Semantic labelling of 3D Point Clouds for Indoor Scenes NIP 2011

$y_i^k \in \{0, 1\}$ ($k \to$ class k, $i \to$ segment i) $\Rightarrow$ $y_1^6 = \{0,0,0,0,0,1,0, \dots K\}$ for segment#1

output $Y = \{ y_1, y_2, y_3 \dots y_N \}$

where $y_1^3 = (0,0,1,0,0,\dots,K)$ segment #1 belongs to class#3

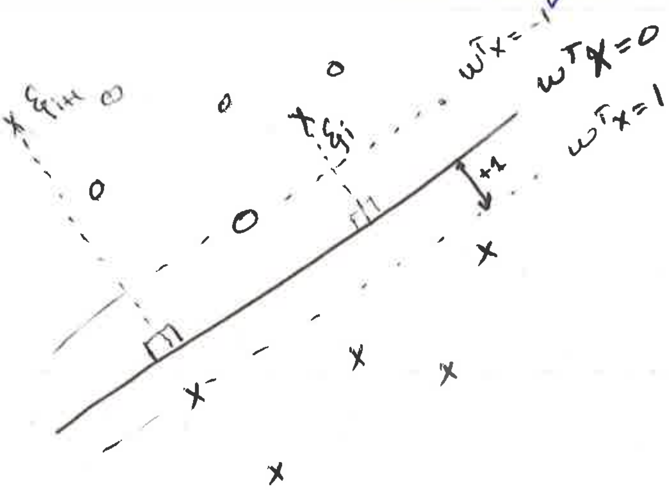
$y_N^2 = (0,1,0,0,\dots K)$ segment#2 belongs to class#2

* MIP solver for prediction with constraint $\forall i \sum_{j=1}^{K} y_i^j = 1$
(see 4 at pg 7)
* Graphcut for solving relaxed MIP

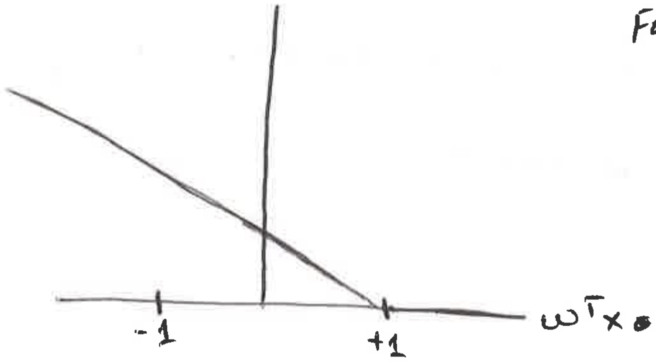
ICRA Papers (Room 7065, level 7, building 2) 12:30 to 3 PM

29/Apr/2013

~~Topic~~



For y = 1



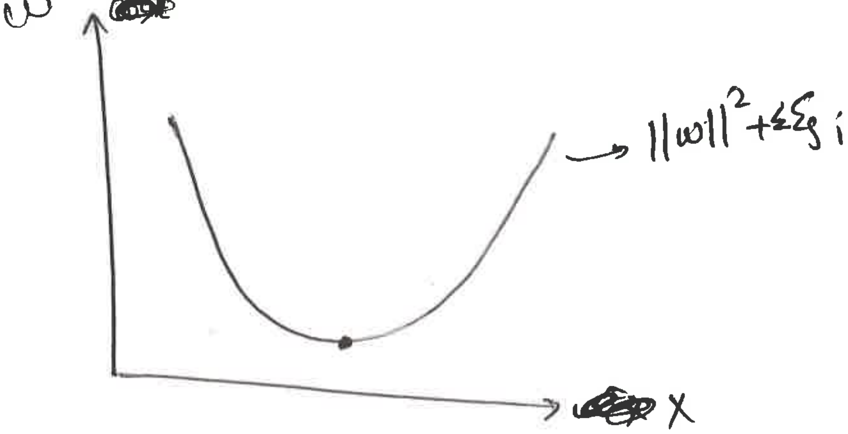
$\omega^T x$	score	loss
+1	3	0
	4	0
	5	0
		0
i	0.5	0.5
i+1	-2	1 - (-2) = 3

