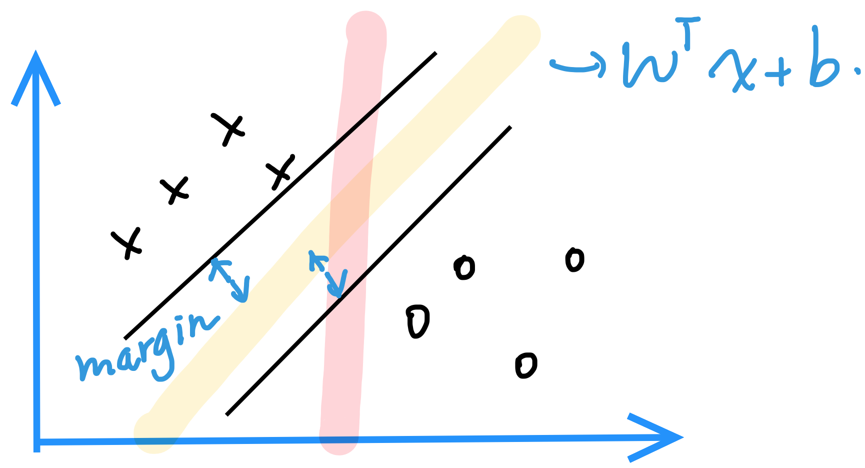
Support Vector Machine.

1. hard-margin SVM.

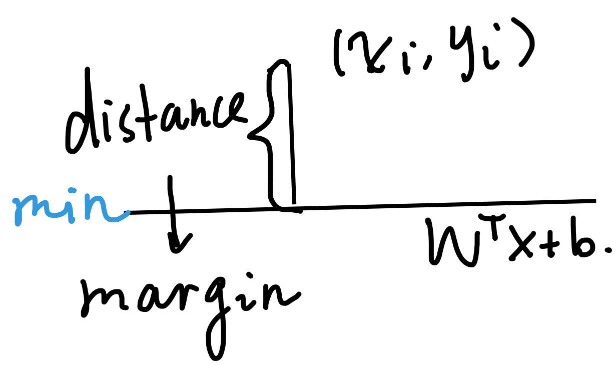


$\{(x_i, y_i)\}_{i=1}^{N}$,

$y_i \in \{-1, 1\}$.

$$\max \ \text{margin}(w,b)$$

$$\text{s.t.} \quad \left.\begin{array}{ll} w^T x_i + b > 0, & y_i = 1. \\ w^T x_i + b < 0, & y_i = -1. \end{array}\right\} \Rightarrow \begin{array}{l} y_i(w^T x_i + b) > 0 \\ \forall i = 1, 2, \cdots N. \end{array}$$



$$= \frac{1}{\|w\|} |w^T x + b|$$

$$\text{margin} = \min_{\substack{w,b \\ x_i. \\ i=1,2,\cdots N}} \text{distance}(w, b, x_i)$$

$$= \min_{\substack{w,b,x_i \\ i=1,2,\cdots N}} \frac{1}{\|w\|} |w^T x + b|$$

$$\Rightarrow \max \ \text{margin}(w,b) = \max_{w,b} \ \min_{\substack{x_i \\ i=1,2,\cdots N}} \frac{1}{\|w\|} |w^T x + b|$$

$$\text{s.t.} \quad y_i(w^T x_i + b) > 0.$$

$$\Rightarrow \max_{w,b} \min_{\substack{x_i \\ i=1,2,\cdots N}} \frac{1}{\|w\|} y_i(w^T x_i + b)$$

($y_i = \pm 1$ and $y_i(w^T x_i + b) > 0$)

$$= \max_{w,b} \frac{1}{\|w\|} \cdot \min_{\substack{x_i \\ i=1,2,\cdots N}} y_i(w^T x_i + b)$$

$$\text{s.t. } y_i(w^T x_i + b) > 0 \Rightarrow \exists\, \gamma > 0, \text{ s.t } \min_{\substack{x_i, y_i \\ i=1,2,\cdots N}} y_i(w^T x_i + b) = \gamma.$$

