*, and Compute

  $w^* = \sum_{i=1}^{n} \lambda_i^* \, y_i x_i$ and $b^* = y_j - \sum_{i=1}^{n} \lambda_i^* y_i \left(x_j{}^T x_i \right) \rightarrow h_{w^*, b^*}(x) = (w^*)^T x + b^*$
- Contruit the separating hyperplane $(w^*)^T x + b^*$ and the decision function is

$$h_S(x) = sign\left(h_{w^*, b^*}(x) \right) \rightarrow L_S(h_S) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{h_{w^*, b^*}(x_i) \neq y_i\}}$$

Soft margin SVC($C - SVC$)



- A margin of error can be introduced ξ_i for classification
- ξ_i are variables that give "soft" to optimal margins

$$\begin{cases} w^T x_i + b \geq 1 - \xi_i \; si \; y_i = 1 \\ \boldsymbol{w^T x_i + b \leq -1 + \xi_i \; si \; y_i = -1} \\ \xi_i \geq 0 \;\; \forall i \end{cases}$$

- The optimization problem becomes

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

- $C - SVC \begin{cases} Min & f(w,\xi) = (f_1(w), f_2(\xi)) \\ s.t & y_i(w^T x_i + b) \geq 1 - \xi_i, i = 1,..,n \\ & \xi_i \geq 0, w \in \mathbb{R}^d, b \in \mathbb{R} \end{cases}$
- $f(w,\xi) \approx \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i$

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

Soft margin SVC($C - SVC$)

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

Theorem:

- There exists solutions to the $C - SVC$ problem w.r.t (w, b)
- The solution $w^*$of the $C - SVC$ problem 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 SVC($C - SVC$): Algorithm

- Input: training set : $\{(x_i, y_i\}_{i=1}^{n}$where $x_i \in \mathbb{R}^d, y_i \in \{1, -1\}$
- Choose an approriate penalty parameter $C > 0$
- Contruct and solve the optimzation problem $\mathrm{C} - SVC$ obtaining (w^*, b^*, ξ^*)

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

- Contruit the separating hyperplane $(w^*)^T x + b^*$ and the decision function is

$$h(x) = sign((w^*)^T x + b^*)$$

Multi objectif optimization problem

- $C - SVC \begin{cases} Min & f = (f_1, f_2) \\ s.t & y_i(w^T x_i + b) \geq 1 - \xi_i, i = 1, .., n \\ & \xi_i \geq 0, w \in \mathbb{R}^d, b \in \mathbb{R} \end{cases}$
- $f_1 = m, f_2 = noise = \sum_{i=1}^{n} \xi_i$

- Agrregation approach: $f = p_1 f_1 + p_2 f_2$
- $\frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \xi_i \rightarrow p_2 = C$

- $C-SVC\begin{cases} Min & f=(f_1,f_2) \\ s.t & y_i(w^T x_i+b)\geq 1-\xi_i, i=1,..,n \\ & \xi_i\geq 0, w\in\mathbb{R}^d, b\in\mathbb{R} \end{cases}$

- $f_1 = marge = \frac{1}{2}\|w\|^2,\;\; f_2 = noise = \sum_{i=1}^{n}\xi_i$
- *Multiobjectif problem optimization*
- *Agregation approch:* $f = C_1 f_1 + C_2 f_2$
- *In Soft-SVC*: $C_1 = 1,\; C_2 = C$

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

Soft-margin $C - SVC$: Dual form

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

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

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

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

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

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

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

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

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

- $\nabla_\xi L_{Soft}(w,\xi,b,\lambda,\mu) = C I_{n\times n} - \lambda - \mu = 0$

- $\nabla_b L_{Soft}(w,\xi,b,\lambda,\mu) = \sum_{i=1}^{n}\lambda_i y_i = 0$

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

Soft-margin $C - SVC$: Dual form

Then

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

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

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

Else

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

Soft-margin $C - SVC$: Dual form

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

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 $(C - SVC)_\lambda$

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

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

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

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

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

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

- Input: training set : $\{(x_i, y_i\}_{i=1}^{n}$where $x_i \in \mathbb{R}^d, y_i \in \{1, -1\}$
- Choose an approriate penalty parameter $C > 0$
- Contruct and solve the optimzation problem Dual: $(C - SVC)_\lambda$obtaining $\lambda^* = (\lambda_1^*, \dots, \lambda_n^*)$

