- Standard Form Problem ①

$$\min f_0(x)$$

$$\text{s.t } f_i(x) \le 0 \ , \ i = 1, \dots, m$$

$$h_i(x) = 0 \ , \ i = 1, \dots, p$$

variable $x \in R^n$, domain D, optimal value p^*

$\Rightarrow$ Lagrangian : $L : R^n \times R^m \times R^p \to R$

with $\operatorname{dom} L = D \times R^m \times R^p$

$$L(x, \lambda, v) = f_0(x) + \sum_{i=1}^{m} \lambda_i \underbrace{f_i(x)}_{\text{constraint function}} + \sum_{i=1}^{p} v_i \underbrace{h_i(x)}_{\text{equality}}$$

λ, v are Lagrange multiplier

- Lagrange dual function ②

$$g : R^m \times R^p \to R$$

$$g(\lambda, v) = \inf_{x \in D} L(x, \lambda, v)$$

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

g is concave, can be $-\infty$ for some λ, ν

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

1, 2 → convex problem

→ For a convex minimization problem with inequality Constru

$$\min_{x} \quad f(x)$$

$$\text{s.t} \quad f_i(x) \preceq 0, \quad i = 1, \dots, m$$

→ The Lagrangian dual problem is

$$\underset{u}{\text{maximize}} \quad \inf_{x} \left(f(x) + \sum^{m} \alpha_i f_i(x) \right)$$

$$\text{s.t} \quad \alpha_i \geq 0, \quad i = 1, \dots, m$$

Then, For soft SVM

→ Primal problem: $\triangle 1$

$$\min_{w, b, \xi} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \xi_i$$

$$\text{s.t} \quad y_i(w \cdot x_i + b) \geq 1 - \xi_i \ , \ i = 1, 2, \ldots, N$$

$$\xi_i \geq 0 \ , \ i = 1, \ldots, N$$

→ Lagrangian dual problem :

$$\min_{\alpha} \quad \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{N} \alpha_i$$

$$\text{s.t} \quad \sum_{i=1}^{N} \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C.$$

Deriv**ation**

Step1 Lagrangian for $\triangle$:

$$L(w, b, \xi, \alpha, \mu) \equiv \underbrace{\frac{1}{2}\|w\|^2 + C \sum_{i=1}^{N} \xi_i}_{f_0(x)} - \sum_{i=1}^{N} \alpha_i \underbrace{(y_i(w \cdot x + b) - 1 + \xi_i)}_{-f_i(x)} - \sum_{i=1}^{N} \mu_i \underbrace{\xi_i}_{h_i(x)}$$

and $\alpha_i \geq 0 \ , \ \mu_i \geq 0$

$$\nabla_w L = w - \sum_{i=1}^{N} \alpha_i y_i x_i \quad \nabla_b L = -\sum_{i=1}^{N} \alpha_i y_i \quad \nabla_\xi L = C - \sum_{i=1}^{N} \mu_i - \sum_{i=1}^{N} \alpha_i$$

let $\nabla_w L = \nabla_b L = \nabla_\xi L = 0$ and get $\min L = -\frac{1}{2}\sum\sum \alpha_i \alpha_j y_i y_j k(x_i, x_j) + \sum \alpha_i$

there is only α in it now, do $\max_\alpha L$, which is $\min -L$

finally we get the dual problem:

$$\min_\alpha \quad \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j k(x_i, x_j) - \sum_{i=1}^{N} \alpha_i$$

Step 2. $\quad$ s.t $\Rightarrow$ $\sum_{i=1}^{N} \alpha_i y_i = 0$

$$C - \alpha_i - \mu_i = 0$$

$$\alpha_i \geq 0$$

$$\mu_i \geq 0, \; i = 1, \dots, N$$

$\Rightarrow$ s.t $\quad 0 \leq \alpha_i \leq C$

P.S $\quad$ we know $w = \sum_{i=1}^{n} \alpha_i y_i \phi(x_i)$

$$w^T w \Rightarrow \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \underbrace{\phi(x_i)\phi(x_j)}_{k(x_i, x_j)}$$

---

kernels:

Linear $\rightarrow$ $k(x, z) = x^T z$

Polynomial $\rightarrow$ $k(x, z) = (1 + x^T z)^d$

Radiu Basis Function (rbf, gaussian kernel)

$\rightarrow$ $k(x, z) = e^{-\frac{\|x - z\|^2}{\sigma^2}}$

$\downarrow$

kernel, x_i is the samplepoint

x_j is the variable.

$Exp \to e^{\frac{-\|x-z\|}{\sigma}}$, Laplacian $\to e^{\frac{-|x-z|}{\sigma}}$, Sigmoid kernel $\to \tanh(ax^T+c)$

Explanation of KKT

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

Gradient vector parraline :

$\nabla f = \lambda \nabla g$ 用f去寻找g的切点

min f 且 $g=0$

case 1: Unconstrained local minimum occurs in the fesible region

$g(x^*) < 0$ (<0 $\boxed{\text{in}}$ fesible region)

$\nabla f(x^*) = 0$

case 2: Unconstraied local minimum lies outside the feasible region

$g(x^*) = 0$

$-\nabla f(x^*) = \lambda \nabla_x g(x^*) \Rightarrow$ 相切

$\lambda \geq 0$

<u>KKT</u>

$\Downarrow$

Given the optimization problem:

$$\min_{x \in R^2} f(x) \quad s.t \quad g(x) \leq 0$$

$\Rightarrow \lambda \geq 0$

Define $L(x, \lambda) = f(x) + \lambda g(x) \Rightarrow \nabla f(x) = \lambda \nabla g(x)$

$\nabla L = 0 \Rightarrow$ 相切.

Then x^* a local minimum $\Longleftrightarrow$ there exists a unique λ^* s.t

① $\nabla_x L(x^*, \lambda^*) = 0 \rightarrow$ 相切 $= \nabla f_0(x) + \sum \lambda g(x) + \underbrace{\sum v h(x)}_{=0}$

② $\lambda^* \geq 0$

③ $\lambda^* g(x^*) = 0 \Longleftrightarrow$ complementary slackness

If a dual variable $\boxed{(\lambda)}$ is greater or equal than zero then the corresponding primal constraint must be tight

④ $g(x^*) \leq 0$

$(\alpha^* g(x^*) = 0)$ $\cup$

kkT : ①, ②, ③, ④

---

by kkT $\Rightarrow w = \sum_{i=1}^{N} \alpha_i y_i x_i$ and $y_i(w_i x_i + b) - 1 = 0$

$\Rightarrow b = y_i - \sum_{i=1}^{N} \alpha_i y_i k(x_i, x) \Rightarrow f(x) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y k(x_i, x) + b\right)$