Support Vector Machine (SVM)

→ optimal hyperplane that separates the two classes $\{+1, -1\}$

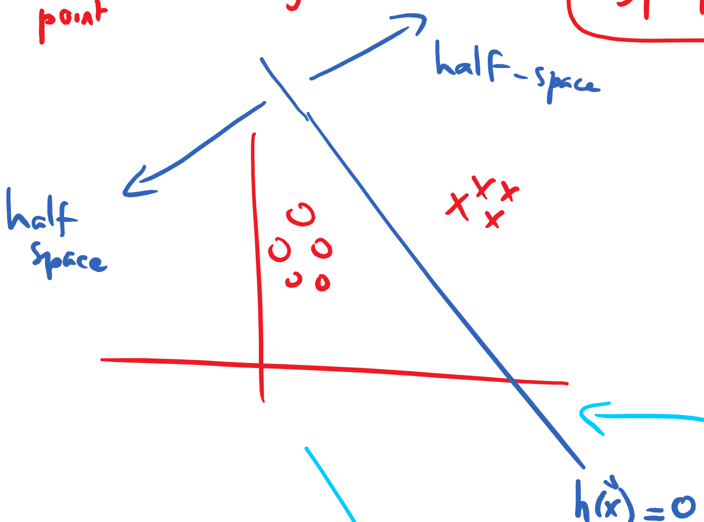
$$h(x) = \underbrace{w_1 x_1 + w_2 x_2 + \dots + w_d x_d}_{} + \underset{\substack{\uparrow \\ \text{scalar} \\ \text{bias}}}{b} = w^T x + b$$

$$\underbrace{\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix}}_{\text{any point}} \qquad \underbrace{\vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{pmatrix}}_{\text{weight vector}}$$

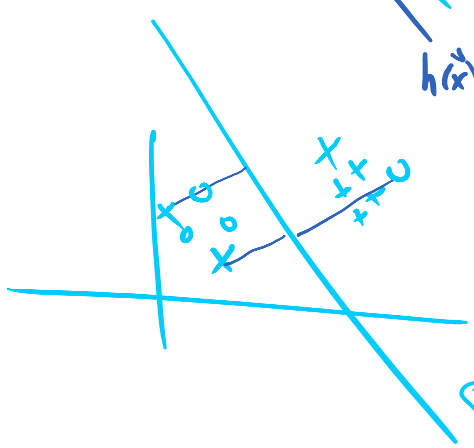
$$\boxed{\text{hyperplane}: \; h(\vec{x}) = 0}$$