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

- Compute $w^* = \sum_{i=1}^{n} \lambda_i^* \, y_i x_i$
- Choose a possitive component of λ^*, $\lambda_j^* \in (0, C)$, and Compute $b^* = y_j - \sum_{i=1}^{n} \lambda_i^* y_i \left({x_j}^T x_i \right)$
- Contruit the separating hyperplane $(w^*)^T x + b^*$ and the decision function is

$$h(x) = sign((w^*)^T x + b^*)$$

Bounded $C - SVC$: BC $- SVC$

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

Theorem:

- The B$C - SVC$ problem is a Convex Quadratic Programming
- There exists a unique solution (w^*, b^*) to the B$C - SVC$ problem

Remarks:

The only difference between $BC - SVC$ and $C - SVC$ is that the term $\frac{1}{2}\|w\|^2$is replaced by $\frac{1}{2}(\|w\|^2 + b^2)$.This difference comes from the maximal principale in different spaces considered; in the objective, the term $\frac{1}{2}\|w\|^2$corresponds to $X - space$ while the term $\frac{1}{2}(\|w\|^2 + b^2)$ to the $X \times \{1\} - space$

$BC - SVC$: Algorithm

- Input: training set : $\{(x_i, y_i\}_{i=1}^{n}$ where $x_i \in \mathbb{R}^d, y_i \in \{1, -1\}$
- Choose an approriate penalty parameter $C > 0$
- Contruct and solve the optimzation problem $\mathrm{C} - SVC$ obtaining (w^*, b^*, ξ^*)

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

- Contruit the separating hyperplane $(w^*)^T x + b^*$ and the decision function is

$$h(x) = sign((w^*)^T x + b^*)$$

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

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

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

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

$\mathrm{B}C - SVC$: Dual form

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

$$
\begin{aligned}
L_{\mathrm{B}CSV}\ (w,\xi,b,\lambda,\mu) &= \frac{1}{2}(\|w\|^2+b^2)+C\sum_{i=1}^{n}\xi_i-\sum_{i=1}^{n}\lambda_i(y_i(w^Tx_i+b)-1+\xi_i)-\sum_{i=1}^{n}\mu_i\xi_i \\
&= \frac{1}{2}(\|w\|^2+b^2)-\sum_{i=1}^{n}\lambda_i(y_i(w^Tx_i+b)-1))-\sum_{i=1}^{n}(-C+\lambda_i+\mu_i)\xi_i
\end{aligned}
$$

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

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

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

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

$BC - SVC$: Dual form

Then

- $\nabla_w L_{\mathrm{B}CSV}\ (w, \xi, b, \lambda, \mu) = w - \sum_{i=1}^{n} \lambda_i y_i x_i = 0 \Longrightarrow w = \sum_{i=1}^{n} \lambda_i y_i x_i$
- $\nabla_\xi L_{\mathrm{B}CSVC}(w, \xi, b, \lambda, \mu) = CI_{n\times n} - \lambda - \mu = 0$
- $\nabla_b L_{\mathrm{B}CSVC}(w, \xi, b, \lambda, \mu) = b - \sum_{i=1}^{n} \lambda_i y_i = 0 \Longrightarrow b = \sum_{i=1}^{n} \lambda_i y_i$

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

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

If $CI_{n\times n} - \lambda - \mu = 0$ then

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

Else

- $\inf\limits_{w\in\mathbb{R}^d, \xi\in\mathbb{R}^n, b\in\mathbb{R}} L_{\mathrm{B}CSVC}(w, \xi, b, \lambda, \mu) = -\infty$

$\mathrm{B}C - SVC$: Dual form

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

Theorem:

- Dual $\mathrm{BC} - SVC$ 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 $(BC - SVC)_\lambda$

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

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

Theorem:

- The $Dual\ (BC - SVC)_\lambda$ poblem is a Convex Quadratic Programming and has a solution λ^*
- For any solution $\lambda^* = (\lambda_1^*, \dots, \lambda_n^*)$, then a solution (w^*, b^*) to the primal problem $(BC - SVC)_\lambda$ w.r.t (w, b) can be obtained in by

$$w^* = \sum_{i=1}^{n} \lambda_i^* y_i x_i \quad and \ \ b^* = \sum_{i=1}^{n} \lambda_i^* y_i$$