Let $\gamma = 1$,

$$\Rightarrow \max_{w,b} \frac{1}{\|w\|}$$

$$\text{s.t. } \min y_i(w^T x_i + b) = 1.$$ ($\Rightarrow y_i(w^T x_i + b) \geq 1$, $i = 1, 2, \cdots N$)

$$\boxed{\begin{aligned} \Rightarrow\ & \min_{w,b} \ \frac{1}{2} w^T w \\ & \text{s.t. } y_i(w^T x_i + b) \geq 1, \ \forall\, i = 1, 2, \ldots N. \end{aligned}}$$

Primal - Dual Problem.

$$\text{Lagrange}(w, b, \lambda) = \frac{1}{2} w^T \cdot w + \sum_{i=1}^{N} \underbrace{\lambda_i}_{\geq 0} \Big(\underbrace{1 - y_i (w^T x_i + b)}_{\leq 0.}\Big)$$

problem with constraints

$$\begin{cases} \min\limits_{w, b} \ \frac{1}{2} w^T w. \\ \text{s.t } y_i (w^T x_i + b) \geq 1 \iff 1 - y_i (w^T x_i + b) \leq 0. \end{cases}$$

Problem without constraints.

$$\begin{cases} \min\limits_{w, b} \ \max\limits_{\lambda} \ \text{Lagrange}(w, b, \lambda) \\ \text{s.t. } \lambda_i \geq 0. \end{cases}$$

$$\begin{cases} \max\limits_{\lambda} \ \min\limits_{w, b} \ \text{Lagrange}(w, b, \lambda) \\ \text{s.t. } \lambda_i \geq 0. \end{cases}$$

$$\min_{w, b} \text{Lagrange}(w, b, \lambda)$$

$$\frac{\partial \text{Lagrange}}{\partial b} = \frac{\partial}{\partial b} \left[\sum_{i=1}^{N} \lambda_i - \sum_{i=1}^{N} \lambda_i y_i (w^T x_i + b) \right]$$

$$= \frac{\partial}{\partial b}\left[-\sum_{i=1}^{N} \lambda_i y_i \cdot b\right]$$

$$= -\sum_{i=1}^{N} \lambda_i y_i$$

$$\frac{\partial}{\partial b} = 0, \quad \Rightarrow -\sum_{i=1}^{N} \lambda_i y_i = 0.$$

Remplace $\sum_{i=1}^{N} \lambda_i y_i$ by 0,

$$\Rightarrow \text{Lagrange}(w, b, \lambda) = \frac{1}{2}w^T w + \sum_{i=1}^{N} \lambda_i - \sum_{i=1}^{N} \lambda_i y_i w^T x_i - \sum_{j=1}^{N} \lambda_i y_i b.$$

$$= \frac{1}{2}w^T w + \sum_{i=1}^{N} \lambda_i - \sum_{i=1}^{N} \lambda_i y_i w^T x_i$$

$$\frac{\partial L}{\partial w} = \frac{1}{2} \cdot 2 \cdot w - \sum_{i=1}^{N} \lambda_i y_i x_i = 0$$

$$\Rightarrow w = \sum_{i=1}^{N} \lambda_i y_i x_i$$

$$\Rightarrow \text{Lagrange}(w, b, \lambda) = \frac{1}{2}\left(\sum_{i=1}^{N} \lambda_i y_i x_i\right)^T \left(\sum_{j=1}^{N} \lambda_j y_j x_j\right) - \sum_{i=1}^{N} \lambda_i y_i \left(\sum_{j=1}^{N} \lambda_j y_j x_j\right)^T x_i + \sum_{i=1}^{N} \lambda_i$$

$$= \frac{1}{2} \cdot \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_i x_i^T x_j -$$

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_i x_j^T x_i + \sum_{i=1}^{N} \lambda_i$$

$(\lambda_i, \lambda_j, y_i, y_i \in \mathbb{R})$

$$= -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_i x_i^T x_j + \sum_{i=1}^{N} \lambda_i$$