$\omega^T x \geq 1$

$\omega^T x \geq 1 + 0.5$

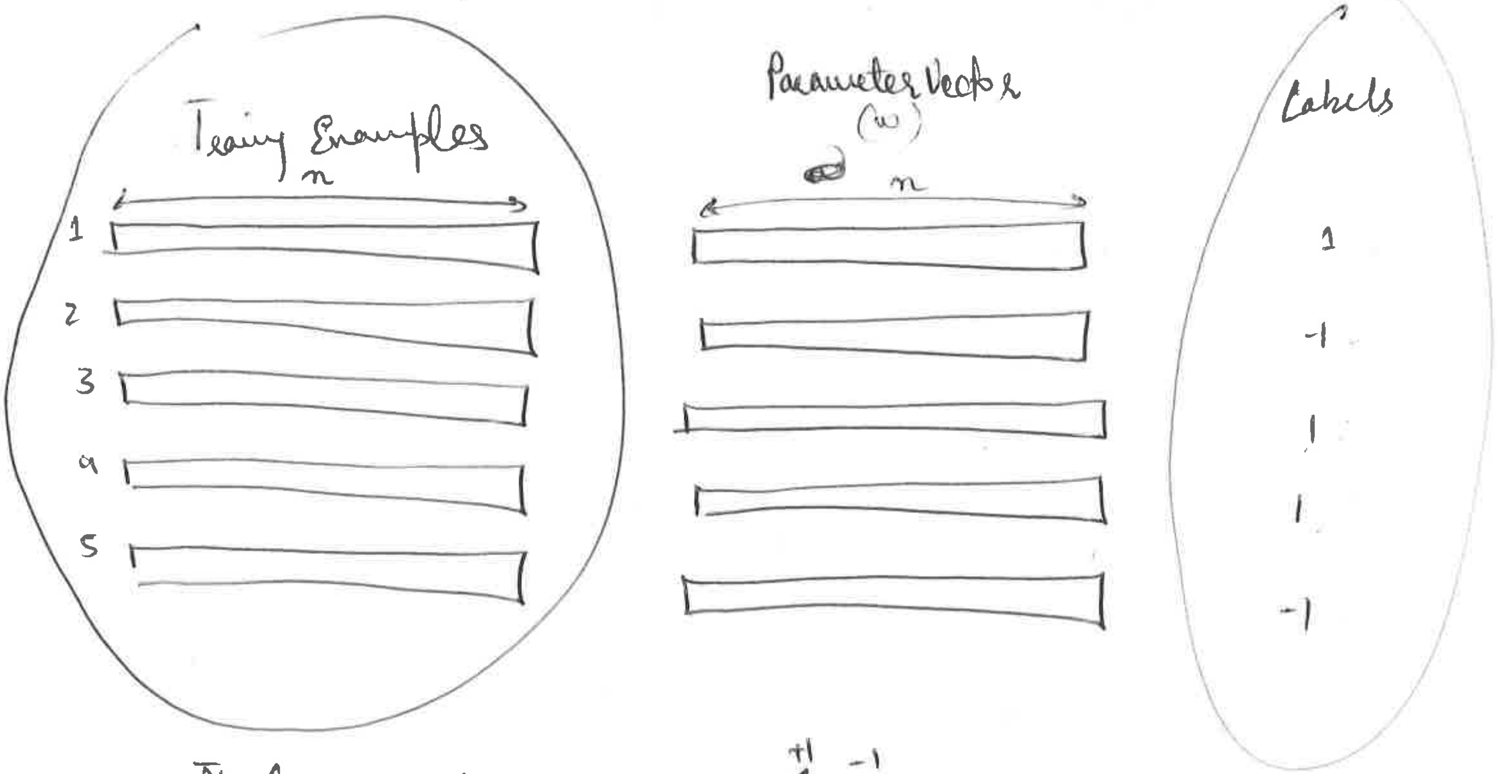
$\omega^T x \geq 0.5$

$\omega^T x_i \geq 3$



$$\bar{x} = (x_1, \dots x_n)$$
$$\bar{y} = (y_1, \dots y_n)$$

$$\underset{y' \in \bar{Y}}{\text{argmax}} \quad w^T \psi(\bar{x}, \bar{y}')$$



$$\bar{X} = (x_1, x_2, \dots x_5) \quad n=5$$

$$w = (\overset{+1}{w_1}, \overset{-1}{w_2})$$
$$w_1 \in \mathbb{R}^n$$
$$w_2 \in \mathbb{R}^n$$

$$\bar{y} = (y_1 \dots y_5) \quad n=5$$

$$\cancel{h_w(\bar{x})}$$

$$\bar{h}_w(\bar{x}) = \underset{y' \in \bar{Y}}{\text{argmax}} \left(w^T \psi(\bar{x}, (1,1,\dots 1)),\ w^T \psi(\bar{x}, (-1,-1,-1,-1,-1)) \dots \right.$$
$$\left. \text{30 more values} \right)$$

$U \rightarrow$ ground truth

$\Delta(y, U) =$ ~~loss~~ risk associated with assignment "y"

$\Delta(y=1, U=1)\, p(U=1|x)$ $\quad$ $\Delta(y=2, U=1)\, p(U=1|x)$ $\quad$ $\Delta(y=3, U=1)\, p(U=1|x)$

$\Delta(y=1, U=2)\, p(U=2|x)$ $\quad$ $\Delta(y=2, U=2)\, p(U=2|x)$ $\quad$ $\Delta(y=3, U=2)\, p(U=2|x)$

$\Delta(y=1, U=3)\, p(U=3|x)$ $\quad$ $\Delta(y=2, U=3)\, p(U=3|x)$ $\quad$ $\Delta(y=3, U=3)\, p(U=3|x)$

y (predicted)

U (True) \ y	1	2	3
1			
2			
3			

Loss $\Delta(\overset{\text{pred}}{y}, \overset{\text{true}}{U})$ $\longrightarrow$ loss of farther classes will be high

U \ y	1	2	3
1	0	1	2
2	1	0	1
3	2	1	0

MAP

$P(Y=1|X) = \frac{1}{1+e^{-\theta^T x}}$

$P(Y=2|X) = \frac{1}{1+e^{-\theta^T x}}$

$P(Y=3|X) = \frac{1}{1+e^{-\theta^T x}}$

$$y^* = \arg\min_y \sum_U \Delta(y, U)\, p(U|x)$$

$$y^* = \arg\min_y \begin{bmatrix} \Delta(y=1, U=1) P(U=1|X) + \Delta(1,2) P(2|X) + \ldots \\ \Delta(2,1) P(1|X) + \Delta(2,2) P(2|X) + \Delta(2,3) P(3|X), \\ \Delta(3,1) P(1|X) + \Delta(3,2) P(2|X) + \Delta(3,3) P(3|X) \end{bmatrix}$$

Min-max Risk

$$y^* = \underset{y}{\operatorname{argmin}} \ \max_{u} \left(\Delta(y,u)\, p(u|X) \right)$$

$$y^* = \underset{y}{\operatorname{argmin}} \ \max_{u} \begin{bmatrix} \Delta(1,1)P(1|X), & \Delta(1,2)P(2|X), & \Delta(1,3)P(3|X) \\ \Delta(2,1)P(1|X), & \Delta(2,2)P(2|X), & \Delta(2,3)P(3|X) \\ \Delta(3,1)P(1|X), & \Delta(3,2)P(2|X), & \Delta(3,3)P(3|X) \end{bmatrix}$$

$X \to$ obs
$C \to$ class

$$\underbrace{P(X,C)}_{\text{Joint}} = \underbrace{P(X|C)}_{\substack{\text{likelihood}\\\text{(Density)}}} \cdot \underbrace{P(C)}_{\text{Prior}} = \underbrace{P(C|X)}_{\substack{\text{Posterior}\\\text{Prob}}} \underbrace{P(X)}_{\substack{\text{Marginal}\\\text{Prob}}}$$

* $P(C) = [0.5, 0.2, 0.3]$

Given Any x

* $P(X) = \sum_C P(X,C)$

$P(X) = P(X,C_1) + P(X,C_2) + P(X,C_3)$

* $P(C_1|X) = \dfrac{P(X,C_1)}{P(X,C_1) + P(X,C_2) + P(X,C_3)}$

$P(C_2|X) = \dfrac{P(X,C_2)}{P(X,C_1) + P(X,C_2) + P(X,C_3)}$

$P(C_3|X) = \dfrac{P(X,C_3)}{P(X,C_1) + P(X,C_2) + P(X,C_3)}$

$$PP = \begin{array}{c} C_1 \quad C_2 \quad C_3 \\ [\ 1 \times 3\ ] \end{array} \quad \text{zero\_one} = \begin{array}{c} P \backslash T \\ \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \end{array}$$

