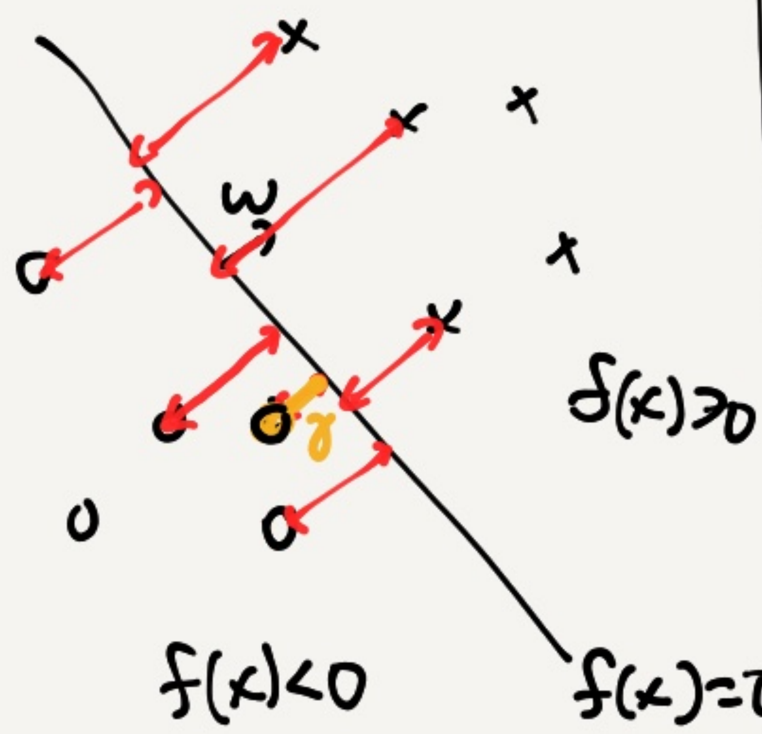
Support Vector Machines (SVM) [Lecture 9]

Linear Classifier

$$f(x) = w^T x + b$$

$$y^* = \begin{cases} +1, & f(x) > 0 \\ -1, & f(x) < 0 \end{cases} = \text{sign}(f(x))$$



Distance from x to boundary is

$$\frac{f(x)}{\|w\|} \qquad \text{(PS 9-1)}$$

- Assume data $\{X, y\}$ is linearly separable:
- "Margin" - distance from the boundary to the closest point in X.

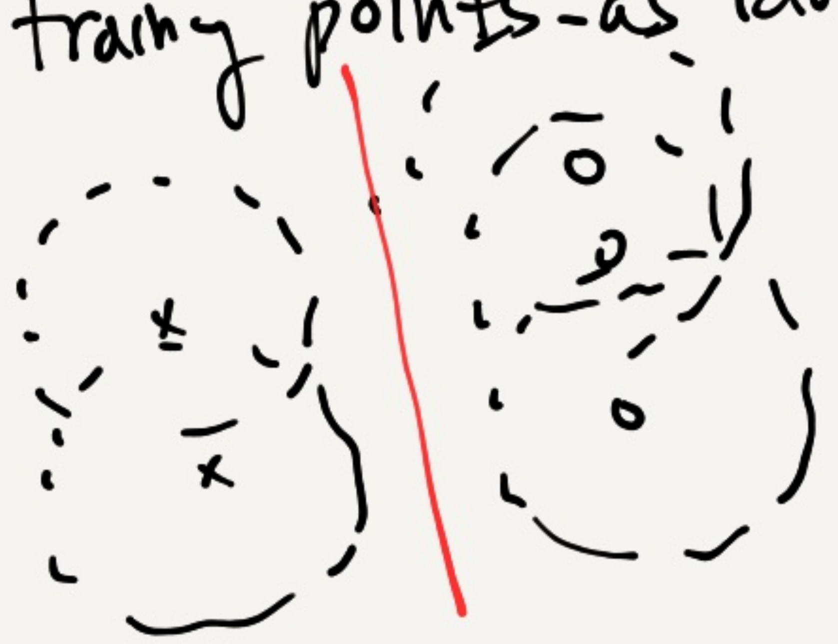
$f(x_i) > 0$ when $y_i = 1$
$f(x_i) < 0$ " $y_i = -1$

$$\gamma = \min_i \frac{|f(x_i)|}{\|w\|} = \min_i \frac{|w^T x_i + b|}{\|w\|} = \min_i \frac{y_i(w^T x_i + b)}{\|w\|}$$

