Iteso Universidad Jesuita de Guadalajara

convex optimization

Exam

Prof. Juan Diego Sanchez T

NAME: ASHWIN BHAT

| | | | |
|---|---|---|---|
| 1. | | 2. | |
| 3. | | 4. | |
| 5. | | 6. | |
| 7. | | 8. | |
| | | TOTAL | |

1 (a)

Solution:

Let $f(x) = \min_{x \in R^n} c^T x$ — (1)

s.t $Ax \leq b$

i.e $b - Ax \geq 0$

$\therefore$ Let $g(x) = b - Ax$

The Lagrangian of (1) is given by

$$L(x,u) = f(x) - u\,g(u) \qquad u \geq 0$$

$$= c^T x - u(b - Ax)$$

$$= \cancel{(c^T + Au)x - ub}$$

$$= c^T x + uAx - ub$$

$$= (c^T + uA)x - ub$$

$$= ((c^T + uA)^T)^T x - b^T u$$

$$= ((c^T)^T + A^T u)^T x - b^T u$$

$$= (c + A^T u)^T x - b^T u$$

Derivating with respect to $x$, we get

$$\frac{\partial L(x,u)}{\partial x} = c + A^T u$$

$$0 = c + A^T u \quad \text{— (2)}$$

Substituting eq (2) in Lagrangian, we get Dual form as

$$g(u) = \min_{x \in R^n} L(x,u) = -b^T u$$

Solution:

1(b) Let $f(x) = \min_{x \in R^n} \frac{1}{2} x^T Q x + c^T x$

The constraints are given as

$$g(x) = x \geq 0$$

$$l(x) = Ax - b$$

The Lagrangian ~~function~~ is given as below

$$L(x,u,v) = f(x) - u\,g(x) + v\,l(x) \quad \text{for } u \geq 0$$

$$= \frac{1}{2} x^T Q x + c^T x - ux + v(Ax - b) \quad \text{for } u \geq 0$$

$$= \frac{1}{2} x^T Q x + c^T x - u^T x + v^T (Ax - b)$$

Derivating with respect to $x$, we get

$$\frac{\partial L}{\partial x}(x,u,v) = Qx + c^T - u^T + v^T A$$

$$0 = Qx + c^T - u^T + v^T A$$

$$\therefore Qx = -(c^T - u^T + v^T A)$$

$$x = -Q^{-1}(c^T - u^T + v^T A)$$

$$x = -Q^{-1}(c - u + A^T v)^T$$

Now derivating w.r.t to $v$, we get

$$\frac{\partial L(x,u,v)}{\partial v} = Ax - b$$

$$0 = Ax - b$$

$$x = -Q^{-1}(c - u + A^T v)^T$$

$$Qx = -(c^T - u^T + v^T A)$$

$$x^T Q x = -x^T(c^T - u^T + v^T A)$$

$$x^T Q x = -(c^T - u^T + v^T A)x$$

$$x^T Q x = -(c^T - u^T)x + v^T A x$$

$$\therefore (c^T - u^T)x = -x^T Q x - v^T b \qquad \because Ax = b$$

$$= -x^T Q x - b^T v$$

So we ~~have the~~ Dual ~~form given as below~~

$\cancel{g(u,v)} =$

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

$$= \frac{1}{2}x^T Q x - x^T Q x - b^T v + v^T(0) \qquad \because Ax = b$$

$$= -\frac{1}{2}x^T Q x - b^T v$$

The Dual form is given as

$$g(u,v) = -\frac{1}{2}(c - u + A^T v)Q^{-1} Q Q^{-1}(c - u + A^T v)^T - b^T v$$

$$= \cancel{-\frac{1}{2}(c^T - u^T + A^T v)^T Q^{-1}} \; (\cdots$$

$$= -\frac{1}{2}(c - u + A^T v)^T Q (c - u + A^T v) - b^T v$$

1 (c) consider the Quadratic optimization problem

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T Q x + c^T x \qquad (3)$$

$$\text{s.t} \quad Ax = 0$$

with $Q \geq 0$ and symmetric. Show that by the KKT conditions for this convex problem with no inequality constraints, $x$ is a solution if satisfies

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix} = \begin{bmatrix} -c \\ 0 \end{bmatrix} \qquad (4)$$

for some $v$. That is, the system (4) provides the optimality condition for (3)

Solution:

Let $f(x) = \min_{x \in \mathbb{R}^n} \frac{1}{2} x^T Q x + c^T x$

and constraint $g(x) = Ax$

The Lagrangian is given by

$$L(x,v) = \frac{1}{2} x^T Q x + c^T x + vAx$$

Derivate with respect to $x$ we get

$$\frac{\partial L}{\partial x} = Qx + c + A^T v$$

$$0 = Qx + c + A^T v$$

$$Qx + A^T v = -c \quad \text{—— (1)}$$

Now derivating with respect to v, we get

$$\frac{\partial L}{\partial v} = Ax$$

$$Ax = 0 \quad \text{—— (2)}$$

Adding (1) and (2) as Matrix form we get

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix} = \begin{bmatrix} -c \\ 0 \end{bmatrix}$$

1d) Consider the quadratic optimization problem

$$\min_{x \in R^n} \frac{1}{2} x^T Q x + c^T x \qquad (5)$$

$$\text{s.t} \quad Ax = b$$

with $Q \succ 0$. Find the optimality conditions for (5)

Let f(x) objective function given as

$$f(x) = \min_{x \in R^n} \frac{1}{2} x^T Q x + c^T x$$

The constraint is given as

$$g(x) = Ax - b$$

The Lagrangian is given as

$$L(x, v) = f(x) + v\, g(x)$$

$$= \frac{1}{2} x^T Q x + c^T x + v(Ax - b)$$

Now derivating with respect to $x$, we get

$$\frac{\partial L}{\partial x} = Qx + c^T + v^T A$$

$$0 = Qx + c^T + A^T v$$

$$\therefore \quad Qx + A^T v = -c \quad \text{——} \quad (1)$$

Now derivating with respect to $v$, we get

$$\frac{\partial L}{\partial v} = Ax - b$$

$$0 = Ax - b$$

$$\therefore \quad Ax - b = 0 \quad \text{——} \quad (2)$$

Adding (1) and (2) as Matrix form, we get

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}$$

2 a) Let the objective function given as below

$$f(x_k) = \min_{w,b,e} p(w,e) = \frac{1}{2}w^Tw + \frac{\gamma}{2}\sum_{k=1}^{N} e_k^2$$