(min)

$$\Rightarrow \begin{cases} \min\limits_{\lambda} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_i x_i^T x_j + \sum_{i=1}^{N} \lambda_i \\ \text{s.t. } \lambda_i \geq 0, \sum_{i=1}^{N} \lambda_i y_i = 0. \end{cases}$$

KKT Condition.

$$\begin{cases} \frac{\partial \text{Lagrange}}{\partial w} = 0, \quad \frac{\partial \text{Lagrange}}{\partial b} = 0, \quad \frac{\partial \text{Lagrange}}{\partial \lambda} = 0. \\ \lambda_i (1 - y_i (w^T x_i + b_i)) = 0. \\ \lambda_i \geq 0. \\ 1 - y_i (w^T x_i + b) \leq 0. \end{cases}$$

slackness complementary

$$w^* = \sum_{i=0}^{N} \lambda_i y_i x_i$$

$\exists\ (x_k, y_k)$, s.t. $1 - y_k(w^T x_k + b) = 0.$

$\Rightarrow y_k(w^T x_k + b) = 1.$

$y_k^2 (w^T x_k + b) = y_k.$, $(y_k^2 = 1.)$

$$\Rightarrow \quad b = y_k - w^T \cdot x_k.$$

$$= y_k - \sum_{i=0}^{N} \lambda_i y_i x_i^T x_k.$$

$$\Rightarrow \quad \boxed{\begin{aligned} w^* &= \sum_{i=0}^{N} \lambda_i y_i x_i \\ b^* &= y_k - \sum_{i=0}^{N} \lambda_i y_i x_i^T x_k. \end{aligned}}$$

2. Soft-margin SVM.

$$\min \ \frac{1}{2} w^T w + loss.$$

(1) $loss = \sum_{i=1}^{N} I\{ y_i (w^T x_i + b) < 1 \}$.

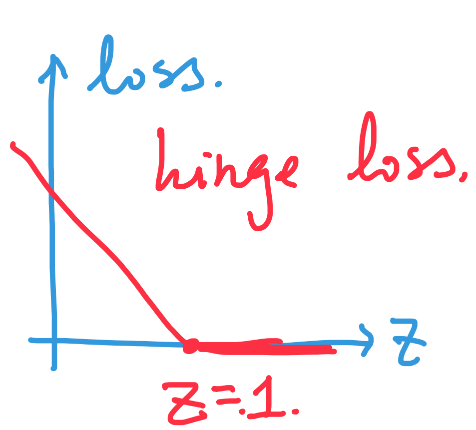
let $z = y(w^T x + b)$ $\quad loss_{0/1} = \begin{cases} 1 & (z<1) \\ 0 & (\text{otherwise}) \end{cases}$

(2) loss distance

if $y_i (w^T x_i + b) \geq 1$, $loss = 0$.

if $y_i (w^T x_i + b) < 1$, $loss = 1 - y_i (w^T x_i + b)$.