$PP = [P(C_1|X) \ P(C_2|X) \ P(C_3|X)]$

Classification

i) MAP
ii) Minimum Risk

$$\text{ZERO\_ONE} = \begin{array}{c} P \backslash T \\ \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \end{array}$$

$$\text{Cost} = P(C|X)\, \Delta(y_{TRUE}, y_{PRED})$$

For x, $C = [C_1, C_2, C_3]$

1) cost for class C_1 : (y_1)

$\cancel{Y}$ $\text{Cost}\_1 = P(C=1|X)\, \Delta(y_{T=1}, y_{P=1}) + \cancel{P}(C=2|X)\, \Delta(\cancel{y}_{T=2}, y_{P=1})$
$+ P(C=3|X)\, \Delta($

$\text{cost}\_1 = \cancel{P(C=1)}\, P(C_1|X)\, \Delta(y_1, \hat{y}_1)$
$+ P(C_2|X)\, \Delta(y_1, \hat{y}_2)$
$+ P(C_3|X)\, \Delta(y_1, \hat{y}_3)$

2) cost for C_2: (y_2) ~~Consider~~ True class = 2

$$cost\_2 = p(c_1|X) \cdot \Delta(y_2, \hat{y_1}) + p(c_2|X) \cdot \Delta(y_2, \hat{y_2}) + p(c_3|X) \cdot \Delta(y_2, \hat{y_3}) +$$

3) cost for C_3: (y_3)

$$cost\_3 = p(c_1|X) \cdot \Delta(y_3, \hat{y_1}) + p(c_2|X) \cdot \Delta(y_3, \hat{y_2}) + p(c_3|X) \cdot \Delta(y_3, \hat{y_3})$$

Minimum Risk dec. rule = min (cost_1, cost_2, cost_3)

$$= \underset{y}{\arg\min} \sum_{\hat{y}} p(\hat{y}|x) \Delta(y, \hat{y})$$

1.

$X = m \times n$

$Y = m \times 1$

$C = 1$

$tol = 1 \times 10^{-3}$

$\alpha = m \times 1$

$b = 0$

$E = m \times 1$

Linear kernel:

$K = X \times X'$

$K = (\underline{m} \times \underline{n})\ (n \times m)$

$$\begin{array}{c|c} & x\ y\ z \\ \hline \begin{array}{c} ex1 \\ ex2 \\ ex3 \end{array} & \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} \end{array} \begin{array}{c|c} & ex1\ ex2\ ex3 \\ \hline \begin{array}{c} x \\ y \\ z \end{array} & \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} \end{array}$$

$K = m \times n$

$$\begin{array}{cccc} & 1st\ ex & 2nd\ ex & \\ K = & \begin{bmatrix} (1,1) & (1,2) & (1,3) \\ (2,1) & (2,2) & (2,3) \\ (3,1) & (3,2) & (3,3) \end{bmatrix} \end{array}$$

kernel computation b/w ex 1 & ex 1

kernel computation b/w ex 1 & ex 2

$= (ex1\_x \cdot ex2\_x) + (ex1\_y \cdot ex2\_y) + (ex1\_z \cdot ex2\_z)$

For $i = 1 : m$

$$E(i) = f(x^{(i)}) - y^{(i)}$$

$$= \sum_{i=1}^{m} \alpha_i y^{(i)} \langle x^{(i)}, x \rangle + b - y^{(i)}$$

$$= b + \sum_{i=1}^{m} \alpha_i y^{(i)} \langle x^{(i)} \cdot x \rangle - y^{(i)}$$

$$= b + \text{sum} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix} [y] K(:, i) - y^{(i)}$$

$$E(i) = \boxed{b + \text{sum} \begin{bmatrix} \alpha_1 y^1 \\ \alpha_2 y^2 \\ \alpha_m y^m \end{bmatrix} \begin{bmatrix} 1,1 \\ 2,1 \\ 3,1 \\ m,m \end{bmatrix}} - y^{(i)}$$

kernel matrix b/w for 1st example & all examples

$$\begin{bmatrix} 1,1 \\ 2,1 \\ 3,1 \\ \vdots \\ m,1 \end{bmatrix}$$

End

$$X = \begin{bmatrix} \text{support vectors} \times 2 \end{bmatrix}$$

$$y = \begin{bmatrix} \cancel{\text{supp}}\, sv \times 1 \end{bmatrix}$$

$$\alpha = \begin{bmatrix} sv \times 1 \end{bmatrix}$$

$$w = \sum_{i=1}^{sv} \alpha_i y_i x_i \in \mathbb{R}^{1 \times 2}$$

$$= \cancel{\alpha_1 y_1 x_1} \left(\left[\overset{\alpha}{\begin{bmatrix} \vdots \end{bmatrix}} \cdot \overset{y}{\begin{bmatrix} \vdots \end{bmatrix}} \right]^T \times \begin{bmatrix} sv \times 2 \end{bmatrix} \right)'$$

$$= \left(\begin{pmatrix} \alpha_1 y_1 \\ \alpha_2 y_2 \\ \alpha_3 y_3 \\ \alpha_{sv} y_{sv} \end{pmatrix}^T \begin{bmatrix} sv \times 2 \end{bmatrix} \right)'$$

$$= \begin{bmatrix} \alpha_1 y_1 & \alpha_2 y_2 & \alpha_3 y_3 & \dots & \alpha_{sv} y_{sv} \end{bmatrix} \left(\begin{bmatrix} sv\_x \end{bmatrix} \begin{bmatrix} sv\_y \end{bmatrix} \right)$$

$$w = \left[\alpha_1 y_1 (sv\_x_1) + \alpha_2 y_2 (sv\_x_2) + \dots \alpha_{sv} y_{sv} (sv\_x_{sv}) \quad \alpha_1 y_1 (sv\_y_1) \dots \alpha_{sv} y_{sv} (sv\_y_{sv}) \right]$$

$$w = \begin{bmatrix} \cdot \\ \cdot \end{bmatrix}$$

$$w(1) = \sum_{i=1}^{sv} \alpha_i y_i X(:,1)$$

$$w(2) = \sum_{i=1}^{sv} \alpha_i y_i X(:,2)$$

Optimal Margin Classifier

$$\min_{\gamma,w,b} \frac{1}{2}\|w\|^2$$

$$\text{s.t } y^{(i)}(w^T x^{(i)}+b) \geq 1 \ , \ i=1,\dots,m$$

Let

$$g_i(w) = y^{(i)}(w^T x^{(i)}+b) \geq 1$$

OR

$$g_i(w): y^{(i)}(w^T x^{(i)}+b) - 1 \geq 0$$

$$\boxed{g_i(w): -y^{(i)}(w^T x^{(i)}+b) + 1 \leq 0}$$

Construct the lagrangian of above problem : $f(w) = \frac{1}{2}\|w\|^2$

$$\leftarrow \mathcal{L}(w,\alpha,\beta) = f(w) + \sum_{i=1}^{K} \alpha_i g_i(w)$$