The constraint is given as

$$g(x_k) = w^T\phi(x_k) + b + e_k - y_k \qquad k = 1 \ldots N$$

The Lagrangian is given as

$$L(w,b,e;\alpha) = \frac{1}{2}w^Tw + \frac{\gamma}{2}\sum_{k=1}^{N} e_k^2 - \sum_{k=1}^{N}\alpha_k\{w^T\phi(x_k) + b + e_k - y_k\}$$

b) KKT 1 is given as

Deriving with respect to $w$, we get

$$\nabla_w L = w - \sum_{k=1}^{N}\alpha_k\phi(x_k)$$

$$0 = w - \sum_{k=1}^{N}\alpha_k\phi(x_k)$$

$$\therefore w = \sum_{k=1}^{N}\alpha_k\phi(x_k) \quad \text{———} \quad (1)$$

Now deriving with respect to $b$, we get

$$\frac{\partial L}{\partial b} = -\sum_{k=1}^{N}\alpha_k$$

$$0 = -\sum_{k=1}^{N}\alpha_k$$

$$\therefore \sum_{k=1}^{N}\alpha_k = 0 \quad \text{———} \quad (2)$$

Derivating with respect to $e_K$, we get

$$\frac{\partial L}{\partial e_K} = \gamma e_K - \alpha_K$$

$$0 = \gamma e_K - \alpha_K$$

$$\therefore \alpha_K = \gamma e_K \quad \text{--- (3)} \quad \text{for } K = 1 \ldots N$$

K.K.T 2

Derivating with respect to $\alpha_K$, we get

$$\frac{\partial L}{\partial \alpha_K} = w^T \phi(x_K) + b + e_K - y_K$$

$$\therefore w^T \phi(x_K) + b + e_K - y_K = 0 \quad \text{--- (4)} \quad \text{for } k = 1 \ldots N$$

(c) substituting (1) and (3) in (4) we get

$$\sum_{l=1}^{N} \alpha_l \phi^T(x_l)\, \phi(x_K) + \underset{1}{b} + \frac{\alpha_K}{\gamma} = y_K \quad \text{for } K = 1 \ldots N$$

for $K = 1$, we get

$$\alpha_1 \phi^T(x_1) \phi(x_1) + \alpha_2 \phi^T(x_2) \phi(x_1) + \cdots + \alpha_N \phi^T(x_N) \phi(x_1) + \frac{\alpha_1}{\gamma} + b = y_1$$

for $K = 2$, we get

$$\alpha_1 \phi^T(x_1) \phi(x_2) + \alpha_2 \phi^T(x_2) \phi(x_2) + \cdots + \alpha_N \phi^T(x_N) \phi(x_2) + \frac{\alpha_2}{\gamma} + b = y_2$$

Similarly for $k = N$, we get

$$\alpha_1 \phi^T(x_1)\phi(x_N) + \alpha_2 \phi^T(x_2)\phi(x_N) + \cdots + \alpha_N \phi^T(x_N)\phi(x_N) + \frac{\alpha_N}{\gamma} + b = y_N$$

Let us consider Kernel Matrix $K$ given as

$$K = \begin{bmatrix} \phi^T(x_1)\phi(x_1) & \cdots & \phi^T(x_N)\phi(x_1) \\ \vdots & & \vdots \\ \phi^T(x_1)\phi(x_N) & & \phi^T(x_N)\phi(x_N) \end{bmatrix}_{N \times N}$$

$$\alpha = [\alpha_1, \alpha_2 \cdots \alpha_N]_{1 \times N}$$

$$1_v = [1, 1 \cdots 1]_{1 \times N}$$

and $Y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}_{N \times 1}$

So we can arrange $N$ equations in below form

$$\cancel{\alpha K^T + b 1_v^T + \frac{1}{\gamma}\alpha = Y}$$

$$\alpha K^T + b 1_v^T + \frac{1}{\gamma}\alpha = Y$$

Now Arranging the equations in Matrix form

$$\begin{bmatrix} 1_v & 0 \\ K^T + \frac{1}{\gamma} I_N & 1_v \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}$$

Also we can infer from the above equation, that it is a regression problem as

$$Y = X\theta \qquad \text{where } \theta = \begin{bmatrix} \alpha \\ b \end{bmatrix}$$

$$\theta = X^{-1} Y$$

i.e

$$\begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} 1_v & 0 \\ K^T + \frac{1}{\gamma} I_N & 1_v \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ y \end{bmatrix}$$

$$X = \begin{bmatrix} 1_v & 0 \\ K^T + \frac{1}{\gamma} I_N & 1_v \end{bmatrix}$$

$$Y = \begin{bmatrix} 0 \\ y \end{bmatrix}$$

3. Consider the following optimization problem:

$$\min_{w,b,e} P(w,e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{k=1}^{N} e_k^2 \quad (7)$$

$$\text{s.t} \quad y_k [w^T \phi(x_k) + b] = 1 - e_k, \quad k = 1 \ldots N$$

where $y_k \in \{-1, 1\}$ is the response (target) variable. Then conduct an analysis of problem (7) by applying the steps (a) to (d), explain how the new problem (7) is related to classification problem. Finally, compare KKT Matrix system obtained for this case with that of problem (6).

Solution:-

a) Let the objective function given by

$$f(x) = \min_{w,b,e} P(w,e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{k=1}^{N} e_k^2$$

and constraint given by

$$g(x) = y_k [w^T \phi(x_k) + b] - 1 + e_k$$

The Lagrangian is given by

$$L(w, b, e; \alpha) = f(x) - \sum_{k=1}^{N} \alpha_k g(x)$$

$$= \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{k=1}^{N} e_k^2 - \sum_{k=1}^{N} \alpha_k \{ y_k [w^T \phi(x_k) + b] - 1 + e_k \}$$

b) KKT 1

Derivating with respect to w, we get

$$\nabla_w L = w - \sum_{k=1}^{N} \alpha_k y_k \phi(x_k)$$

$$0 =$$

$$\therefore w = \sum_{k=1}^{N} \alpha_k y_k \phi(x_k) \quad \text{———— (1)}$$

Derivating with respect to $e_k$, we get

$$\frac{\partial L}{\partial e_k} = \gamma e_k - \alpha_k$$

$$\therefore 0 = \gamma e_k - \alpha_k$$

$$\alpha_k = \gamma e_k \quad \text{———— (2)} \quad \text{for } k = 1 \dots N$$

Derivating with respect to b, we get

$$\frac{\partial L}{\partial b} = \sum_{k=1}^{N} \alpha_k y_k$$

$$\sum_{k=1}^{N} \alpha_k y_k = 0 \quad \text{———— (3)}$$

KKT 2

Derivating with respect to $\alpha_k$, we get

$$\frac{\partial L}{\partial \alpha_k} = y_k [w^T \phi(x_k) + b] + e_k - 1$$

$$0 = y_k [w^T \phi(x_k) + b] + e_k - 1$$

$$\therefore y_k [w^T \phi(x_k) + b] + e_k = 1 \quad \text{———— (4)}$$

for $k = 1 \dots N$

c) Substituting (1) and (2) in (4) we get

$$y_k \left[ \sum_{l=1}^{N} \alpha_l y_l \phi(x_l) \phi(x_k) + b \right] + \frac{1}{\gamma} \alpha_k = 1 \quad \text{for } k = 1 \text{ -- } N$$

for $k = 1$, we get

$$y_1 \alpha_1 y_1 \phi(x_1) \phi(x_1) + \text{----} + y_1 \alpha_N y_N \phi(x_N) \phi(x_1) + b y_1 + \frac{1}{\gamma} \alpha_1 = 1$$

for $k = 2$, we get

$$y_2 \alpha_1 y_1 \phi(x_1) \phi(x_2) + \text{-----} + y_2 \alpha_N y_N \phi(x_N) \phi(x_2) + b y_2 + \frac{1}{\gamma} \alpha_2 = 1$$

