$X_1, X_2, X_3 \in a, b, c$

$X_1 \quad X_2 \quad X_3 \quad Y$

$IG(X_1 == a) = \text{Entropy}(\quad) - [\quad]$

" $(X_1 == b)$

$(X_1 == c)$

$(X_2 == a)$

$(X_2 == b)$

$(X_2 == c)$

$(X_3 == a)$

$(X_3 == b)$

$(X_3 == c)$

Objective

$$\text{Min } \frac{1}{2}\|w\|^2 \quad \text{s.t}$$

$$y(\vec{w}\cdot\vec{x} + b) \geq +1 \text{ for all points}$$

Dual Problem

$$\text{Max} \sum_{i=1}^{m} \alpha_i - \sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y^i y^j \langle x^i \cdot x^j \rangle$$

$$\text{s.t} \quad \alpha_i \geq 0$$

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

$$x^i \cdot x^j$$

$$\phi(x^i) \cdot \phi(x^j)$$

$$\underset{1D}{x} \longrightarrow \overset{dD}{x^1, x^2, x^3, x^4 \ldots x^d}$$

Kernel Function

①

$$K(x^i, x^j) = \langle \phi(x^i) \cdot \phi(x^j) \rangle$$

$$K(X, Z) = (\langle X \cdot Z \rangle)^2$$

$$= (X^T \cdot Z)^2$$

$$X = [X_1 \quad X_2] \cdots n$$

$$Z = [Z_1 \quad Z_2)$$

$O(n)$

$$(X^T \cdot Z)^2 = (X_1 Z_1 + X_2 Z_2)^2$$

$$X_1 Z_1 X_1 Z_1 + X_2 Z_2 X_2 Z_2 + 2 X_1 Z_1 X_2 Z_2$$

2

$$K(x, z) = (x^T \cdot z)^2$$

$$x = [x_1 \quad x_2] \qquad z = [z_1 \quad z_2]$$

$$\phi(x) = [x_1 x_1 \quad x_1 x_2 \quad x_2 x_2 \quad x_2 x_1]$$

$$\phi(z) = [z_1 z_1 \quad z_1 z_2 \quad z_2 z_2 \quad z_2 z_1]$$

$$= (x_1 x_1 z_1 z_1 + x_1 x_2 z_1 z_2 + x_2 x_2 z_2 z_2 + x_2 x_1 z_2 z_1) \qquad O(n^2)$$

Quadratic Kernel

$$k(X, Z) = (X^T \cdot Z + c)^{2 = d}$$

$$X = [x_1 \; x_2] \qquad Z = [z_1 \; z_2]$$

$$(\underbrace{x_1 z_1 + x_2 z_2}_{} + c)^2 \qquad O(n)$$

$$\left\{ \begin{array}{l} x_1 z_1 x_1 z_1 + x_2 z_2 x_2 z_2 + c^2 + \\ 2 x_1 z_1 c + 2 x_2 z_2 c + 2 x_1 z_1 x_2 z_2 \end{array} \right\}$$

$$(X^T \cdot Z)$$

$$X \cdot Z = [x_1 \; x_2] \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = x_1 z_1 + x_2 z_2$$

$$X^T = [x_1 \ x_2] \qquad Z^T = [z_1 \ z_2]$$

$$\phi(x) = \begin{bmatrix} x_1 x_1 \\ \sqrt{2}\, x_1 x_2 \\ x_2 x_2 \\ \sqrt{2c}\, x_1 \\ \sqrt{2c}\, x_2 \\ c \end{bmatrix} \qquad \phi(z) = \begin{bmatrix} z_1 z_1 \\ \sqrt{2}\, z_1 z_2 \\ z_2 z_2 \\ \sqrt{2c}\, z_1 \\ \sqrt{2c}\, z_2 \\ c \end{bmatrix}$$