$$\mathcal{L}(w,\alpha,\beta) = \frac{1}{2}\|w\|^2 + \sum_{i=1}^{m} \alpha_i g_i(w)$$

$$\mathcal{L}(w,\alpha,\beta) = \frac{1}{2}\|w\|^2 + \sum_{i=1}^{m} \alpha_i \left[-y^{(i)}(w^T x^{(i)}+b) + 1\right]$$

$$\mathcal{L}(w,\alpha,\beta) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \alpha_i y^{(i)}(w^T x^{(i)}+b) + \sum_{i=1}^{m} \alpha_i$$

$\mathcal{L}(w,b,\alpha) \leftarrow$ OR

$$\boxed{\mathcal{L}(w,\alpha,\beta) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \alpha_i \left[y^{(i)}(w^T x^{(i)}+b) - 1\right]} \rightarrow (A)$$

differential wrt "w"

$$\nabla_w \mathcal{L}(w,\alpha,\beta) = \frac{d}{dw}\left[\frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \alpha_i \left[y^{(i)}(w^T x^{(i)}+b) - 1\right]\right]$$

$$= \frac{d}{dw}\frac{1}{2}\|w\|^2 - \frac{d}{dw}\sum_{i=1}^{m} \alpha_i y^{(i)} w^T x^{(i)} - \frac{d}{dw}\sum_{i=1}^{m} \alpha_i y^{(i)} b$$

$$+ \frac{d}{dw}\sum_{i=1}^{m} \alpha_i$$

$$\nabla_w \mathcal{L}(w,\alpha,\beta) = w - \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)} - 0 + 0$$

As $\nabla_w \mathcal{L}(w,\alpha,\beta) = 0$

$$\Rightarrow w - \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)} = 0$$

$$w = \sum_{i=1}^{m} \alpha_i y_i x_i$$

$$\text{OR} \quad \boxed{w = \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)}} \longrightarrow (9)$$

Diff (A) w.r.t B

$$\cancel{\nabla_B \mathcal{L}(w, b, \alpha)} \quad \nabla_b \mathcal{L}(w, b, \alpha) = \frac{d}{db}\left\{ \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \alpha_i \left[y_i (w^T x^{(i)} + b) - 1 \right] \right\}$$

$$= 0 - \frac{d}{db} \sum_{i=1}^{m} \alpha_i y_i w^T x^{(i)} - \frac{d}{db} \sum_{i=1}^{m} \alpha_i y_i b + \frac{d}{db} \sum_{i=1}^{m} \alpha_i$$

Set $\nabla_b \mathcal{L}(w, b, \alpha) = 0$

$$\Rightarrow -\frac{d}{db} \sum_{i=1}^{m} \alpha_i y_i b = 0$$

$$\boxed{\sum_{i=1}^{m} \alpha_i y_i = 0} \longrightarrow (10)$$

As (A):

$$\mathcal{L}(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \alpha_i \left[y^{(i)} (w^T x^{(i)} + b) - 1 \right]$$

$$= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \left[\alpha_i y^{(i)} w^T x^{(i)} + \alpha_i y_i b - \alpha_i \right]$$

$$= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \alpha_i y^{(i)} w^T x^{(i)} - \sum_{i=1}^{m} \alpha_i y_i b + \sum_{i=1}^{m} \alpha_i \longrightarrow (10A)$$

As $w = \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)}$, substitute in above equation

$$= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \alpha_i y^{(i)} \left(\sum_{j=1}^{m} \alpha_j y^{(j)} x^{(j)} \right)^T x^{(i)} - b \underbrace{\sum_{i=1}^{m} \alpha_i y_i}_{\to 0} + \sum_{i=1}^{m} \alpha_i$$

$$= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \alpha_i y^{(i)} \left(\sum_{j=1}^{m} \alpha_j y^{(j)} \right) \langle x^{(i)}, x^{(j)} \rangle - b(0) + \sum_{i=1}^{m} \alpha_i$$

$$= \frac{1}{2} \sum_{i=1}^{m} \alpha_i^2 y^{(i)2} \langle x^{(i)}, x^{(i)} \rangle - \sum_{\substack{i=1 \\ j=1}}^{m} \alpha_i y^{(i)} \alpha_j y^{(j)} \langle x^{(i)}, x^{(j)} \rangle + \sum_{i=1}^{m} \alpha_i$$

Rewrite (10A):

$$\mathcal{L}(\omega,b,\alpha) = \frac{1}{2}\omega^T\cdot\omega - \sum_{i=1}^{m}\alpha_i y^{(i)}\omega^T x^{(i)} - \sum_{i=1}^{m}\alpha_i y_i b + \sum_{i=1}^{m}\alpha_i$$

As $\omega = \sum_{i=1}^{m}\alpha_i y^{(i)} x^{(i)}$

$$\mathcal{L}(\omega,b,\alpha) = \frac{1}{2}\left(\sum_{i=1}^{m}\alpha_i y^{(i)} x^{(i)}\right)^T\left(\sum_{j=1}^{m}\alpha_j y^{(j)} x^{(j)}\right) - \sum_{i=1}^{m}\alpha_i y^{(i)}\left(\sum_{j=1}^{m}\alpha_j y^{(j)} x^{(j)}\right)^T x^{(i)}$$

$$- b\underbrace{\sum_{i=1}^{m}\alpha_i y_i}_{\to 0} + \sum_{i=1}^{m}\alpha_i$$

$$= \frac{1}{2}\sum_{\substack{i=1\\j=1}}^{m}\alpha_i\alpha_j y^{(i)}y^{(j)}\langle x^{(i)},x^{(j)}\rangle - \sum_{\substack{i=1\\j=1}}^{m}\alpha_i\alpha_j y^{(i)}y^{(j)}\langle x^{(i)},x^{(j)}\rangle$$

$$+ \sum_{i=1}^{m}\alpha_i$$

$$\boxed{\mathcal{L}(\omega,b,\alpha) = \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{m}\alpha_i\alpha_j y^{(i)}y^{(j)}\langle x^{(i)},x^{(j)}\rangle}$$

we get this equation by minimizing $\mathcal{L}$ w.r.t ω & b

Putting this thing together with $\alpha_i \geq 0$ & $\sum_{i=1}^{m}\alpha_i y_i = 0$, we obtain the dual optimization prob as:

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{m}\alpha_i\alpha_j y^{(i)}y^{(j)}\langle x^{(i)},x^{(j)}\rangle$$

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

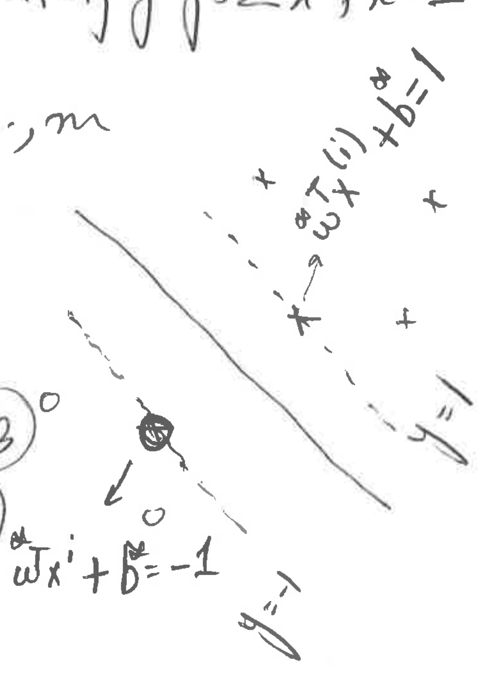
$$\sum_{i=1}^{m}\alpha_i y_i = 0$$

<u>For b</u> once we have ω^*,

For +ve Support vector: $\min_{i:y^{(i)}=1}\left[\omega^{*T}x^{(i)}\right] + b^* = 1 \longrightarrow$ (10B)

For −ve SV: $\max_{i:y^{(i)}=-1}\left[\omega^{*T}x^{(i)}\right] + b^* = -1 \longrightarrow$ (10C)

Add (10B) & (10C)



$$\min_{i:y^{(i)}=1}\left[w^{*T}x^{(i)}\right]+\max_{i:y^{(i)}=-1}\left[w^{*T}x^{(i)}\right]+2b=0$$

$$\min_{i:y^{(i)}=1} w^{*T}x^{(i)}+\max_{i:y^{(i)}=-1} w^{*T}x^{(i)}=-2b^{*}$$

$$\boxed{b^{*}=-\frac{\left[\max_{i:y^{(i)}=-1} w^{*T}x^{(i)}+\min_{i:y^{(i)}=1} w^{*T}x^{(i)}\right]}{2}} \rightarrow 11$$

Non-Separable CASE with Regulariser

$$\min_{\gamma,\omega,b} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{m}\xi_i$$

$$\text{s.t} \quad y^{(i)}(\omega^T x^{(i)} + b) \geq 1 - \xi_i \quad , i = 1, \dots, m$$

$$\xi_i \geq 0$$

Forming the Lagrangian of above primal problem:

$$\mathcal{L}(\omega, b, \xi, \alpha, \gamma) = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{m}\xi_i + \alpha_i(\underbrace{\text{constraint}}_{g}) + \gamma_i(\underbrace{\text{constraint}}_{h})$$

$$g: \; y^{(i)}(\omega^T x^{(i)} + b) - 1 + \xi_i \geq 0$$

For Lagrangian, ~~they~~ constraints are transformed into ≤ 0

$$\left.\begin{aligned} g: &\quad -y^{(i)}(\omega^T x^{(i)} + b) + 1 - \xi_i \leq 0 \\ \text{Similarly } h: &\quad -\xi_i \leq 0 \end{aligned}\right\} \; i = 1 \dots m$$

$$\Rightarrow \mathcal{L}(\omega, b, \xi, \alpha, \gamma) = \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{m}\xi_i + \sum_{i=1}^{m}\alpha_i\left[-y^{(i)}(\omega^T x^{(i)} + b) + 1 - \xi_i\right] + \sum_{i=1}^{m}\gamma_i(-\xi_i)$$

<u>Primal variables</u>: ω, ξ, b

<u>Dual variables</u>: α, γ

$$\boxed{\begin{aligned} \Rightarrow \mathcal{L}(\omega, b, \xi, \alpha, \gamma) &= \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{m}\xi_i - \sum_{i=1}^{m}\alpha_i\left[y^{(i)}(\omega^T x^{(i)} + b) - 1 + \xi_i\right] - \sum_{i=1}^{m}\gamma_i\xi_i \\ \text{And} \quad &\alpha_i \geq 0 \quad i = 1, \dots, m \\ &\gamma_i \geq 0 \quad i = 1, \dots, m \end{aligned}} \rightarrow (A)$$

$$(A) \Rightarrow \mathcal{L}(\omega, b, \xi, \alpha, \gamma) = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{m}\xi_i - \sum_{i=1}^{m}\alpha_i y^{(i)}\omega^T x^{(i)} - \sum_{i=1}^{m}\alpha_i y^{(i)} b + \sum_{i=1}^{m}\alpha_i - \sum_{i=1}^{m}\alpha_i\xi_i - \sum_{i=1}^{m}\gamma_i\xi_i$$

Minimizing (A) w.r.t w, ξ, b while keeping α, γ fixed

$$\frac{d}{dw} \mathcal{L}(w, b, \xi, \alpha, \gamma) = 0 \longrightarrow (A1)$$

$$\frac{d}{db} \mathcal{L}(w, b, \xi, \alpha, \gamma) = 0 \longrightarrow (A2)$$

$$\frac{d}{d\xi} \mathcal{L}(w, b, \xi, \alpha, \gamma) = 0 \longrightarrow (A3)$$

(A1) $$\frac{d}{dw} \mathcal{L}(w, b, \xi, \alpha, \gamma) = \frac{d}{dw} 2\|w\|^2 + 0 - \frac{d}{dw} \sum_{i=1}^{m} \alpha_i \left[y^{(i)} [w^T x^{(i)} + b] - 1 + \xi_i \right] - 0$$

$$0 = w - \frac{d}{dw} \sum_{i=1}^{m} \left[\alpha_i y^{(i)} w^T x^{(i)} + \alpha_i b y^{(i)} - \alpha_i + \alpha_i \xi_i \right]$$

$$\Rightarrow w - \frac{d}{dw} \sum_{i=1}^{m} \alpha_i y^{(i)} w^T x^{(i)} - \frac{d}{dw} \sum_{i=1}^{m} b y^{(i)} \alpha_i + \frac{d}{dw} \sum_{i=1}^{m} \alpha_i - \frac{d}{dw} \sum_{i=1}^{m} \alpha_i \xi_i = 0$$

$$w - \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)} - 0 + 0 - 0 = 0$$

$$\boxed{w = \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)}} \longrightarrow RA1$$