$\vdots$

for $k = N$, we get

$$y_N \alpha_1 y_1 \phi(x_1) \phi(x_N) + \text{-----} + y_N \alpha_N y_N \phi(x_N) \phi(x_N) + b y_N + \frac{1}{\gamma} \alpha_N = 1$$

Let us consider $\Omega$ be kernel Matrix given by

$$\Omega = \begin{bmatrix} y_1 y_1 \phi(x_1) \phi(x_1) & \text{----} & y_1 y_N \phi(x_N) \phi(x_1) \\ \vdots & & \\ y_N y_1 \phi(x_1) \phi(x_N) & \text{----} & y_N y_N \phi(x_N) \phi(x_N) \end{bmatrix}_{N \times N}$$

$$\alpha = [\alpha_1, \alpha_2 \cdots \alpha_N]_{1 \times N}$$

~~$y = [y_1, y_2, \ldots, y_N]$ and $1_v =$~~

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}_{N \times 1} \quad \text{and} \quad 1_v = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{N \times 1}$$

So we can arrange N equations as below

$$\alpha \Omega^T + bY + \frac{1}{\gamma} \alpha I = 1_v \quad \text{———} \quad (5)$$

Now Arranging eq (3) and (5) in Matrix form, we get

$$\begin{bmatrix} \Omega^T + \frac{1}{\gamma} I & Y \\ Y & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} 1_v \\ 0 \end{bmatrix}$$

prove It is a classification problem

Eq (5) resolves as below

$$\left(\Omega^T + \frac{1}{\gamma} I\right) \alpha = 1_v - bY$$

$$\left(\Omega^T + \frac{1}{\gamma} I\right) \alpha = \{1-b, 1+b\} 1_v$$

Since $y \in \{-1, 1\}$

$\therefore$ we have binary classification of 2 possible values of $(1-b, 1+b)$

The regression problem gives below Matrix form

$$\begin{bmatrix} 1_v & 0 \\ K^T + \frac{1}{\gamma} I & 1_v \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix}$$

and the classification problem results in below Matrix

$$\begin{bmatrix} Y & 0 \\ \Omega^T + \frac{1}{\gamma} I & Y \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 1_v \end{bmatrix}$$

comparing above two Matrix equations, we see Y is replaced by $1_v$ from regression to classification.

4

(a)

Tikhonov:

Let $f(x) = \min_{w,b,e_i} J_T(w, e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{k=1}^{N} e_k^2$

$g(x) = w^T \phi(x_k) + b + e_k - y_k \quad k = 1 \dots N$

$\therefore$ The Lagrangian is given by

$$L_T(w, b, e; \alpha) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{k=1}^{N} e_k^2 - \sum_{k=1}^{N} \alpha_k (w^T \phi(x_k) + b + e_k - y_k)$$

Mozorov:

Let $f(x) = \frac{1}{2} w^T w$

$g(x) = w^T \phi(x_k) + b + e_k - y_k \quad k = 1 \dots N$

$l(x) = N\sigma^2 - \sum_{k=1}^{N} e_k^2$

$$L_M(w, b, e; \alpha, \xi) = \frac{1}{2} w^T w + \sum_{k=1}^{N} \alpha_k (w^T \phi(x_k) + b + e_k - y_k) + \xi (N\sigma^2 - \sum_{k=1}^{N} e_k^2)$$

Ivanov:

Let $f(x) = \frac{1}{2} e^T e$

$g(x) = w^T \phi(x_k) + b + e_k - y_k \quad k = 1 \dots N$

$l(x) = \pi^2 - w^T w$

$$L_I(w, b, e; \alpha, \xi) = \frac{1}{2} e^T e - \sum_{k=1}^{N} \alpha_k (w^T \phi(x_k) + b + e_k - y_k) - \xi (\pi^2 - w^T w)$$

Tikhonov conditions

KKT 1

$$\frac{\partial L}{\partial w} = w - \sum_{k=1}^{N} \alpha_k \phi(x_k)$$

$$\therefore w = \sum_{k=1}^{N} \alpha_k \phi(x_k) \qquad \because \frac{\partial L}{\partial w} = 0$$

$$\frac{\partial L}{\partial b} = -\sum_{k=1}^{N} \alpha_k$$

$$\therefore \sum_{k=1}^{N} \alpha_k = 0 \qquad \because \frac{\partial L}{\partial b} = 0 \quad \text{for } k = 1 \dots N$$

$$\frac{\partial L}{\partial e_k} = \gamma e_k - \alpha_k$$

$$\therefore \gamma e_k = \alpha_k \qquad \because \frac{\partial L}{\partial e_k} = 0 \quad \text{for } k = 1 \dots N$$

KKT 2

$$\frac{\partial L}{\partial \alpha_k} = w^T \phi(x_k) + b + e_k - y_k$$

$$\therefore w^T \phi(x_k) + b + e_k = y_k \quad -(1) \qquad \because \frac{\partial L}{\partial \alpha_k} = 0 \quad \text{for } k = 1 \dots N$$

Morozov conditions

KKT 1

$$\frac{\partial L}{\partial w} = w - \sum_{k=1}^{N} \alpha_k \phi(x_k)$$

$$\therefore w = \sum_{k=1}^{N} \alpha_k \phi(x_k) \qquad \because \frac{\partial L}{\partial w} = 0$$

$$\frac{\partial L}{\partial b} = -\sum_{k=1}^{N} \alpha_k$$

$$\therefore \sum_{k=1}^{N} \alpha_k = 0 \qquad \because \frac{\partial L}{\partial b} = 0$$

$$\frac{\partial L}{\partial e_k} = \alpha_k - 2\xi e_k$$

$\therefore \quad \alpha_k = 2\xi e_k \qquad \because \frac{\partial L}{\partial e_k} = 0$