$\mathrm{B}C - SVC$: Dual $(BC - SVC)_\lambda$ Algorithm

- Input: training set : $\{(x_i, y_i\}_{i=1}^{n}$where $x_i \in \mathbb{R}^d, y_i \in \{1, -1\}$
- Choose an approriate penalty parameter $C > 0$
- Contruct and solve the optimzation problem Dual: $(C - SVC)_\lambda$obtaining $\lambda^* = (\lambda_1^*, \dots, \lambda_n^*)$

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

- Compute $w^* = \sum_{i=1}^{n} \lambda_i^* \, y_i x_i$
- Compute $b^* = \sum_{i=1}^{n} \lambda_i^* y_i$
- Contruit the separating hyperplane $(w^*)^T x + b^*$ and the decision function is

$$h(x) = sign((w^*)^T x + b^*)$$

Least Squares $C-SVC$(LS$C-SVC$)

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

Theorem:

- The LS$C-SVC$ problem is a Convex Quadratic Programming
- There exists solutions to the LS$C-SVC$ problem w.r.t (w, b)
- The solution $w^*$of the LS$C-SVC$ problem w.r.t w is unique

$LSC - SVC$: Algorithm

- Input: training set : $\{(x_i, y_i\}_{i=1}^{n}$ where $x_i \in \mathbb{R}^d, y_i \in \{1, -1\}$
- Choose an approriate penalty parameter $C > 0$
- Contruct and solve the optimzation problem $\text{C} - SVC$ obtaining (w^*, b^*, ξ_i^*)

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

- Contruit the separating hyperplane $(w^*)^T x + b^*$ and the decision function is

$$h(x) = sign((w^*)^T x + b^*)$$

$LSC - SVC$: Dual form

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

$$L_{\mathrm{LS}CSVC}(w,\xi,b,\lambda) \;=\; \frac{1}{2}\|w\|^2 + \frac{C}{2}\sum_{i=1}^{n}(\xi_i)^2 + \sum_{i=1}^{n}\lambda_i(y_i(w^T x_i + b) - 1 + \xi_i)$$

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

$$Dual\ C - SVC\begin{cases} Max \quad g(\mu) \overset{(w^*,b^*,\xi^*)}{=} \inf\limits_{w\in\mathbb{R}^d,\xi\in\mathbb{R}^n,b\in\mathbb{R}} L_{\mathrm{LS}CSVC}(w,\xi,b,\lambda) \\ s.t \qquad\qquad \lambda \in \mathbb{R}^n \end{cases}$$

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

$$\nabla_{w,\xi,b} L_{\mathrm{LS}CSVC}(w,\xi,b,\lambda) = 0$$

$\mathrm{LSC}-SVC$: Dual form

Then

- $\nabla_w L_{\mathrm{LS}CSVC}(w,\xi,b,\lambda) = w - \sum_{i=1}^{n} \lambda_i y_i x_i = 0 \Longrightarrow w = \sum_{i=1}^{n} \lambda_i y_i x_i$
- $\nabla_\xi L_{\mathrm{LS}CSVC}(w,\xi,b,\lambda) = C\xi - \lambda = 0 \Longrightarrow C\xi = \lambda$
- $\nabla_b L_{\mathrm{LS}CSVC}(w,\xi,b,\lambda) = \sum_{i=1}^{n} \lambda_i y_i = 0$

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

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

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

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

Else

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

$\mathrm{LS}C - SVC$: Dual form

$$\mathrm{DLS}C - SVC \begin{cases} Max \quad g(\lambda) = \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_j \lambda_i y_j y_i \left(\left({x_j}^T x_i \right) + \frac{\delta_{ij}}{C} \right) \\ s.t \quad \sum_{i=1}^{n} \lambda_i y_i = 0 \\ \quad \lambda \in \mathbb{R}^n \end{cases}$$

Theorem:

- $\mathrm{DLS}C - SVC$ poblem has a solution λ^*
- then a solution (w^*, b^*) to the primal problem $C - SVC$ w.r.t (w, b) can be obtained in by