(A2) $$\Rightarrow \frac{d}{db} (\text{eq } A) = 0 + 0 - 0 - \sum_{i=1}^{m} \alpha_i y^{(i)} + 0 - 0 - 0$$

$$\Rightarrow \boxed{\sum_{i=1}^{m} \alpha_i y^{(i)} = 0} \longrightarrow RA2$$

(A3) $$\Rightarrow \frac{d}{d\xi} (\text{eq } A) = 0 + C \frac{d}{d\xi_i} \sum_{i=1}^{m} \xi_i - 0 - 0 + 0 - \frac{d}{d\xi_i} \sum_{i=1}^{m} \alpha_i \xi_i - \frac{d}{d\xi_i} \sum_{i=1}^{m} \gamma_i \xi_i$$

$$0 = C \frac{d}{d\xi_i} \sum_{i=1}^{m} \xi_i - \frac{d}{d\xi_i} \sum_{i=1}^{m} \alpha_i \xi_i - \frac{d}{d\xi_i} \sum_{i=1}^{m} \gamma_i \xi_i$$

$\rightarrow 1 \quad 0 = C - \alpha_1 - \gamma_1$
$\rightarrow 2 \quad 0 = C - \alpha_2 - \gamma_2$
$\rightarrow 3$
$\vdots$
$\rightarrow m \quad 0 = C - \alpha_m - \gamma_m$

$$0 = C - \alpha_i - \gamma_i$$

$$\Rightarrow \boxed{C - \alpha_i - \gamma_i = 0} \longrightarrow RA3 \quad i = 1, \ldots, m$$

OR

$$\boxed{C = \alpha_i + \gamma_i} \longrightarrow RA3$$

(A) $\Rightarrow \mathcal{L}(w,b,\xi,\alpha,\gamma) = \frac{1}{2}w^Tw + C\sum_{i=1}^{m}\xi_i - \sum_{i=1}^{m}\alpha_i y^{(i)} w^T x^{(i)}$

$$- \sum_{i=1}^{m}\alpha_i y^{(i)} b + \sum_{i=1}^{m}\alpha_i - \sum_{i=1}^{m}\alpha_i\xi_i - \sum_{i=1}^{m}\gamma_i\xi_i$$

And $w = \sum_{i=1}^{m}\alpha_i y^{(i)} x^{(i)} \longrightarrow$ RA1

$\sum_{i=1}^{m}\alpha_i y^{(i)} = 0 \longrightarrow$ RA2

$C - \alpha_i - \gamma_i = 0 \longrightarrow$ RA3 OR $\boxed{\gamma_i = C - \alpha_i}$

Substitute w in above $\mathcal{L}(w,b,\xi,\alpha,\gamma)$

$$\mathcal{L}(w,b,\xi,\alpha,\gamma) = \frac{1}{2}\left[\sum_{i=1}^{m}\alpha_i y^{(i)} x^{(i)}\right]^T \left[\sum_{j=1}^{m}\alpha_j y^{(j)} x^{(j)}\right] + C\sum_{i=1}^{m}\xi_i - \sum_{i=1}^{m}\alpha_i y^{(i)} \underbrace{\left(\sum_{j=1}^{m}\alpha_j y^{(j)} x^{(j)}\right)^T}_{} x^{(i)}$$

$$- b\underbrace{\sum_{i=1}^{m}\alpha_i y^{(i)}}_{0} + \sum_{i=1}^{m}\alpha_i - \sum_{i=1}^{m}\xi_i\underbrace{(\alpha_i + \gamma_i)}_{C}$$

$$= \frac{1}{2}\sum_{\substack{i=1\\j=1}}^{m}\alpha_i\alpha_j y^{(i)}y^{(j)}\langle x^{(i)}, x^{(j)}\rangle + \cancel{C\sum_{i=1}^{m}\xi_i} - \sum_{\substack{i=1\\j=1}}^{m}\alpha_i\alpha_j y^{(i)}y^{(j)}\langle x^{(i)}, x^{(j)}\rangle$$

$$+ \sum_{i=1}^{m}\alpha_i - \cancel{C\sum_{i=1}^{m}\xi_i}$$

$$\mathcal{L}(w,b,\xi,\alpha,\gamma) = \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{m}\alpha_i\alpha_j y^{(i)}y^{(j)}\langle x^{(i)}, x^{(j)}\rangle$$

s.t $\boxed{\sum_{i=1}^{m}\alpha_i y^{(i)} = 0}$

Now $C - \alpha_i - \gamma_i = 0$

$\Rightarrow \alpha_i = C - \gamma_i \rightarrow \leq C$

As $\gamma_i \geq 0$

& $\alpha_i \geq 0$

$\Rightarrow \gamma_i \leq C$

or $\gamma_i = C - \alpha_i$ As $\gamma_i, \alpha_i \geq 0$

To keep $\gamma_i \geq 0$

$\boxed{\alpha_i \leq C}$

<u>Derived</u>

& $\boxed{\alpha_i \geq 0}$ condition of Lagrange multipliers

$\Rightarrow \boxed{0 \leq \alpha_i \leq C}$ $i = 1, \ldots, m$

Thus Dual of non-separable case is:

$$\max_{\alpha} W(\alpha) = \max_{\alpha} \mathcal{L}(\omega, b, \xi, \alpha, \gamma)$$

$$\cancel{\max_{\alpha} W(\alpha) =} \quad \boxed{\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} \langle x^{(i)}, x^{(j)} \rangle \\ \text{s.t} \quad & \sum_{i=1}^{m} \alpha_i y^{(i)} = 0 \\ & 0 \leq \alpha_i \leq C \ , \ \forall i = 1, \cdots, m \end{aligned}}$$

Constraint : $C - \alpha_i - \gamma_i = 0$ was used to derive bounds on α_i i.e $0 \leq \alpha_i \leq C \ \forall i = 1, \ldots, m$