$$loss = \max\{0, 1 - \underbrace{y_i (w^T x_i + b)}_{z}\}.$$



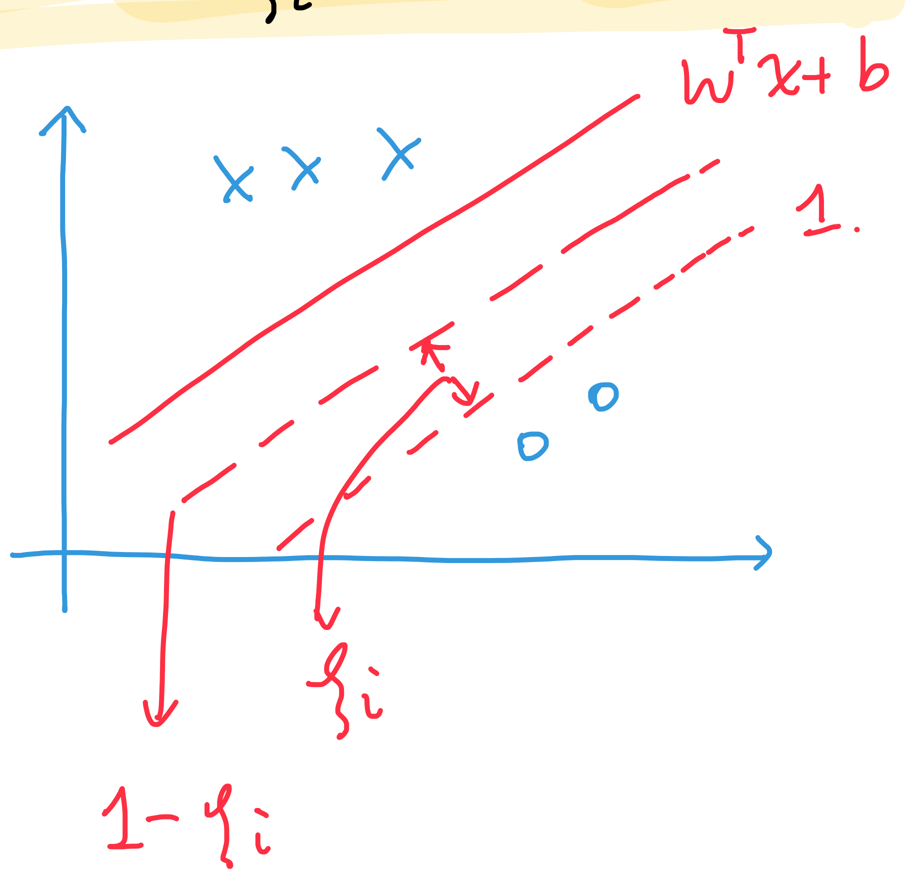
Soft-margin SVM.

$$\begin{cases} \min \ \frac{1}{2}w^Tw + C \cdot \sum_{i=1}^{N} \max\{0, 1-y_i(w^Tx_i+b)\}. \\ \text{s.t.} \quad y_i(w^Tx_i+b) \geq 1. \end{cases}$$

$i = 1, 2, \cdots N.$

Let $\xi_i = 1 - y_i(w^Tx_i+b)$, $\xi_i \geq 0.$

$$\begin{cases} \min\limits_{w,b} \ \frac{1}{2}w^Tw + C \cdot \sum_{i=1}^{N} \xi_i \\ \text{s.t.} \quad y_i(w^Tx_i+b) \geq 1 - \xi_i \\ \qquad \xi_i \geq 0. \end{cases}$$



3. Constraint Optimatization - Weak Duality.

primal problem

$$\begin{cases} \min\limits_{x \in \mathbb{R}^p} \quad f(x) \\ \\ \text{s.t.} \quad m_i(x) \leq 0, \quad i = 1, \dots M. \\ \qquad\;\; n_j(x) = 0, \quad j = 1, \dots N. \end{cases}$$

$$\mathcal{L}agrange(x, \lambda, \eta) = f(x) + \sum_{i=1}^{N} \lambda_i m_i + \sum_{j=1}^{N} \eta_j n_j$$

$$\begin{cases} \min\limits_{x} \quad \max\limits_{\lambda, \eta} \; \mathcal{L}agrange(x, \lambda, \eta) \\ \\ \text{s.t.} \quad \lambda_i \geq 0 \end{cases}$$

dual problem.

$$\begin{cases} \max\limits_{\lambda, \eta} \quad \min\limits_{x} \; \mathcal{L}agrange(x, \lambda, \eta) \\ \\ \text{s.t.} \quad \lambda_i \geq 0. \end{cases}$$

proof: $\max\limits_{\lambda, \eta} \min\limits_{x} \mathcal{L} \leq \min\limits_{x} \max\limits_{\lambda, \eta} \mathcal{L}.$

$$\underbrace{\min_{x} \mathcal{L}(x, \lambda, \eta)}_{A(\lambda, \eta)} \leq \mathcal{L}(x, \lambda, \eta) \leq \underbrace{\max_{\lambda, \eta} \mathcal{L}(x, \lambda, \eta)}_{B(x)}$$