KKT2

$$\frac{\partial L}{\partial \alpha_k} = w^T \phi(x_k) + b + e_k - y_k$$

$\therefore \quad w^T \phi(x_k) + b + e_k = y_k \quad \text{—(2)} \qquad k = 1 \cdots N$

$$\frac{\partial L}{\partial \xi} = N\sigma^2 - \sum_{k=1}^{N} e_k^2$$

$\therefore \quad N\sigma^2 = \sum_{k=1}^{N} e_k^2 \qquad \because \frac{\partial L}{\partial \xi} = 0$

Ivanov conditions

KKT1

$$\frac{\partial L}{\partial w} = -\sum_{k=1}^{N} \alpha_k \phi(x_k) + 2\xi w$$

$2\xi w = \sum_{k=1}^{N} \alpha_k \phi(x_k) \qquad \because \frac{\partial L}{\partial w} = 0$

$$\frac{\partial L}{\partial b} = -\sum_{k=1}^{N} \alpha_k$$

$\therefore \quad \sum_{k=1}^{N} \alpha_k = 0 \qquad \because \frac{\partial L}{\partial b} = 0$

$$\frac{\partial L}{\partial e_k} = e_k - \alpha_k$$

$\therefore \quad e_k = \alpha_k \qquad \because \frac{\partial L}{\partial e_k} = 0$

KKT 2

$$\frac{\partial L}{\partial \alpha_K} = w^T\varphi(x_K) + b + e_K - y_K$$

$$\therefore w^T\phi(x_K) + b + e_K = y_K \quad —(3) \quad \because \frac{\partial L}{\partial \alpha_K} = 0$$

$$\frac{\partial L}{\partial \xi} = \Pi^2 - w^T w$$

$$\therefore w^T w = \Pi^2 \quad \because \frac{\partial L}{\partial \xi} = 0$$

(b) Tikhonov :

Substituting the value of $w$ and $e_K$ in (1), we get

$$\sum_{l=1}^{N} \alpha_l \phi(x_l).\phi(x_K) + b + \frac{1}{\gamma}\alpha_K = y_K \quad \text{for } k = 1 \ldots N$$

for $k=1$, we get

$$\alpha_1\phi(x_1)\phi(x_1) + \alpha_2\phi(x_2)\phi(x_1) + \cdots + \alpha_N\phi(x_N)\phi(x_1) + b + \frac{1}{\gamma}\alpha_1 = y_1$$

for $k=2$, we get

$$\alpha_1\phi(x_1)\phi(x_2) + \alpha_2\phi(x_2).\phi(x_2) + \cdots + \alpha_N\phi(x_N)\phi(x_2) + b + \frac{1}{\gamma}\alpha_2 = y_2$$

$\vdots$

for $k=N$, we get

$$\alpha_1\phi(x_1)\phi(x_N) + \alpha_2\phi(x_2)\phi(x_N) + \cdots + \alpha_N\phi(x_N)\phi(x_N) + b + \frac{1}{\gamma}\alpha_N = y_N$$

We know $\Omega_{KT} = K(x_K, x_t) = \phi(x_K)^T \phi(x_t)$

So we can have N equations as below

$\cancel{\alpha\Omega + bI_N + \frac{1}{\gamma}I_N}$

$$\alpha\Omega + b1_v + \frac{1}{\gamma} I_N \alpha = Y$$

Where $Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix}$

So we arrange the above equation

and equation of $\sum_{K=1}^{N} \alpha_K = 0$ in Matrix form

$$\begin{bmatrix} 0 & 1_N^T \\ 1_N & \Omega + \frac{1}{\gamma} I_N \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}$$

(c) <u>Mozorov</u>

Substituting the value of $w$ in eq (2), we get

$$\sum_{k=1}^{N} \alpha_l \phi(x_l) \phi(x_k) + b + \frac{1}{2\zeta} \alpha_K = y_K \quad \text{for } k = 1 \cdots N$$

for $K=1$

$\cancel{\alpha_1 \phi(x_1) \phi(x_1) + b + \frac{1}{2\zeta}\alpha_1 = y_1}$

$$\alpha_1 \phi(x_1) \phi(x_1) + \alpha_2 \phi(x_2) \phi(x_1) + \cdots + \alpha_N \phi(x_N)$$

$$+ b + \frac{1}{2\zeta} \alpha_K = y_1$$

for $k=2$, we get

$$\alpha_1 \phi(x_1)\phi(x_2) + \alpha_2 \phi(x_2)\phi(x_2) + \cdots + \alpha_N \phi(x_N)\phi(x_2) + b + \frac{1}{2\xi}\alpha_2 = y_2$$

$\vdots$

Similarly for $k=N$, we get

$$\alpha_1 \phi(x_1)\phi(x_N) + \alpha_2 \phi(x_2)\phi(x_N) + \cdots + \alpha_N \phi(x_N)\phi(x_N) + b + \frac{1}{2\xi}\alpha_N = y_N$$

Substituting $\Omega$ in equation, we get

$$\alpha\Omega + b1_N + \frac{1}{2\xi} I_N \alpha = y$$

Also we have $\sum_{k=1}^{N} \alpha_k = 0$

Arranging above 2 equations, we get

$$\begin{bmatrix} 0 & 1_N^T \\ 1_N & \frac{1}{2\xi} I_N + \Omega \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}$$

s.t $\quad N\sigma^2 = \sum_{i=1}^{N} e_i^2$

$$N\sigma^2 = \frac{1}{4\xi^2} \sum_{i=1}^{N} \alpha_i^2 \qquad \therefore \; e_i = \frac{1}{2\xi}\alpha_i$$

Let us $\sigma' = 2\xi\sigma \quad \because \xi$ is constant

$\therefore$ we get $N\sigma'^2 = \sum_{i=1}^{N} \alpha_i^2$

s.t $\quad N\sigma'^2 = \alpha^T \alpha \qquad$ where $\alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}_{N\times 1}$

d) <u>Ivanov</u>

Substituting the value of $w$ and $e_K$ in (3) we get

$$\frac{1}{2\xi}\sum_{l=1}^{N}\alpha_l\phi(x_l)\phi(x_K) + b + \cancel{t}\alpha_K = y_K \qquad K=1\ldots N$$

for $K=1$, we get

$$\frac{1}{2\xi}\left[\alpha_1\phi(x_1)\phi(x_1) + \alpha_2\phi(x_2)\phi(x_1) + \cdots + \alpha_N\phi(x_N)\phi(x_1)\right] + b + \cancel{t}\alpha_1 = y_1$$

$\vdots$

for $K=N$, we get

$$\frac{1}{2\xi}\left[\alpha_1\phi(x_1)\phi(x_N) + \alpha_2\phi(x_2)\phi(x_N) + \cdots + \alpha_N\phi(x_N)\phi(x_N)\right] + b + \cancel{t}\alpha_N = y_N$$