Lets Assume we have 5 data points. Then objective becomes:

$$\text{Expanded objective} \left\{ \begin{aligned} \max_{\alpha} \Big\{ & (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5) - \frac{1}{2} \Big[\overbrace{(\alpha_1 \alpha_1 y^{(1)} y^{(1)} \langle x^{(1)}, x^{(1)} \rangle)}^{i=1,\ j=1} \\ & + \underbrace{(\alpha_1 \alpha_2 y^{(1)} y^{(2)} \langle x^{(1)}, x^{(2)} \rangle)}_{i=1\ j=2} + \underbrace{(\alpha_1 \alpha_3 y^{(1)} y^{(3)} \langle x^{(1)}, x^{(3)} \rangle)}_{i=1,\ j=3} \\ & + \ldots + \underbrace{(\alpha_5 \alpha_5 y^{(1)} y^{(5)} \langle x^{(5)}, x^{(5)} \rangle)}_{i=5,\ j=5} \Big] \Big\} \end{aligned} \right.$$

$$\text{Expanded constraints} \left\{ \begin{aligned} \text{s.t} \quad & \alpha_1 y^{(1)} + \alpha_2 y^{(2)} + \alpha_3 y^{(3)} + \alpha_4 y^{(4)} + \alpha_5 y^{(5)} = 0 \\ & 0 \leq \alpha_1 \leq C \\ & 0 \leq \alpha_2 \leq C \\ & 0 \leq \alpha_3 \leq C \\ & 0 \leq \alpha_4 \leq C \\ & 0 \leq \alpha_5 \leq C \end{aligned} \right.$$

Kernel trick

$\vec{x} = (\vec{x_1}, \vec{x_2})$
$\vec{z} = (\vec{z_1}, \vec{z_2})$

We are in 2D-space. Lets consider $K(x,z) = (\vec{x}\cdot\vec{z})^2$

$$
\begin{aligned}
K(x,z) &= (\vec{x}\cdot\vec{z})^2 \\
&= (\vec{x}^T\vec{z})^2 \\
&= (x_1 \cdot z_1 + x_2 \cdot z_2)^2 \\
&= (x_1 z_1)^2 + 2(x_1 z_1)(x_2 z_2) + (x_2 z_2)^2 \\
&= (x_1 z_1)^2 + 2(x_1 x_2)(z_1 z_2) + (x_2 z_2)^2 \longrightarrow \text{eq } (1)
\end{aligned}
$$

Let $\vec{x} = x_1 \hat{i} + x_2 \hat{j}$ (2D-space)

From 2D-space, we would like to learn polynomial of degree-2. Thus we define ~~$\phi(x) = (x_1 + x_2)^2$~~ following feature map to learn non-linear boundary.

$$\phi(x) = (x_1^2, \sqrt{2}\, x_1 x_2, x_2^2)$$

Now in dual of SVM, we need to compute $\phi(x)^T \phi(z)$ which would be:

$$
\begin{aligned}
\phi(x)^T \phi(z) &= \begin{bmatrix} x_1^2 \\ \sqrt{2} x_1 x_2 \\ x_2^2 \end{bmatrix}^T \begin{bmatrix} z_1^2 \\ \sqrt{2} z_1 z_2 \\ z_2^2 \end{bmatrix} \\
&= x_1^2 z_1^2 + 2 x_1 x_2 z_1 z_2 + x_2^2 z_2^2 \\
&= (x_1 z_1)^2 + 2(x_1 x_2)(z_1 z_2) + (x_2 z_2)^2 \longrightarrow \text{eq } (2)
\end{aligned}
$$

Eq (1) & Eq (2) are similar. This implies that if we want to learn non-linear decision boundary by transforming

$$\vec{x} : (x_1, x_2) \longrightarrow \phi(x) : (x_1^2, \sqrt{2} x_1 x_2, x_2^2)$$

Then we need to compute $\phi(x)$ for every sample. However, we can learn the same decision boundary by eq (1) without calculating $\phi(x)$. using valid kernel function [$(\vec{x}\cdot\vec{z})^2$, in our case].

Using representer theorem:

$$w = \sum_{i=1}^{n} \alpha_i y_i x_i$$

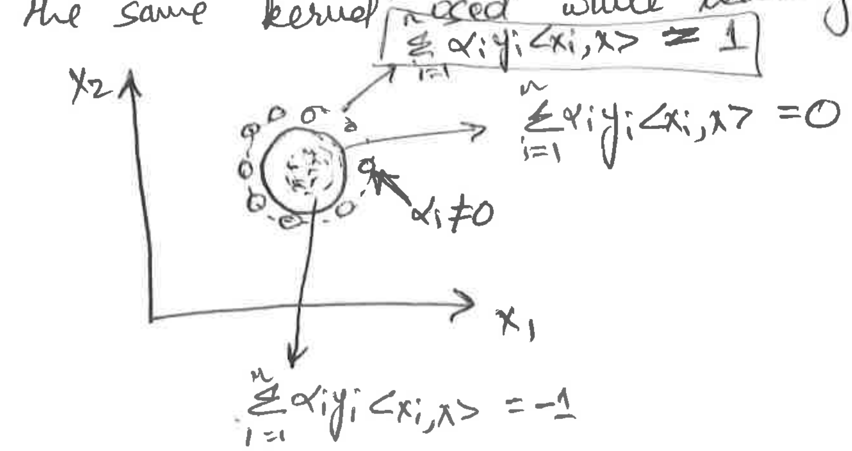
Our prediction function is:

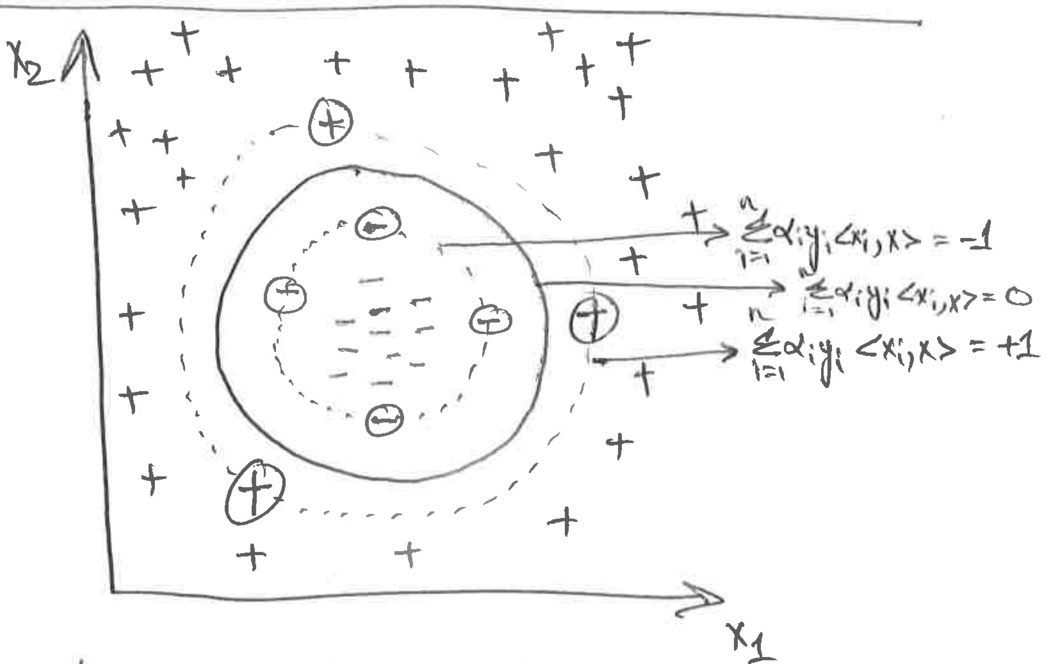
$$w^T x + b \quad \text{or} \quad w^T x.$$

$$w^T x = \sum_{i=1}^{n} \alpha_i y_i x_i \cdot x$$

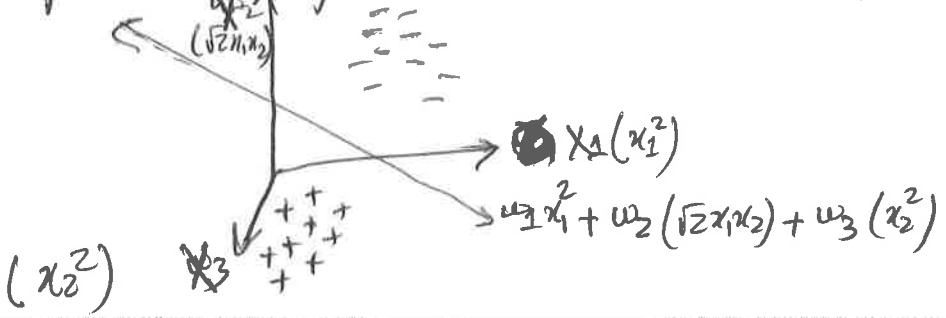
$$\boxed{w^T x = \sum_{i=1}^{n} \alpha_i y_i \langle x_i, x \rangle}$$

Using learned α_i, we can plot non-linear decision boundary using the same kernel used while learning.





In 2D, we can not separate above data by linear function. However, if we do $(x_1, x_2) \longmapsto (x_1^2, \sqrt{2}x_1x_2, x_2^2)$ then we are actually transforming 2D to 3D where it might look like:



i.e we can easily separate our data using hyperplane in the transformed 3D-space. However, if we look at equation of separating hyperplane in 2D, it will trace an ellipse because:

$w_1 x_1^2 + w_2 \sqrt{2}(x_1 x_2) + w_3 x_2^2 = 0$ is an equation of ellipse in R^2.

Quadratic term for dual of SSVM

$$\sum_{i,y}\sum_{j,\bar{y}} \alpha(iy)\,\alpha(j\bar{y})\,\bar{J}_{(iy)(j\bar{y})}$$

Let x sequence of 2 nodes
x_1 " " 3 "
be training set

$$= \sum_{y\bar{y}}\Big[\underbrace{\alpha_{(1y)}\alpha_{(1\bar{y})}\bar{J}_{(1y)(1\bar{y})}}_{4\times 4 \text{ terms}} + \underbrace{\alpha_{(1y)}\alpha_{(2\bar{y})}\bar{J}_{(1y)(2\bar{y})}}_{4*8 \text{ terms}}$$

$$+ \underbrace{\alpha_{(2y)}\alpha_{(1\bar{y})}\bar{J}_{(2y)(2\bar{y})}}_{8*4 \text{ terms}} + \underbrace{\alpha_{(2y)}\alpha_{(2\bar{y})}\bar{J}_{(2y)(2\bar{y})}}_{8*8 \text{ terms}}\Big]$$

Ex:

$$\begin{array}{c|cc} & 1 & 2 \\ \hline 1 & \bar{J}_{(1y)(1\bar{y})} & \bar{J}_{(1y)(2\bar{y})} \\ 2 & \bar{J}_{(2y)(2\bar{y})} & \bar{J}_{(2y)(2\bar{y})} \end{array}$$

$$\sum_{y\bar{y}} \alpha_{1(y)}\alpha_{(1\bar{y})} = \alpha_{1y_1}\alpha_{1\bar{y}_1} + \alpha_{1y_1}\alpha_{1\bar{y}_2} + \alpha_{1y_1}\alpha_{1\bar{y}_3} + \alpha_{1y_1}\alpha_{1\bar{y}_4}$$
$$+ \alpha_{1y_2}\alpha_{1\bar{y}_1} + \alpha_{1y_2}\alpha_{1\bar{y}_2} + \alpha_{1y_2}\alpha_{1\bar{y}_3} + \alpha_{1y_2}\alpha_{1\bar{y}_4}$$
$$+ \alpha_{1y_3}\alpha_{1\bar{y}_1} + \alpha_{1y_3}\alpha_{1\bar{y}_2} + \alpha_{1y_3}\alpha_{1\bar{y}_3} + \alpha_{1y_3}\alpha_{1\bar{y}_4}$$
$$+ \alpha_{1y_4}\alpha_{1\bar{y}_1} + \alpha_{1y_4}\alpha_{1\bar{y}_2} + \alpha_{1y_4}\alpha_{1\bar{y}_3} + \alpha_{1y_4}\alpha_{1\bar{y}_4}$$

$$\sum_{y\bar{y}} \alpha_{(1y)}\alpha_{(1\bar{y})} = \begin{array}{c|cccc} & \bar{y}_1 & \bar{y}_2 & \bar{y}_3 & \bar{y}_4 \\ \hline y_1 & \alpha_{y_1\bar{y}_1} & \alpha_{y_1\bar{y}_2} & \alpha_{y_1\bar{y}_3} & \alpha_{y_1\bar{y}_4} \\ y_2 & & \cdots & & \\ y_3 & & & \cdots & \\ y_4 & & & & \alpha_{y_4\bar{y}_4} \end{array}$$

Sum all entries of above matrix.

Here y & $\bar{y}$ are violated labelings