$$X^T = [x_1 \ x_2 \ \cdots \ x_n]$$

$$\phi(x) = \Big[\underset{i}{\text{all}}\ x_i^2,\ \ \overset{n-1}{\underset{i=1}{\text{all}}}\ \overset{n}{\underset{j=i}{\text{all}}} \sqrt{2}\, x_i x_j,\ \ \underset{i}{\text{all}}\ \sqrt{2c}\, x_i,\ \ c\Big] \qquad O(n^2)$$

$$\binom{n+d}{d}$$

$$K(x,z) = \exp\left(\frac{||x-z||^2}{2\sigma^2}\right)$$

$$K(x,z) = \langle \Phi(x) \cdot \phi(z) \rangle$$

$K(x,z)$ - large if x & z are similar

small if dissimilar

Linear kernel

$$K(x,z) = (x^T z + \underline{c})$$

Valid kernels = Mercer kernel

$K(x, z)$ is valid if there exists ϕ function s.t

$$K(x, z) = \langle \phi(x) \cdot \phi(z) \rangle$$

Data set $= \{ x^{(1)}, x^{(2)}, \ldots \ldots x^{(m)} \}$

Kernel Matrix $= K$

$$\begin{bmatrix} \underset{11}{x^{(1)} \cdot x^{(1)}} & \underset{12}{x^{(1)} \cdot x^{(2)}} & \underset{13}{x^{(1)} x^{(3)}} & \cdots & \underset{1m}{x^{(1)} \cdot x^{(m)}} \\ \underset{21}{x^{(2)} \cdot x^{(1)}} & \underset{22}{x^{(2)} \cdot x^{(2)}} & & & \underset{2m}{x^{(2)} \cdot x^{(m)}} \\ \vdots & & & & \\ \vdots & & & & \\ \underset{m1}{x^{m} \cdot x^{1}} & \cdots & \cdots & \cdots & \underset{mm}{x^{(m)} \cdot x^{(m)}} \end{bmatrix}$$

K = Kernel Matrix $m \times m$

Then for any vector $z \in \mathbb{R}^m$

$$\boxed{z^T K z} \geqslant 0 \qquad K(x,z) = \phi(x)\phi(z)$$

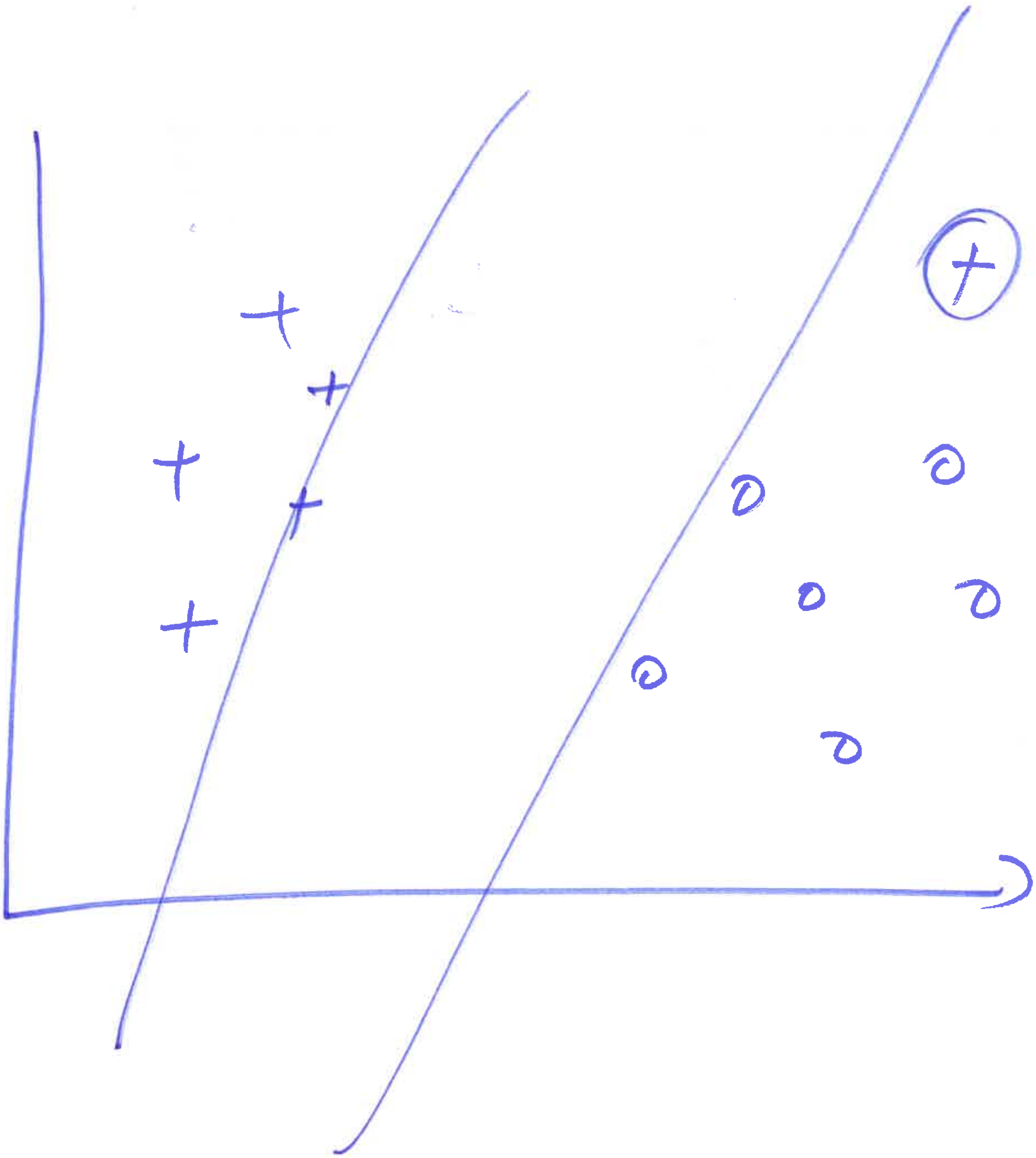
~~m x m~~
1 x ~~m~~
m x n
1 x ~~m~~
~~m~~ x 1
1 x 1