Substituting $\Omega_{KT} = k(x_K, x_T) = \phi(x_K)^T\phi(x_T)$, we can put

N equations as

$$\frac{1}{2\xi}\Omega\alpha + b1_N + \cancel{t}I_N\alpha = Y \qquad \text{where } Y = \begin{bmatrix} y_1 \\ \vdots \\ \vdots \\ y_N \end{bmatrix}_{N\times 1}$$

and we have other equation as

$$\sum_{K=1}^{N}\alpha_K = 0$$

and $\alpha = [\alpha_1 \cdots \alpha_N]$

Arranging above 2 equations as Matrix form we get

$$\begin{bmatrix} 0 & 1_N^T \\ 1_N & \frac{1}{2\xi}\Omega + \cancel{t}I_N \end{bmatrix}\begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix}$$

s.t $\cancel{t}^2 = w^Tw$

we know $w^T w = \sum_{l=1}^{N} \alpha_l \phi^T(x_l) \sum_{k=1}^{N} \alpha_k \phi(x_k)$

$= \sum_{l,k=1}^{N} \alpha_l \phi^T(x_l) \phi(x_k) \alpha_k$

$= \alpha^T \Omega \alpha$ where $\Omega = K(x_l, x_k)$

and $\alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}$

$\therefore$ we have

$$\begin{bmatrix} 0 & 1_N^T \\ 1_N & \frac{1}{2\xi}\Omega + I_N \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}$$

s.t $\pi^2 = \alpha^T \Omega \alpha$

5

a) Let the objective function be

$$f(x) = \min_{w,b,\xi} P(w, \xi) = \frac{1}{2} w^T w + c \sum_{k=1}^{N} \xi_k$$

and constraints given as

$$g(x) = y_k \left[w^T \phi(x_k) + b\right] - 1 + \xi_k$$

$$l(x) = \xi_k$$

Since $y_k \in \{-1, 1\}$

when $w^T \phi(x_k) + b < 0$ $\quad y_k = -1$

and $w^T \phi(x_k) + b > 0$ $\quad y_k = 1$

hence It is a classification problem, as output is binary $\{-1, 1\}$

(b) Let the objective function be

$$f(x) = \min_{w,b,\xi} P(w,\xi) = \frac{1}{2} w^T w + c \sum_{k=1}^{N} \xi_k$$

and constraint be given as

$$g(x) = y_k [w^T \phi(x_k) + b] - 1 + \xi_k \geq 0$$

$$l(x) = \xi_k \geq 0 \qquad k = 1, \dots, N$$

The Lagrangian would be given as

$$L(w, b; \alpha) = \frac{1}{2} w^T w + c \sum_{k=1}^{N} \xi_k + \sum_{k=1}^{N} \alpha_k \{ y_k [w^T \phi(x_k) + b] - 1 + \xi_k \} - \sum_{k=1}^{N} v_k \xi_k$$

Deriving with respect to w, we get

$$\frac{\partial L}{\partial w} = w - \sum_{k=1}^{N} \alpha_k y_k \phi(x_k)$$

$$\therefore w = \sum_{k=1}^{N} \alpha_k y_k \phi(x_k) \qquad \because \frac{\partial L}{\partial w} = 0$$

Derivating with respect to b, we get

$$\frac{\partial L}{\partial b} = -\sum_{k=1}^{N} \alpha_k y_k$$

$$\therefore \sum_{k=1}^{N} \alpha_k y_k = 0 \qquad \because \frac{\partial L}{\partial b} = 0$$

Derivating with respect to $\xi_k$, we get

$$\frac{\partial L}{\partial \xi_k} = c - \alpha_k - \gamma_k$$

$$\therefore \gamma_k = c - \alpha_k \qquad \because \frac{\partial L}{\partial \xi_k} = 0$$

Substituting the above derived values in Lagrangian equation we get

(c)

$$D(\alpha) = \frac{1}{2} \sum_{k=1}^{N} \alpha_k y_k \phi^T(x_k) \sum_{l=1}^{N} \alpha_l y_l \phi(x_l) + c \sum_{k=1}^{N} \xi_k$$
$$- \sum_{k=1}^{N} \alpha_k \left\{ y_k \left( \sum_{l=1}^{N} \alpha_l y_l \phi^T(x_l) \phi(x_k) + b \right) - 1 + \xi_k \right\} - \sum_{k=1}^{N} (c - \alpha_k) \xi_k$$

$$= \frac{1}{2} \sum_{k,l=1}^{N} \alpha_k \alpha_l y_k y_l \phi^T(x_k) \phi(x_l) + \cancel{c \sum_{k=1}^{N} \xi_k}$$
$$- \sum_{k,l=1}^{N} \alpha_k \alpha_l y_k y_l \phi^T(x_k) \phi(x_l) - b \sum_{k=1}^{N} \alpha_k y_k$$
$$+ \sum_{k=1}^{N} \alpha_k - \cancel{\sum_{k=1}^{N} \alpha_k \xi_k} - \cancel{c \sum_{k=1}^{N} \xi_k} + \cancel{\sum_{k=1}^{N} \alpha_k \xi_k}$$

$$D(\alpha) = \frac{-1}{2} \sum_{l,k=1}^{N} \alpha_l \alpha_k y_l y_k \phi^T(x_l)\phi(x_k) + \sum_{k=1}^{N} \alpha_k$$

$$\text{s.t} \quad \sum_{k=1}^{N} \alpha_k y_k = 0 \qquad k = 1, \ldots, N$$

(d) we have $D(\alpha) = -\frac{1}{2} \sum_{l,k=1}^{N} \alpha_l \alpha_k y_l y_k \phi^T(x_l)\phi(x_k) + \sum_{k=1}^{N} \alpha_k$

Let $K(x_k, x_l) = \phi^T(x_k)\phi(x_l)$ is Quadratic as it is multiplication of two terms

$\therefore$ we get $D(\alpha) = \frac{-1}{2} \sum_{l,k=1}^{N} \alpha_k \alpha_l y_k y_l K(x_k, x_l) + \sum_{k=1}^{N} \alpha_k$

which becomes a Quadratic equation.