$$\boxed{h(\vec{x} \mid \vec{w}, b) = 0}$$

$\vec{w}, b$: parameters of h



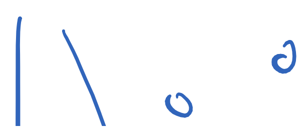
← separating hyperplane



← optimal "sep" hyperplane

Maximum margin classification

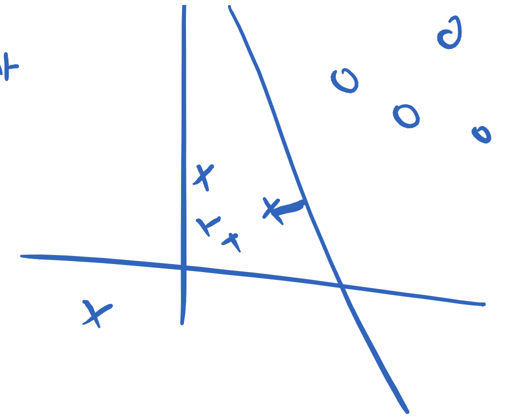
margin ← distance to the closest point

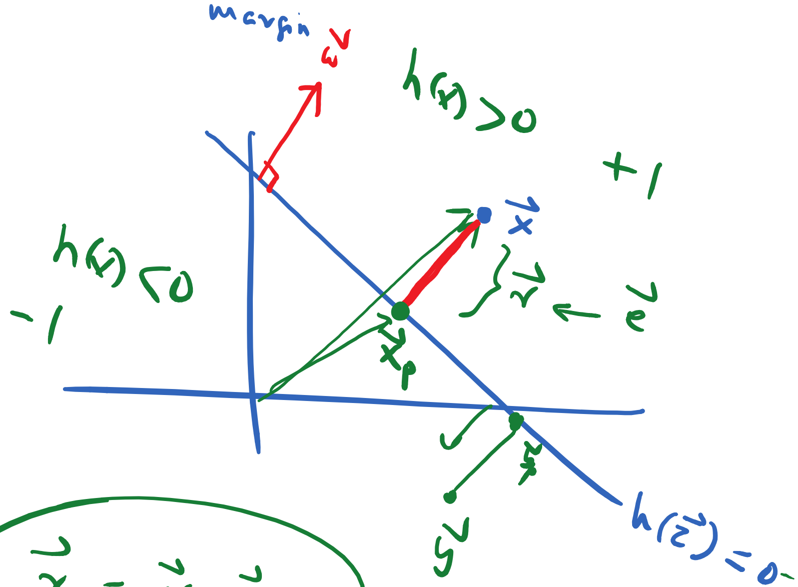


distance to the closest point

$$\max_{h:\,\vec{w},b}\left\{\text{min distance to } h(\vec{x})\right\}$$

margin





$\vec{w}$ is orthogonal to the hyperplane

"normal" vector

$$\vec{r} = \vec{x} - \vec{x}_p$$

$$\vec{r} = \alpha \cdot \frac{\vec{w}}{\|\vec{w}\|}$$

distance

$$\vec{x} = \vec{r} + \vec{x}_p = \left(\vec{x}_p + \alpha \cdot \frac{\vec{w}}{\|\vec{w}\|}\right)$$

$$h(\vec{x}) = \vec{w}^T \vec{x} + b$$

$$= \vec{w}^T\left(\vec{x}_p + \alpha \cdot \frac{\vec{w}}{\|\vec{w}\|}\right) + b$$

$$= \underbrace{\vec{w}^T x_p + b} + \alpha \cdot \underbrace{\frac{\vec{w}^T\vec{w}}{\|\vec{w}\|}}$$

$$w^T w = \|w\|^2$$

$$\boxed{h(\vec{x}) = \underbrace{\cancel{h(\vec{x}_p)}}_{0} + \alpha \cdot \|\vec{w}\|}$$

$$\boxed{\alpha = \frac{h(\vec{x})}{\|\vec{w}\|}}$$

signed distance

$(\vec{x}, y)$

feature vector

true class

$y = \{+1, -1\}$

gned distance $\underbrace{\qquad\qquad}_{\|\vec{w}\|}$

$$\alpha = \frac{y \cdot h(\vec{x})}{\|\vec{w}\|}$$

non-negative distance

$h(\vec{x}) < 0 \quad y = -1$

$h(\vec{x}) > 0 \quad y = +1$

SVM objective:

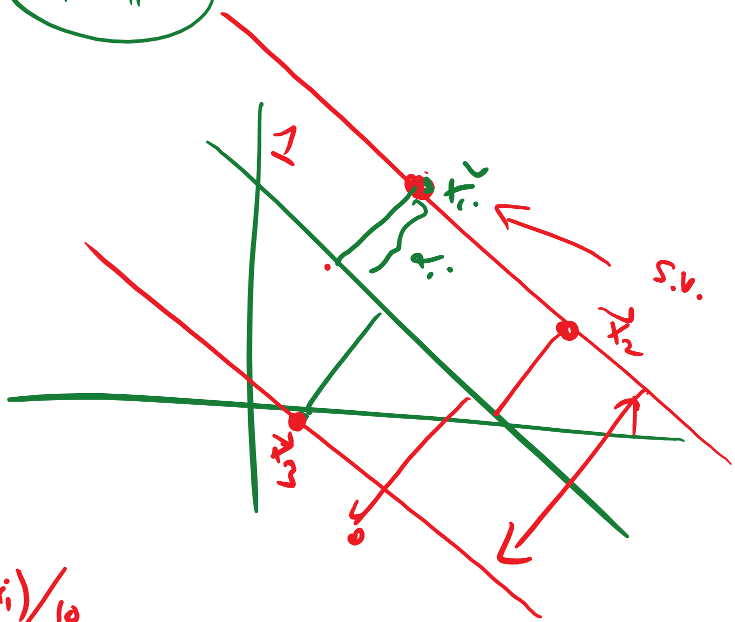
$$\max_{\substack{h \\ \vec{w}, b}} \left\{ \underbrace{\min_{i=1}^{n} \frac{y_i \, h(\vec{x}_i)}{\|\vec{w}\|}}_{\text{margin}} \right\}$$

$D = \vec{x}_1, y_1$
$\quad \vec{x}_2, y_2$
$\quad \vdots$
$\quad \vec{x}_n, y_n$

$$\alpha_i = \frac{y_i \left(w^T \vec{x}_i + b \right)}{\|\vec{w}\|}$$

distance of $\vec{x}_i$ to h

$\|\vec{w}\|$ to be small



$$\boxed{s:\ h(\vec{x}) = 0.5}$$

scalar s

$$s(w^T x + b) = 0$$

$$(sw)^T x + (sb) = 0$$

$$(w')^T x + b' = 0$$

$$\alpha_i = \frac{y_i h(x_i)/10}{\|w\|}$$

$$y_i h(x_i) = 10$$

$h(x)$ for which the margin $= 1$ is called the canonical hyperplane

$$\max_{\vec{\omega}, b} \left\{ \frac{1}{\|\omega\|} \right\}$$

subject to the constraint that all points should be at a d of 1 or more from h.