$$= \sum\sum z_i k_{ij} z_j$$

$$k_{ij} = \phi(x^i) \cdot \phi(x^j)$$

$$\geqslant 0$$

Soft Margin SVM



Min

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i$$

s.t

$$\begin{cases} y(w^T x + b) - 1 + \xi_i \geq 0 \\ \xi_i \geq 0 \end{cases}$$

$$\mathcal{L}(w, b, \alpha, r) +$$

$$= \frac{1}{2}w^2 + C\sum \xi_i - \sum_{i=1}^{m}\alpha_i\left(y(w^T x + b) - 1 + \xi_i\right)$$

$$- \sum_{i=1}^{m} r_i \xi_i$$

$$\frac{\delta L}{\delta w} \Rightarrow \underbrace{w = \sum_{i=1}^{m} \alpha_i y^i x^i}$$

$$\frac{\delta L}{\delta b} = \underbrace{\sum \alpha_i y^i = 0}$$

$$\frac{\delta L}{\xi_i} = C - \alpha_i - r_i = 0$$

$$\underbrace{\alpha_i = C - r_i}$$

max

$$\sum \alpha_i - \frac{1}{2} \sum \sum \alpha_i \alpha_j y^i y^j \langle x^i \cdot x^j \rangle$$

$$\text{s.t } C \geq \alpha_i \geq 0 , \quad \sum \alpha_i y^i = 0$$

$$\alpha = 0 \Rightarrow y(w^T x + b) \geq 1$$

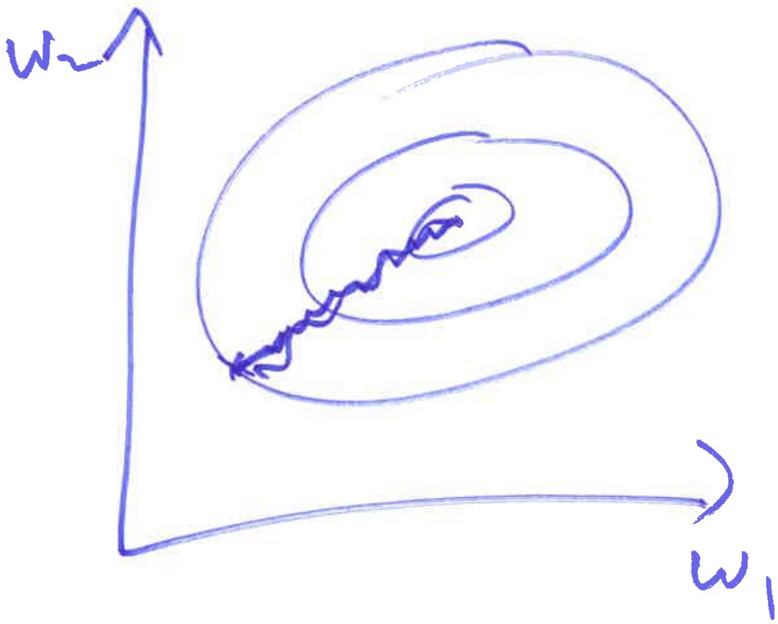
$$0 < \alpha < C \Rightarrow y(w^T x + b) = 1$$

$$\alpha = C \Rightarrow y^i(w^T x + b) \leq 1$$

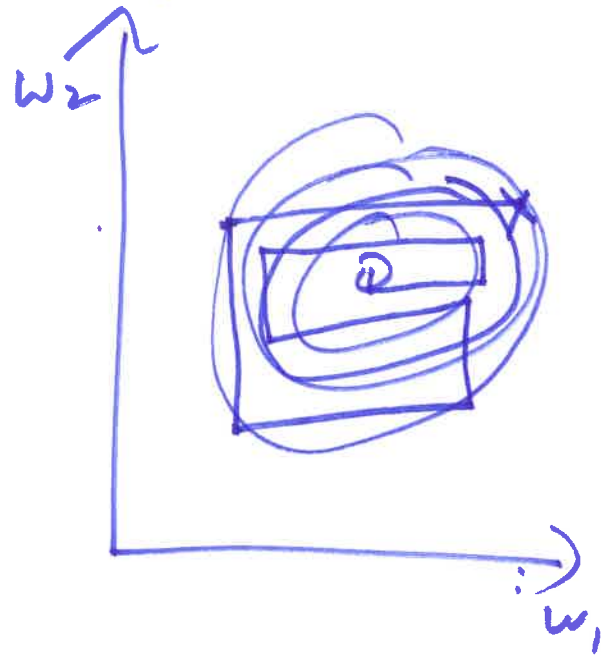
SMO → Sequential Minimum Optimization

Co-ordinate Accent

Gradient Decent



Co-ordinate Decent



max $W(\alpha_1, \alpha_2 \ldots \ldots \alpha_m)$

no constraints on α_i's

Repeat { For $i = 1$ to m

$$\alpha_i = \arg\max_{\alpha_i} W(\alpha_i, \underbrace{\alpha_1 \cdots \alpha_{i-1}, \alpha_{i+1} \cdots \alpha_m}_{\text{consts}})$$

}

SVM

We have constrain $\sum \alpha_i y^i = 0$

$\alpha_1 \quad \alpha_2$

$$\sum \alpha_i y^i = 0$$

$$\alpha_1 y^1 + \alpha_2 y^2 + \alpha_3 y^3 + \cdots \alpha_m y^m = 0$$

$$\alpha_1 y^1 + \alpha_2 y^2 = -(\alpha_3 y^3 + \cdots \alpha_m y^m)$$

$$= -\xi$$

$$\alpha_1 = +y^1(-\xi - \alpha_2 y^2)$$

$$a\alpha_2^2 + b\alpha_2 + c = 0$$

$$ax^2 + by + c = 0$$

0

A

$$x = \cancel{\pm b} -$$

$$= \frac{-b \pm \sqrt{b^2 - 4ac}}{\cancel{4ac}\ 2a}$$

$$c < \alpha > 0$$