$A(\lambda, \eta) \leq B(x) \Rightarrow A(\lambda, \eta) \leq \min B(x).$

$\Rightarrow \max A(\lambda, \eta) \leq \min B(x)$

$\Rightarrow \max\limits_{\lambda, \eta} \min\limits_{x} \mathcal{L} \leq \min\limits_{x} \max\limits_{\lambda, \eta} \mathcal{L}$

4. Slater Condition.

$$\begin{cases} \min f(x) \\ x \in \mathbb{R}^p \\ \text{s.t. } m_1(x) \leq 0. \end{cases}$$

$D = \operatorname{dom} f \cap \operatorname{dom} m_1.$

$\mathcal{L}agrange(x, \lambda) = f(x) + \lambda m_1(x), \ \lambda \geq 0.$

$p^* = \min f(x)$

$d^* = \max\limits_{\lambda} \min\limits_{x} \mathcal{L}agrange(x, \lambda)$

$G = \{ (m_1(x), f(x)) \mid x \in D \}.$

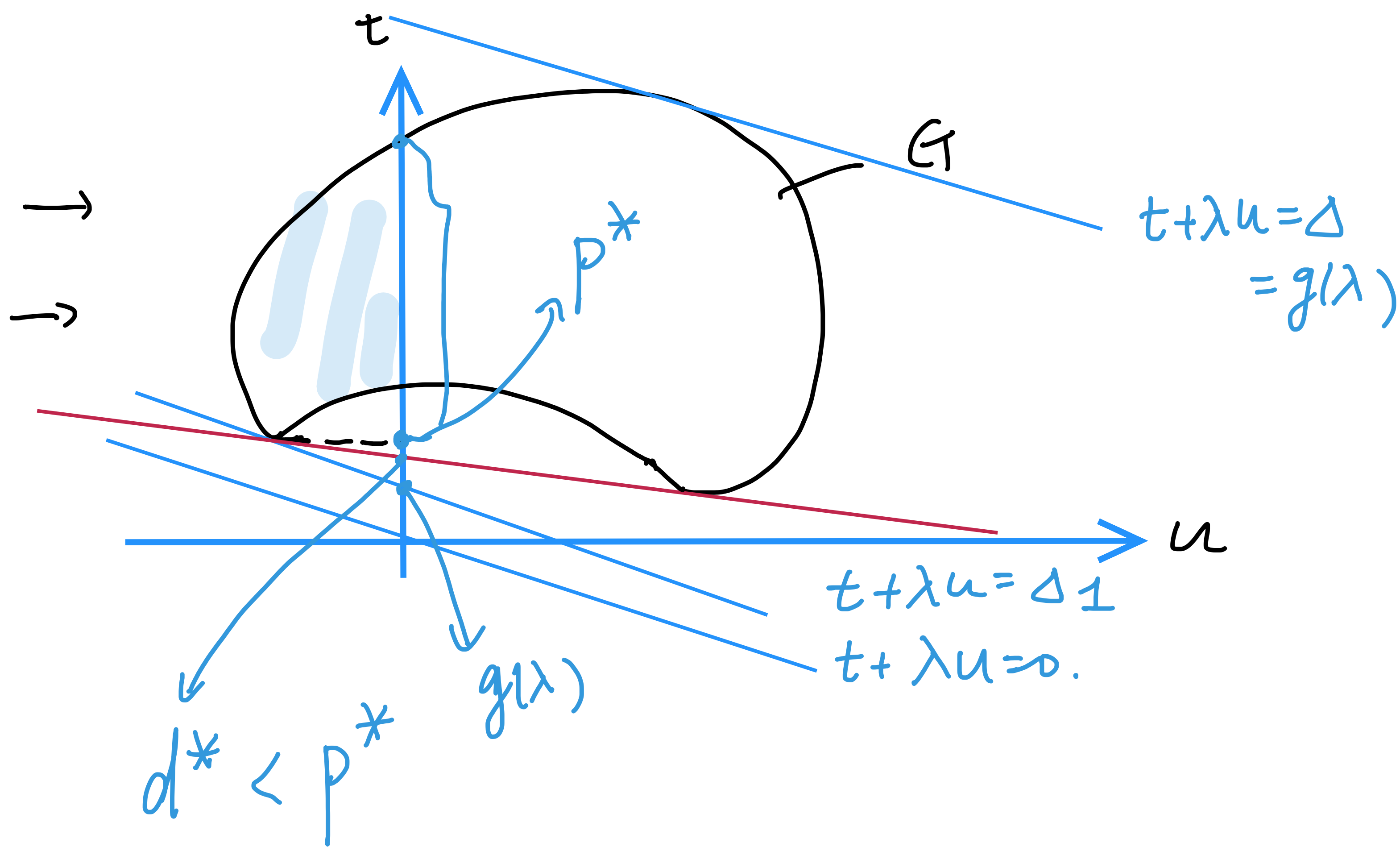
$= \{ (u, t) \mid x \in D \}.$

$p^* = \inf \{ t \mid (u, t) \in G, u \leq 0 \}.$

$$d^* = \max_{\lambda} \underbrace{\min (t + \lambda u)}_{g(\lambda)}$$

$$= \max_{\lambda} g(\lambda)$$

$$g(\lambda) = \inf \{ t + \lambda u \mid (u, t) \in G \}.$$



Slater condition:

$\exists \hat{x} \in$ relative interior D,

s.t $\forall\, i = 1, \dots, M.$ $m_i(\hat{x}) < 0.$

Conclusion: Convex + Slater $\Rightarrow$ Strong Duality.

5. KKT Condition.

Primal problem.	Dual Problem.
$\begin{cases} \min f(x) \\ \text{s.t. } m_i(x) \le 0, \quad i=1,2,\dots m. \\ \qquad n_j(x) \le 0, \quad j=1,2,\dots m. \end{cases}$	$\begin{cases} \max\limits_{\lambda,\eta} g(\lambda,\eta) \\ \text{s.t. } \lambda \geq 0. \end{cases}$

$$Lagrange(x, \lambda, \eta) = f(x) + \sum_i \lambda_i m_i + \sum_j n_j \eta_j$$

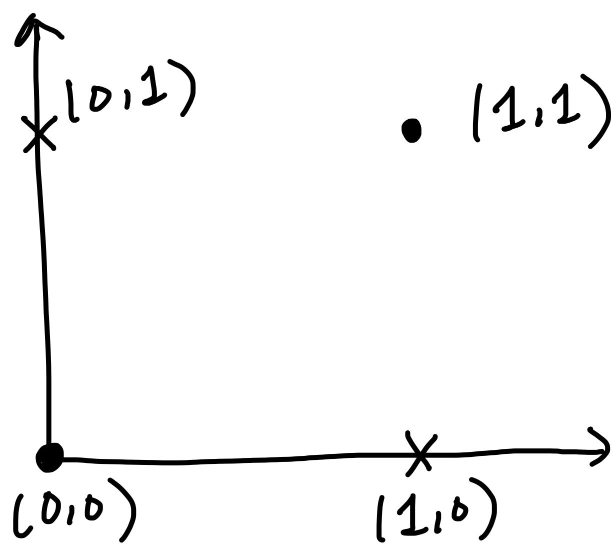
$$g(\lambda, \eta) = \min_x \mathcal{L}(x, \lambda, \eta)$$

$$\begin{cases} \text{primal feasibility} \rightarrow \begin{cases} m_i(x^*) \le 0 \\ n_j(x^*) = 0 \\ \lambda^* \geq 0 \end{cases} \\ \\ \text{Complementary slackness: } \lambda_i^* m_i = 0. \\ \\ \text{gradient} = 0 : \left. \dfrac{\partial Lagrange(x, \lambda^*, \eta^*)}{\partial x} \right|_{x = x^*} = 0. \end{cases}$$