$$y_i \, h(x_i) \geq 1$$

$$y_i (\omega^T x_i + b) \geq 1 \qquad n \text{ constraints}$$

$$\frac{1}{\sqrt{\omega_1^2 + \omega_2^2 + \dots + \omega_d^2}}$$

$$\min_{\vec{\omega}, b} \left\{ \|\vec{\omega}\| \right\}$$

$$\min_{\vec{\omega}, b} \left\{ \frac{1}{2} \|\vec{\omega}\|^2 \right\}$$

$$\frac{d\, x^2}{dx} = 2x$$

PRIMAL FORMULATION

SVM:
$$\min_{\vec{\omega}, b} \left\{ \frac{1}{2} \|\vec{\omega}\|^2 \right\}$$

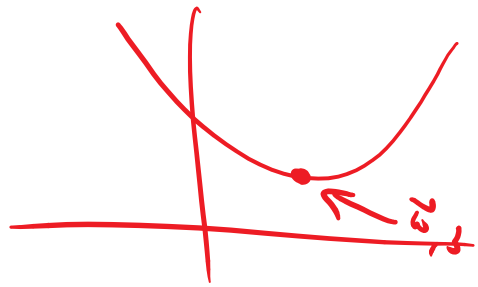
$$\|\vec{\omega}\|^2 = \vec{\omega}^T \vec{\omega}$$

$$\boxed{\text{s.t.} \quad y_i (\omega^T x_i + b) \geq 1} \qquad n \text{ equations/constraints}$$

Convex Quadratic optimization with linear constraints

... quadratic optimization with linear constraints

$\Rightarrow$ find the optimal $\vec{w}$, & b



$$\alpha_i \left(y_i (w^T x_i + b) - 1 \right) \geq 0$$

lagrange multiplier

$$\min_{\vec{w}, b} J = \frac{1}{2} w^T w - \sum_{i=1}^{n} \alpha_i \left(y_i (w^T x_i + b) - 1 \right)$$

$$\frac{\partial J}{\partial \vec{w}} = \frac{1}{\cancel{2}} \cdot \cancel{2} \vec{w} - \sum_{i=1}^{n} \alpha_i y_i x_i = 0$$

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

$$\frac{\partial J}{\partial b} = \boxed{\sum_{i=1}^{n} \alpha_i y_i = 0}$$

$$\alpha_i \geq 0$$

$\alpha_i y_i w^T x_i$

T

$\otimes$

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

dual formulation

$$\min J = \sum \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$\min_{\alpha_i} J = \sum \alpha_i - \frac{1}{2}\sum_{i=1}\sum_{j=1} \alpha_i\alpha_j y_i y_j x_i^T x_j$$

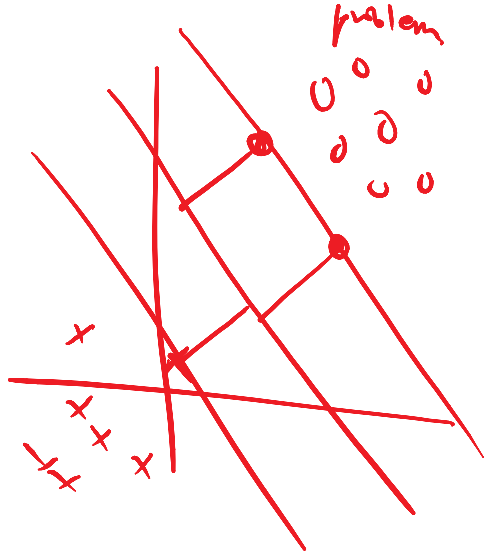
$\alpha_i \geq 0$

Convex quadratic optimization problem

Solve for α_i (n values)

$\vec{w}$, b

$$\boxed{\vec{w} = \sum_{\alpha_i > 0} \alpha_i y_i \vec{x}_i}$$



Only the support vectors have $\alpha_i > 0$

all other points have $\alpha_i = 0$

b ?

$$\boxed{\alpha_i > 0, \quad y_i h(x_i) = 1}$$

$$y_i(w^T x_i + b) = 1$$

$$y_i b = 1 - y_i w^T x_i$$

$$\boxed{b = \frac{1}{y_i} - w^T x_i}$$

take the avg of all b values to set final b.