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

$\mathrm{DLS}C - SVC$ Algorithm

- Input: training set : $\{(x_i, y_i\}_{i=1}^{n}$where $x_i \in \mathbb{R}^d, y_i \in \{1, -1\}$
- Choose an approrиate penalty parameter $C > 0$
- Contruct and solve the optimzation problem Dual: $\mathrm{DLS}C - SVC$ obtaining $\lambda^* = (\lambda_1^*, \dots, \lambda_n^*)$

$$\mathrm{DLS}C - SVC \begin{cases} Max \quad g(\lambda) = \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_j \lambda_i y_j y_i \left(\left({x_j}^T x_i \right) + \frac{\delta_{ij}}{C} \right) \\ s.t \quad \sum_{i=1}^{n} \lambda_i y_i = 0 \\ \quad \lambda \in \mathbb{R}^n \end{cases}$$

- Compute $w^* = \sum_{i=1}^{n} \lambda_i^* \, y_i x_i$
- Choose a possitive component of λ^*, λ_j^*, and Compute $b^* = y_j \left(1 - \frac{\lambda_j^*}{C} \right) - \sum_{i=1}^{n} \lambda_i^* y_i \left({x_j}^T x_i \right)$
- Contruit the separating hyperplane $(w^*)^T x + b^*$ and the decision function is

$$h(x) = sign((w^*)^T x + b^*)$$

Proximal $C - SVC$(P$C - SVC$)

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

Theorem:

- The P$C - SVC$ problem is a Convex Quadratic Programming
- There exists a unique solution to the P$C - SVC$ problem(w^*, b^*, ξ^*)

$PC - SVC$: Algorithm

- Input: training set : $\{(x_i, y_i\}_{i=1}^{n}$ where $x_i \in \mathbb{R}^d, y_i \in \{1, -1\}$
- Choose an approriate penalty parameter $C > 0$
- Contruct and solve the optimzation problem $\mathrm{C} - SVC$ obtaining (w^*, b^*, ξ^*)

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

- Contruit the separating hyperplane $(w^*)^T x + b^*$ and the decision function is

$$h(x) = sign((w^*)^T x + b^*)$$

$PC - SVC$: Dual form

$$\text{D}PC - SVC \begin{cases} Max \quad g(\lambda) = \sum_{i=1}^{n} \lambda_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_j \lambda_i y_j y_i \left(\left({x_j}^T x_i \right) + 1 \right) - \frac{1}{2C}\sum_{i=1}^{n} (\lambda_i)^2 \\ s.t \qquad\qquad\qquad\qquad\qquad\qquad \lambda \in \mathbb{R}^n \end{cases}$$

Theorem:

- D$PC - SVC$ poblem has a solution λ^*
- then a solution (w^*, b^*, ξ^*) to the primal problem P$C - SVC$ can be obtained in by

  - $w^* = \sum_{i=1}^{n} \lambda_i^* \, y_i x_i$
  - $b^* = \sum_{i=1}^{n} \lambda_i^* y_i$
  - $\xi^* = \frac{\lambda^*}{C}$

$\mathrm{DP}C - SVC$ Algorithm

- Input: training set : $\{(x_i, y_i\}_{i=1}^{n}$where $x_i \in \mathbb{R}^d, y_i \in \{1, -1\}$
- Choose an approriate penalty parameter $C > 0$
- Contruct and solve the optimzation problem Dual: $\mathrm{D}PC - SVC$obtaining $\lambda^* = (\lambda_1^*, \dots, \lambda_n^*)$

$$\mathrm{D}PC - SVC \begin{cases} Max \quad g(\lambda) = \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_j \lambda_i y_j y_i \left(\left({x_j}^T x_i \right) + 1 \right) - \frac{1}{2C} \sum_{i=1}^{n} (\lambda_i)^2 \\ s.t \qquad \lambda \in \mathbb{R}^n \end{cases}$$

- Compute $w^* = \sum_{i=1}^{n} \lambda_i^* \, y_i x_i$
- Compute $b^* = \sum_{i=1}^{n} \lambda_i^* y_i$
- Contruit the separating hyperplane $(w^*)^T x + b^*$ and the decision function is

$$h(x) = sign((w^*)^T x + b^*)$$

ν-Soft margin SVC($\nu - SVC$)