6. Kernel Method.

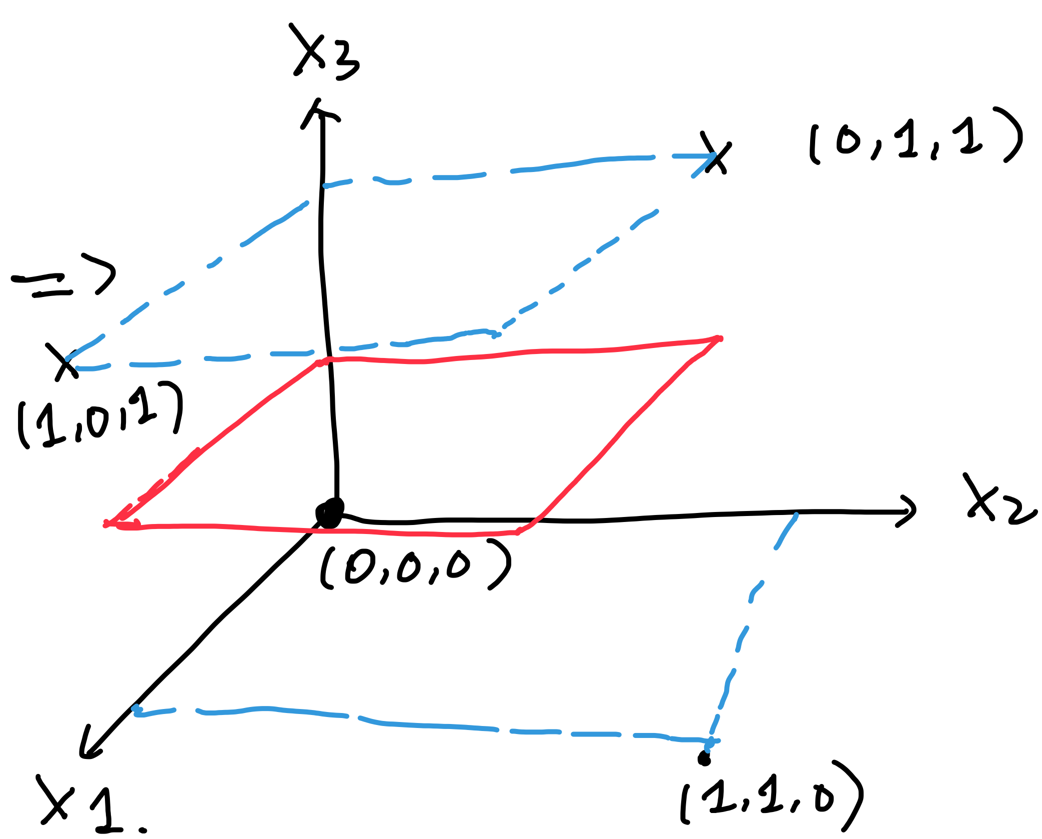
$$\underset{\text{input space}}{x} \xrightarrow{\phi(x)} \underset{\text{feature space.}}{z}$$

example:



$$2D \xrightarrow{\phi(x)} 3D.$$

$$(x_1, x_2) \qquad z = (x_1, x_2, (x_1 - x_2)^2)$$



kernel function:

$$K(x_1, x_2) = \phi(x_1)^T \phi(x_2) = \langle \phi(x_1), \phi(x_2) \rangle$$

$$\text{s.t. } K(x_1, x_2) = \phi(x_1)^T \cdot \phi(x_2)$$

$$k(x_1, x_2) = \exp\left(- \frac{(x_1 - x_2)^2}{2\sigma^2}\right)$$

STT3795 - Ex3

svm_train:

N: number of training samples.

$$\min \quad \frac{1}{2} \sum_{i,j=1}^{N} c_i c_j \lambda_i \lambda_j k_{ij} - \sum_{i=1}^{N} \lambda_i$$

$$\text{s.t.} \quad -\lambda_i \leq 0 \quad (\forall \; 1 \leq i \leq N)$$

$$\sum_{i=1}^{N} c_i \lambda = 0.$$

$$\frac{1}{2} \sum_{i,j=1}^{N} \lambda_i \lambda_j \cdot \underbrace{\sum_{i,j=1}^{N} c_i c_j}_{\substack{\text{np.outer}(c, c) \\ (N \times N)}} \cdot \underbrace{\sum_{i,j=1}^{N} k_{ij}}_{\substack{*\ \text{kernel matrice.} \\ (N \times N)}} - \sum_{i=1}^{N} \lambda_i$$