N : # of support vectors

SVM : $h(\vec{z}) = \begin{cases} +1 & h(\vec{z}) > 0 \end{cases}$

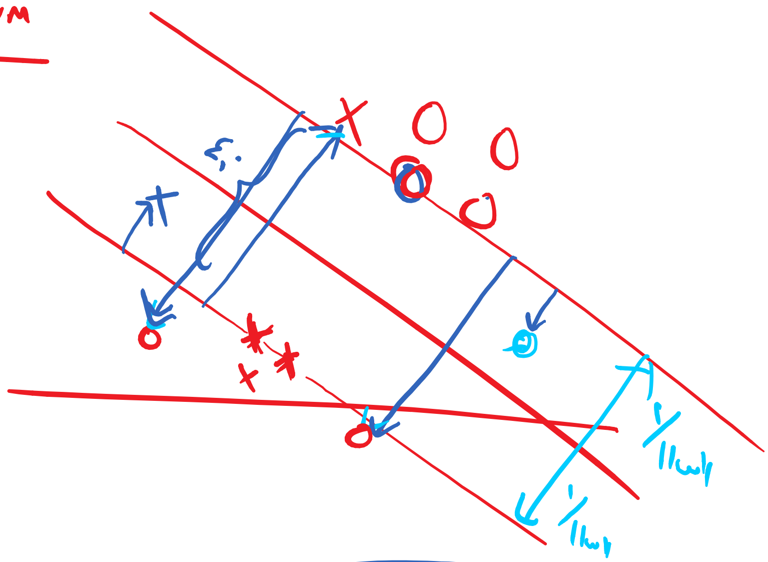
SVM: testing $\vec{z}$

$$h(\vec{z}) = \begin{cases} +1 & h(\vec{z}) > 0 \\ -1 & h(\vec{z}) < 0 \end{cases}$$

$$= \boxed{\operatorname{sign}(h(\vec{z}))}$$

---

soft SVM



max margin

min $\frac{1}{2}\|w\|^2$

+

loss

$$\min_{\vec{w}, b} \left\{ \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n} (\xi_i)^k \right\}$$

k = 1 hinge

k = 2 quadratic

s.t. $y_i(w^T x_i + b) \geq 1 - \xi_i$ n constraints

slack variables

$\xi_i \geq 0$

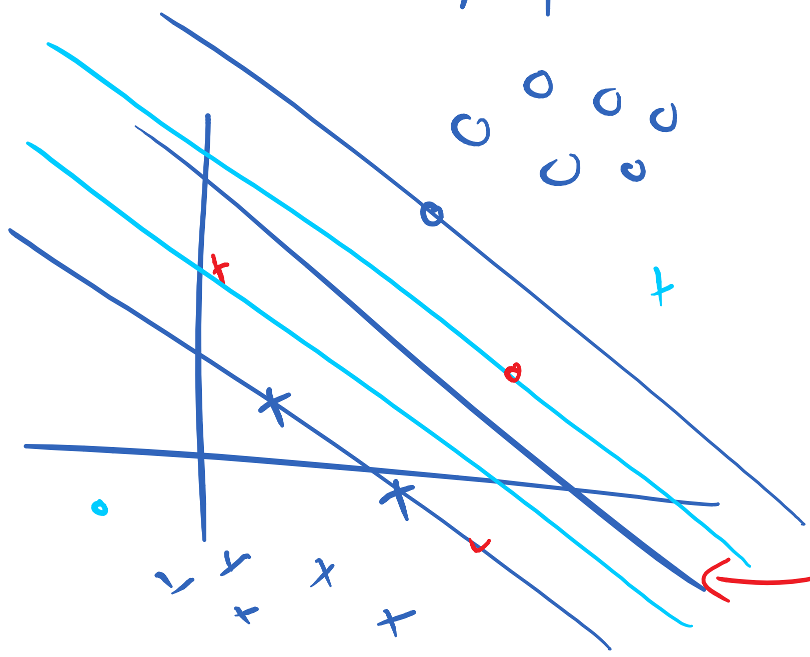
1) $\xi_i = 0$, points with no violation

2) $0 < \xi_i \leq 1$: points within the margin (no real problem)

3) $\xi_i > 1$: ...

(no real problem)

3) $\xi_i > 1$: misclassified points



when C is small we set large margin

C is large then margin will be small

dual : $\sum \alpha_i - \sum \sum \alpha_i \alpha_j y_i y_j x_i^T x_j$

$0 \leq \alpha_i \leq C$

α_i : lagrange multiplier (NOT DISTANCE)

Kernels: non-linear svm

$$\phi(x_i)^T \phi(x_j) = k(x_i, x_j)$$