(e) <u>KKT1</u> (stationary condition)

$$D(\alpha) = \frac{-1}{2} \sum_{l,k=1}^{N} \alpha_l \alpha_k y_l y_k \phi^T(x_l)\phi(x_k) + \sum_{k=1}^{N} \alpha_k$$

$$\text{s.t} \quad \sum_{k=1}^{N} \alpha_k y_k = 0$$

<u>KKT2</u> (primal feasibility condition)

$$y_k [w^T \phi(x_k) + b] \geq 1 - \xi_k \qquad k = 1, \ldots, N$$

$$\xi_k \geq 0 \qquad k = 1, \ldots, N$$

KKT3 (Dual feasibility condition)

$$0 \le \alpha_k \le C \quad , \quad k = 1, \dots, N$$

$$\nu_k \geq 0 \quad k = 1, \dots, N$$

KKT4 (complementary slackness condition)

$$\alpha_k \left[ y_k \{ w^T \phi(x_k) + b \} - 1 + \xi_k \right] = 0 \quad k = 1, \dots, N$$

$$\& \ \nu_k \xi_k = 0 \quad k = 1, \dots, N$$

(f)

When $\alpha_k = 0$

$$y_k \{ w^T \phi(x_k) + b \} - 1 + \xi_k \geq 0$$

$$\nu_k = C \quad \therefore \xi_k = 0$$

$$\therefore y_k \{ w^T \phi(x_k) + b \} \geq 1$$

here $x_k$ does not become support vector and It is correctly classified.

when $\alpha_K = C$

$$y_K \{ w^T \phi(x_K) + b \} - 1 + \xi_K = 0$$

$$V_K = 0 \qquad \therefore \xi_K \geq 0$$

$$\therefore y_K \{ w^T \phi(x_K) + b \} = 1 - \xi_K \qquad \xi_K \geq 0$$

$$\therefore y_K \{ w^T \phi(x_K) + b \} \leq 1$$

$x_K$ is support vector when $0 < \xi_K < 1$ and It is bounded, however when $\xi_K \geq 1$ It is wrongly clanifred.

when $0 < \alpha_K < C$

$$y_K \{ w^T \phi(x_K) + b \} - 1 + \xi_K = 0$$

$$V_K > 0 \qquad \therefore \xi_K = 0$$

$$\therefore y_K \{ w^T \phi(x_K) + b \} = 1$$

$x_K$ is support vector not bounded by $\xi_K$ and is clanified good.

5(g)

we have $y_K [w^T \phi(x_k) + b] = 1$ when $0 < \alpha_K < C$

$\therefore$ we get $b = \frac{1}{y_K} - w^T \phi(x_K)$, $K \in U$

$U$ is subset of $I = \{1, 2, \cdots N\}$ which is not bounded

Since $y_k \in \{-1, 1\}$, so we get mean of $|\bar{b}|$ as

$$\bar{b} = \frac{1}{|U|} \sum_{K=1}^{N} (y_K - w^T \phi(x_K))$$

$$\therefore D(x) = \sum_{K \in U} \alpha_K y_K \phi(x_K)^T \phi(x_K) + \bar{b}$$

5(h) ~~Let $K(x_k, x_l) = \phi(x_k)^T \phi(x_l)$~~

(i) Let $K(x_K, x_l) = \phi(x_K)^T \phi(x_l)$ for $k, l = 1, \cdots, N$

we have $D(\alpha) = \sum_{k,l=1}^{N} \alpha_K \alpha_l y_K y_l \phi(x_K)^T \phi(x_l) + \sum_{K=1}^{N} \alpha_K$

$$D(\alpha) = \sum_{K,l=1}^{N} \alpha_K \alpha_l y_K y_l K(x_K, x_l) + \sum_{K=1}^{N} \alpha_K$$

$$\text{s.t} \quad \sum_{K=1}^{N} \alpha_K y_K = 0$$

(ii) we have value of $w$ as

$$w = \sum_{k=1}^{N} \alpha_k y_k \phi(x_k)$$

and we have $D(\alpha) = \sum_{k,l=1}^{N} \alpha_k \alpha_l y_k y_l K(x_k, x_l) + \sum_{k=1}^{N} \alpha_k$

$$\text{s.t} \quad \sum_{k=1}^{N} \alpha_k y_k = 0$$

for $\alpha_k = 0$ we have $\xi_k = 0$

~~and $0 < \alpha_k < C$ we have $\xi_k = 0$~~

So for certain values of $\alpha_k$, we get $\xi_k = 0$ hence we get Sparseness

(iii) when $\alpha_k > 0$

we get $\xi_k = 0$ for $0 < \alpha_k < C$

So we get support vectors $x_k$, not bounded with solution vector $w$ and $\alpha_k$ are support values

Similiarly for $\alpha_k = C$, we get support vectors $x_k$ bounded as $0 < \xi_k < 1$ and we get only good classification if $\xi_k$ is bounded $[0,1]$

6.

Let us consider analogous form of regression from previous exercise as below

$$\min_{w,\xi} P(w,\xi) = \frac{1}{2} w^T w + c \sum_{k=1}^{N} \xi_k$$

$$\text{s.t} \quad y_k - w^T \phi(x_k) - b \leq \epsilon + \xi_k$$
$$\xi_k \geq 0$$

$\phi(\cdot) : R^N \to R^M$

Let $f(x) = \min_{w,\xi} (P(w,\xi)) = \frac{1}{2} w^T w + c \sum_{k=1}^{N} \xi_k$

and constraints be given as

$$g(x) = w^T \phi(x_k) + b - y_k + \epsilon + \xi_k \geq 0$$

and $l(x) = \xi_k \geq 0$

The Lagrangian is given as

$$L(w,b,\xi;\alpha) = \frac{1}{2} w^T w + c \sum_{k=1}^{N} \xi_k - \sum_{k=1}^{N} \alpha_k (w^T \phi(x_k) + b + y_k + \epsilon + \xi_k) - \sum_{k=1}^{N} \eta_k \xi_k$$

Derivating with respect to $w$, we get

$$\frac{\partial L}{\partial w} = w - \sum_{k=1}^{N} \alpha_k \phi(x_k)$$

$$\therefore w = \sum_{k=1}^{N} \alpha_k \phi(x_k) \qquad \because \frac{\partial L}{\partial w} = 0$$

Derivating with respect to b, we get

$$\frac{\partial L}{\partial b} = -\sum_{k=1}^{N} \alpha_k \qquad \because \frac{\partial L}{\partial b} = 0$$

$$\therefore \sum_{k=1}^{N} \alpha_k = 0$$

Derivating with respect to $\xi_k$, we get

$$\frac{\partial L}{\partial \xi_k} = c - \alpha_k - \eta_k \qquad \because \frac{\partial L}{\partial \xi_k} = 0$$

$$\therefore \eta_k = c - \alpha_k$$

Now substituting the values in Lagrangian, we get the dual form as below

$$D(\alpha) = \frac{1}{2} \sum_{k=1}^{N} \alpha_k \phi^T(x_k) \sum_{l=1}^{N} \alpha_l \phi(x_l) + c \sum_{k=1}^{N} \alpha_k$$

$$- \sum_{k=1}^{N} \alpha_k \left( \left( \sum_{l=1}^{N} \alpha_l \phi^T(x_l) \right) \phi(x_k) + b + y_k + \epsilon + \xi_k \right)$$

$$- \sum_{k=1}^{N} (c - \alpha_k) \xi_k$$

$$= \frac{1}{2} \sum_{k,l=1}^{N} \alpha_k \alpha_l \phi^T(x_k) \phi(x_l) + c \cancel{\sum_{k=1}^{N} \alpha_k}$$

$$- \sum_{k,l=1}^{N} \alpha_k \alpha_l \phi^T(x_l) \phi(x_k) - b \sum_{k=1}^{N} \alpha_k + \sum_{k=1}^{N} \alpha_k y_k - \sum_{k=1}^{N} \epsilon \alpha_k$$

$$- \cancel{\sum_{k=1}^{N} \alpha_k \xi_k} - \cancel{c \sum_{k=1}^{N} \xi_k} + \cancel{\sum_{k=1}^{N} \alpha_k \xi_k}$$