Résoudre le problème de maximisation :

$$\underset{\lambda_1, \dots, \lambda_N}{\arg\max} \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i,j=1}^{N} c_i c_j \lambda_i \lambda_j K_{ij}$$

$$\text{s.t.} \quad \forall\, i \leq i \leq N. \quad \lambda_i \geqslant 0.$$

$$\sum_{i=1}^{N} c_i \lambda_i = 0.$$

$\Rightarrow$ Transformer le problème de maximisation vers le problème de minimisation

$$\underset{\lambda_1, \dots \lambda_N}{\arg\min} \quad \frac{1}{2} \sum_{i,j=1}^{N} c_i c_j \lambda_i \lambda_j K_{ij} - \sum_{i=1}^{N} \lambda_i$$

$$\text{s.t.} \quad \forall\, i \leq i \leq N \qquad -\lambda_i \leqslant 0$$

$$\sum_{i=1}^{N} c_i \lambda_i = 0$$

$$\min \frac{1}{2} \sum_{i,j=1}^{N} c_i c_j \lambda_i \lambda_j k_{ij} - \sum_{i=1}^{N} \lambda_i$$

$$\Rightarrow \min \frac{1}{2} \underbrace{\sum_{i=1}^{N} \lambda_i}_{\lambda^T} \sum_{i,j=1}^{N} c_i c_j k_{ij} \underbrace{\sum_{j=1}^{N} \lambda_j}_{\lambda} - \sum_{i=1}^{N} \lambda_i$$

Selon user guide de cvxopt,

$$\min \quad \underbrace{\frac{1}{2} x^T P x + q^T x.}_{} \longrightarrow -\sum_{i=1}^{N} \lambda_i \Rightarrow$$

$$\text{s.t.} \quad Gx \leq h.$$

$$Ax + b.$$

$$\overset{\longleftarrow N \longrightarrow}{\underbrace{(-1, -1, \cdots -1)}_{q^T}} \cdot \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \end{pmatrix}$$

$\Rightarrow$ q = - np.ones(N)

$$P = \sum_{i,j=1}^{N} c_i c_j * \sum_{i,j=1}^{N} K_{ij}$$

$$\underbrace{\begin{matrix} [[\, c_0 c_0 & c_0 c_1 & \cdots & c_0 c_N] \\ [\, c_1 c_0 & c_1 c_1 & \cdots & c_1 c_N] \\ [\, \vdots & & & \vdots\ ] \\ [\, c_N c_0 & c_N c_1 & \cdots & c_N c_N\,]] \end{matrix}}_{\sum_{i,j=1}^{N} c_i c_j} * \underbrace{\begin{bmatrix} K_{00} & K_{01} & \cdots & K_{0N} \\ K_{10} & K_{11} & \cdots & K_{1N} \\ \vdots & & & \\ K_{N0} & K_{N1} & \cdots & K_{NN} \end{bmatrix}}_{\sum_{i,j=1}^{N} K_{ij}}$$

$\Downarrow$ numpy.outer(c, c) $\qquad\qquad$ $\Downarrow$ kernel matrix

$\Rightarrow$ P = kernel matrix * np.outer(c, c)

$$Gx \leq h \Rightarrow \forall\ 1 \leq i \leq N \quad -\lambda_i \leq 0$$

$$\underbrace{\begin{bmatrix} -1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix}}_{G} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_N \end{pmatrix} \leqslant \underbrace{\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{h}$$

N (columns), N (rows)

$$\Rightarrow G = -\text{np.eye}(N)$$

$$h = \text{np.zeros}(N)$$

$$Ax = b. \Rightarrow \sum_{i=1}^{N} c_i \lambda_i = 0.$$

$$\underbrace{(c_1, c_2, \cdots c_N)}_{C^T} \cdot \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \end{pmatrix} = \underset{\substack{\downarrow \\ 0}}{0.}$$

$$\Rightarrow A = C^T \quad (\dim = 1 * N)$$

$$b = 0.$$