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

Where $\nu \in]0,1]$is a preselected parameter. Its dual is following

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

$v - SVC$:

the primal problem $v - SVC$ and the dual problem $Dv - SVC$ are the convex quadratic programming

Theorem:

Suppose that $\lambda^* = (\lambda_1^*, \ldots, \lambda_n^*)$is any solution to the dual problem $Dv - SVC$. If there exists two component of λ^*, λ_j^* and λ_k^*, such that :

- $\lambda_j^* \in \left]0, \frac{1}{n}\right[$and $y_j = 1$
- $\lambda_k^* \in \left]0, \frac{1}{n}\right[$and $y_k = -1$

Then a solution (w^*, b^*, ρ^*) to the primal problem $v - SVC$ w.r.t. (w, b, ρ) can be obtained by

- $w^* = \sum_{i=1}^{n} \lambda_i^* \, y_i x_i$
- $b^* = -\frac{1}{2} \sum_{i=1}^{n} \lambda_i^* y_i ((x_i)^T x_j + (x_i)^T x_k)$
- $\rho^* = \sum_{i=1}^{n} \lambda_i^* y_i (x_i)^T x_j + b^* = -\sum_{i=1}^{n} \lambda_i^* y_i (x_i)^T x_k - b^*$

Relationship $v - SVC$ and $C - SVC$

Theorem:

There exists a non-increasing function φ: $]0, +\infty[\rightarrow]0,1]$ $v = \varphi(C)$ such that

$$\forall \bar{C} \in]0, +\infty[, \varphi(\bar{C}) = \bar{v} \in]0,1]$$

The decision functions obtained by $v - SVC$ with $v = \bar{v}$ and $C - SVC$ with $\mathrm{C} = \bar{C}$ are identical if they can be computed by both of them , i.e.

- For $v - SVC$ with $v = \bar{v}$, two component of λ^*, λ_j^* and λ_k^*, such that :
  - $\lambda_j^* \in \left]0, \frac{1}{n}\right[$ and $y_j = 1$
  - $\lambda_k^* \in \left]0, \frac{1}{n}\right[$ and $y_k = -1$
- For $C - SVC$ with $\mathrm{C} = \bar{C}$, one component λ_j^* of λ^* can be chosen such that $\lambda_j^* \in]0, C[$

$v - SVC$: Significance of the parameter v

The significance of the parameter v is concerned with the terms of Support Vector and Training set with margin error.

Definition:

Suppose that $\lambda^* = (\lambda_1^*, \dots, \lambda_n^*)$ is the solution to the D$v - SVC$, and the corresponding solution to the $v - SVC$ is $(w^*, b^*, \rho^*, \xi^*)$. The training set $\{(x_i, y_i)\}_{i=1}^{n}$ is called training set with margin error if $y_j((w^*)^T x_j + b^*) < \rho^*$

Theorem

If $\rho^* > 0$ then

$$\frac{p}{n} \leq v \leq \frac{q}{n}$$

- p: the number of the training points with margin error
- q: the number of support vectors

$D\nu - SVC$: Algorithm

- Input: training set : $\{(x_i, y_i\}_{i=1}^{n}$ where $x_i \in \mathbb{R}^d, y_i \in \{1, -1\}$
- Choose an approriate penalty parameter $\nu \in]0,1]$
- Contruct and solve the convex quadratic programming $\mathrm{D}\nu - SVC$ obtaining $\lambda^* = (\lambda_1^*, \ldots, \lambda_n^*)$

- Compute b^*: Choose tow components of λ^*, λ_j^* and λ_k^*, such that :
  - $\lambda_j^* \in \left]0, \frac{1}{n}\right[$ and $y_j = 1$
  - $\lambda_k^* \in \left]0, \frac{1}{n}\right[$ and $y_k = -1$
  - And compute $b^* = -\frac{1}{2}\sum_{i=1}^{n} \lambda_i^* y_i((x_i)^T x_j + (x_i)^T x_k)$
- Compute $w^* = \sum_{i=1}^{n} \lambda_i^* \, y_i x_i$

- Contruit the separating hyperplane $(w^*)^T x + b^*$ and the decision function is

$$h(x) = sign((w^*)^T x + b^*)$$