## KKT 1

$$D(\alpha) = -\frac{1}{2}\sum_{k,l=1}^{N} \alpha_k \alpha_l \phi^T(x_k)\phi(x_l) + \sum_{k=1}^{N}\alpha_k y_k - \sum_{k=1}^{N}\alpha_k \epsilon$$

$$D(\alpha) = -\frac{1}{2}\sum_{k,l=1}^{N} \alpha_k \alpha_l \phi^T(x_k)\phi(x_l) + \sum_{k=1}^{N}\alpha_k (y_k - \epsilon)$$

$$\text{s.t} \quad \sum_{k=1}^{N}\alpha_k = 0$$

## KKT 2

$$y_k - w^T\phi(x_k) - b \leq \epsilon + \xi_k$$

$$\xi_k \geq 0$$

## KKT 3

$$0 \leq \alpha_k \leq C$$

$$\eta_k \geq 0$$

## KKT 4

$$\alpha_k\left[w^T\phi(x_k) + b - y_k + \epsilon + \xi_k\right] = 0 \qquad k = 1, \ldots, N$$

$$\eta_k \xi_k = 0$$

7.

Let the objective function be

$$f(x) = \min_{w,b,\xi} P(w,\xi) = \frac{1}{2} w^T w + c\left\{ v\epsilon + \frac{1}{N}\sum_{k=1}^{N}(\xi_k + \xi_k^*)\right)$$

and constraints

$$g(x) = w^T \phi(x_k) + b - y_k + \epsilon + \xi_k \geq 0$$

$$l(x) = \epsilon + \xi_k^* + y_k - w^T\phi(x_k) \cancel{+} - b \geq 0$$

$$n(x) = \xi_k \geq 0$$

$$m(x) = \xi_k^* \geq 0$$

The Lagrangian is given by

$$L(w,b,\xi;\alpha) = \frac{1}{2} w^T w + c\left\{ v\epsilon + \frac{1}{N}\sum_{k=1}^{N}(\xi_k + \xi_k^*)\right\}$$

$$- \sum_{k=1}^{N} \alpha_k (w^T\phi(x_k) + b - y_k + \epsilon + \xi_k)$$

$$- \sum_{k=1}^{N} \eta_k \xi_k$$

$$- \sum_{k=1}^{N} \alpha_k^* (-w^T\phi(x_k) - b + \epsilon + \xi_k^* + y_k)$$

$$- \sum_{k=1}^{N} \eta_k^* \xi_k^*$$

Deriving with respect to $w$, we get

$$\frac{\partial L}{\partial w} = w - \sum_{k=1}^{N} \alpha_k \phi(x_k) + \sum_{k=1}^{N} \alpha_k^* \phi(x_k)$$

$$\therefore w = \sum_{k=1}^{N} (\alpha_k - \alpha_k^*) \phi(x_k)$$

Derivating with respect to b, we get

$$\frac{\partial L}{\partial b} = -\sum_{k=1}^{N} \alpha_k + \sum_{k=1}^{N} \alpha_k^*$$

$$\therefore \sum_{k=1}^{N} (\alpha_k - \alpha_k^*) = 0$$

Derivating with respect to $\xi_k$, we get

$$\frac{\partial L}{\partial \xi_k} = \frac{C}{N} - \alpha_k - \eta_k$$

$$\eta_k = \frac{C}{N} - \alpha_k$$

Derivating with respect to $\xi_k^*$, we get

$$\frac{\partial L}{\partial \xi_k^*} = \frac{C}{N} - \alpha_k^* - \eta_k^*$$

$$\therefore \eta_k^* = \frac{C}{N} - \alpha_k^*$$

Substituting the values to get the Dual form

$$D(\alpha, \alpha^*) = \frac{1}{2} \sum_{k=1}^{N} (\alpha_k - \alpha_k^*) \phi^T(x_k) \sum_{l=1}^{N} (\alpha_l - \alpha_l^*) \phi(x_l) + C\left(\nu\epsilon + \frac{1}{N} \sum_{k=1}^{N} (\xi_k + \xi_k^*)\right)$$

$$- \sum_{k=1}^{N} \alpha_k \left( \sum_{l=1}^{N} (\alpha_l - \alpha_l^*) \phi^T(x_l) \phi(x_k) + b - y_k + \epsilon + \xi_k \right) - \sum_{k=1}^{N} \left(\frac{C}{N} - \alpha_k\right) \xi_k$$

$$+ \sum_{k=1}^{N} \alpha_k^* \left( \sum_{l=1}^{N} (\alpha_l - \alpha_l^*) \phi^T(x_l) \phi(x_k) + b - y_k \cancel{+ \epsilon} + \xi_k^* \right) - \sum_{k=1}^{N} \left(\frac{C}{N} - \alpha_k^*\right) \xi_k^*$$

$$= \frac{-1}{2} \sum_{k,l=1}^{N} (\alpha_k - \alpha_k^*)(\alpha_l - \alpha_l^*) \phi^T(x_k) \phi(x_l) + C\nu\epsilon + \frac{C}{N} \cancel{\sum_{k=1}^{N} (\xi_k + \xi_k^*)}$$

$$- \frac{C}{N} \cancel{\sum_{k=1}^{N} (\xi_k + \xi_k^*)} + \sum_{k=1}^{N} \alpha_k y_k - \sum_{k=1}^{N} \alpha_k^* y_k + b \sum_{k=1}^{N} (\alpha_k^* - \alpha_k)$$

$$+ \epsilon \sum_{k=1}^{N} -(\alpha_k^* + \alpha_k) + \cancel{\sum_{k=1}^{N} \alpha_k^* \xi_k^*} - \cancel{\sum_{k=1}^{N} \alpha_k \xi_k} + \cancel{\sum_{k=1}^{N} \alpha_k \xi_k}$$

$$+ \cancel{\sum_{k=1}^{N} \alpha_k^* \xi_k^*}$$

KKT1

$$D(\alpha,\alpha^*) = -\frac{1}{2}\sum_{k,l=1}^{N}(\alpha_k-\alpha_k^*)(\alpha_l-\alpha_l^*)\,\phi^T(x_k)\,\phi(x_l) + C\nu\epsilon$$

$$+\sum_{k=1}^{N}(\alpha_k-\alpha_k^*)\,y_k - \epsilon\sum_{k=1}^{N}(\alpha_k+\alpha_k^*)$$

$$\text{s.t.}\quad \sum_{k=1}^{N}(\alpha_k-\alpha_k^*) = 0$$

KKT2

$$y_k - w^T\phi(x_k) - b \le \epsilon + \xi_k \qquad k=1,\dots,N$$

$$w^T\phi(x_k) + b - y_k \le \epsilon + \xi_k^*, \qquad k=1,\dots,N$$

$$\xi_k,\ \xi_k^* \geq 0$$

KKT3

$$0 \le \alpha_k \le \frac{C}{N}$$

$$\eta_k \geq 0$$

$$0 \le \alpha_k^* \le C/N$$

$$\eta_k^* \geq 0$$

KKT4

$$\alpha_k\,[w^T\phi(x_k) + b - y_k + \epsilon + \xi_k] = 0$$

$$\eta_k\xi_k = 0$$

$$\alpha_k^*\,[-w^T\phi(x_k) - b + y_k + \epsilon + \xi_k^*] = 0$$

$$\eta_k^*\xi_k^* = 0$$

8. For the classification problem the model would be

$$f(x) = \min_{w,b,\xi} P(w,\xi) = \frac{1}{2} w^T w + c\{ v\epsilon + \frac{1}{N} \sum_{k=1}^{N} \xi_k \}$$

$$\text{s.t } y_k [w^T \phi(x_k) + b] \geq 1 - \xi_k \quad k = 1, \dots, N$$

$$\xi_k \geq 0 \quad k = 1, \dots, N$$

The Lagrangian would be given as

$$L(w, b; \alpha) = \frac{1}{2} w^T w + c\{ v\epsilon + \frac{1}{N} \sum_{k=1}^{N} \xi_k \} - \sum_{k=1}^{N} \alpha_k ( y_k (w^T \phi(x_k) + b) - 1 + \xi_k ) - \sum_{k=1}^{N} \eta_k \xi_k$$

Derivating with respect to w, we get

$$\frac{\partial L}{\partial w} = w - \sum_{k=1}^{N} \alpha_k y_k \phi(x_k)$$

$$\therefore w = \sum_{k=1}^{N} \alpha_k y_k \phi(x_k)$$

Derivating with respect to b, we get

$$\frac{\partial L}{\partial b} = - \sum_{k=1}^{N} \alpha_k y_k$$

$$\therefore \sum_{k=1}^{N} \alpha_k y_k = 0$$

Derivating with respect to $\xi_K$, we get

$$\frac{\partial L}{\partial \xi_K} = \frac{C}{N} - \alpha_K - \eta_K$$

$$\therefore \eta_K = \frac{C}{N} - \alpha_K$$

Substituting above values to get Dual form

$$D(\alpha) = \frac{1}{2} \sum_{K=1}^{N} \alpha_K y_K \phi^T(x_K) \sum_{l=1}^{N} \alpha_l y_l \phi(x_l) + C\left\{\nu\epsilon + \frac{1}{N} \sum_{K=1}^{N} \xi_K \right\}$$

$$- \sum_{K=1}^{N} \alpha_K \left( y_K \left( \sum_{l=1}^{N} \alpha_l y_l \phi^T(x_l) \phi(x_K) + b \right) - 1 + \xi_K \right)$$

$$- \sum_{K=1}^{N} \left( \frac{C}{N} - \alpha_K \right) \xi_K$$

$$= -\frac{1}{2} \sum_{K,l=1}^{N} \alpha_K y_K \alpha_l y_l \phi^T(x_K) \phi(x_l) + C\nu\epsilon + \cancel{\frac{C}{N} \sum_{K=1}^{N} \xi_K}$$

$$- b \sum_{K=1}^{N} \alpha_K y_K + \sum_{K=1}^{N} \alpha_K - \cancel{\sum_{K=1}^{N} \alpha_K \xi_K} - \cancel{\frac{C}{N} \sum_{K=1}^{N} \xi_K} + \cancel{\sum_{K=1}^{N} \alpha_K \xi_K}$$

KKT1

$$D(\alpha) = -\frac{1}{2} \sum_{K,l=1}^{N} \alpha_K \alpha_l y_K y_l \phi^T(x_K) \phi(x_l) + C\nu\epsilon + \sum_{K=1}^{N} \alpha_K$$

$$\text{s.t} \quad \sum_{K=1}^{N} \alpha_K y_K = 0$$

KKT2

$$\cancel{\min_{w,b,\xi} P(w,\xi) = \frac{1}{2}w^T w + c\left\{\nu\rho + \frac{1}{N}\sum_{k=1}^{N}\xi_k\right\}}$$

$$\cancel{ST}\; y_k\left[w^T\phi(x_k) + b\right] \geq 1 - \xi_k \qquad k = 1, \cdots, N$$

$$\xi_k \geq 0 \qquad k = 1, \cdots, N$$

KKT3

$$\sum_{k=1}^{N}\alpha_k y_k = 0$$

$$0 \leq \alpha_k \leq \frac{C}{N}$$

$$\eta_k \geq 0$$

KKT4 complementary

$$\alpha_k\left\{y_k\left(w^T\phi(x_k) + b\right) - 1 + \xi_k\right\} = 0$$

$$\eta_k\xi_k = 0$$

$$\xi_k \geq 0 \qquad \eta_k = \frac{C}{N} - \alpha_k$$

when $\alpha_k = 0$

then $\eta_k = \frac{C}{N}$ and $\xi_k = 0$

Also $y_k\left(w^T\phi(x_k) + b\right) - 1 + \xi_k \geq 0$

i.e $y_k\left(w^T\phi(x_k) + b\right) \geq 1 \quad \because \xi_k = 0$

$x_k$ is not a support vector, as it is correctly classified.

when $0 \leq \alpha_K < \frac{C}{N}$

then $\alpha_K > 0$ $\therefore y_K(w^T\phi(x_K)+b) - 1 + \xi_K = 0$

$\eta_K > 0$ $\therefore \xi_K = 0$

$\therefore y_K(w^T\phi(x_K)+b) = 1$

$x_K$ is support vector not bounded and It is good clanified

when $\alpha_K = \frac{C}{N}$

then $\eta_K = 0$ $\xi_K \geq 0$

$y_K(w^T\phi(x_K)+b) = 1-\xi_K$

for $\xi_K$ $0 < \xi_K < 1$, It is clanified good and $x_K$ is support vector bounded for $\xi_K \geq 1$ It is bad clanified.