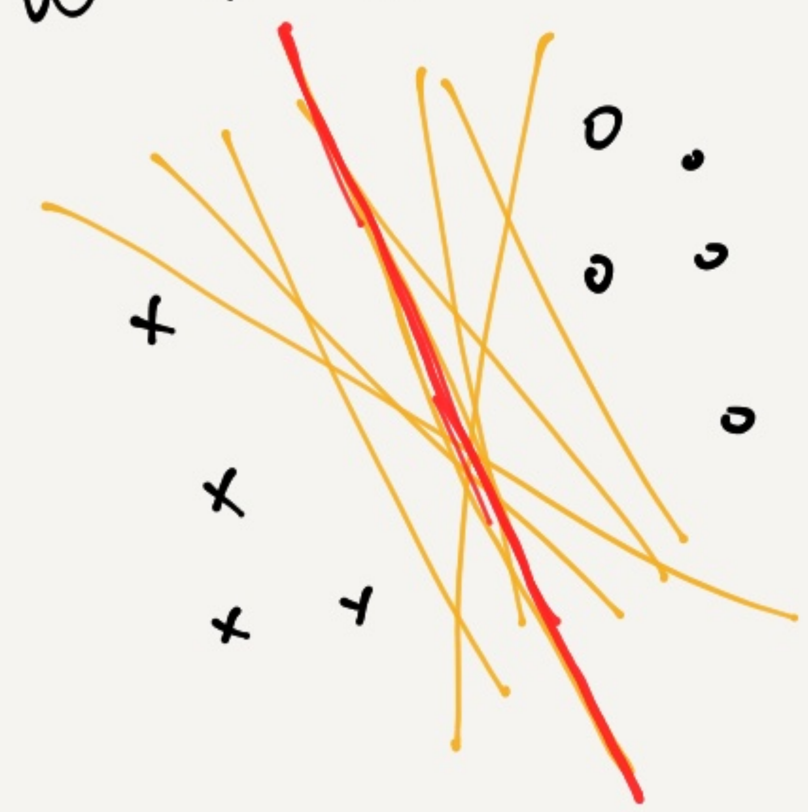
<u>Idea:</u> Maximize the margin, i.e. the separation btwn the boundary & closest point.

<u>Why?</u>
1) w/ perceptron - <u>margin</u> determines the learning complexity.

2) training points - as random samples → leave a margin to be safe from uncertainty.



3) w is an uncertain estimate



max margin → allow more variance in w estimate w/o causing errors.

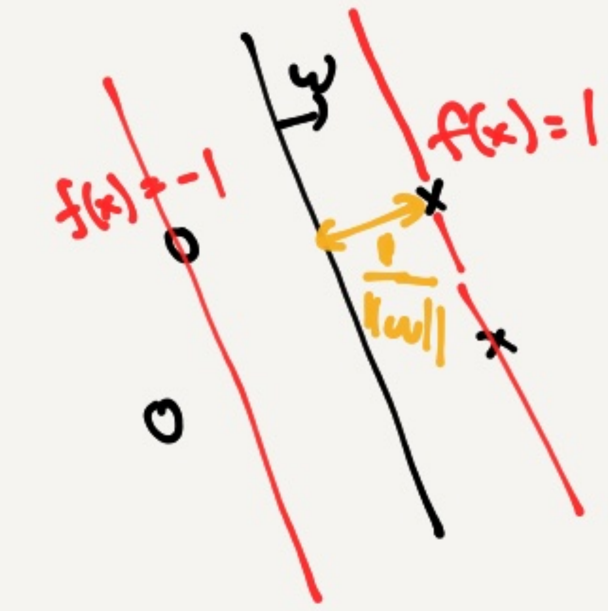
CS5487 Lecture Notes (2020)
Dr. Antoni B. Chan
Dept of Computer Science
City University of Hong Kong

Formulation

Need a normalization ⇒ fix the numerator.

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

$$\Rightarrow \gamma = \frac{1}{\|w\|}$$



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Find w:

$$\underset{w,b}{\text{argmax}}\ \gamma \quad \text{s.t.} \quad \min_i y_i(w^T x_i + b) = 1$$

$$\text{argmax}\ \frac{1}{\|w\|} \quad \text{s.t.} \quad \min_i y_i(w^T x_i + b) = 1$$

apply xform: $\text{argmax}\ \frac{1}{x} = \text{argmin}\ x^2$

$$\text{argmin}\ \frac{1}{2}\|w\|^2 \quad \text{s.t.} \quad \min_i y_i(w^T x_i + b) = 1$$

since the minimum is 1 so other x_i will have > 1

$$\text{argmin}\ \frac{1}{2}\|w\|^2 \quad \text{s.t.} \quad y_i(w^T x_i + b) \geq 1 \quad \forall i$$

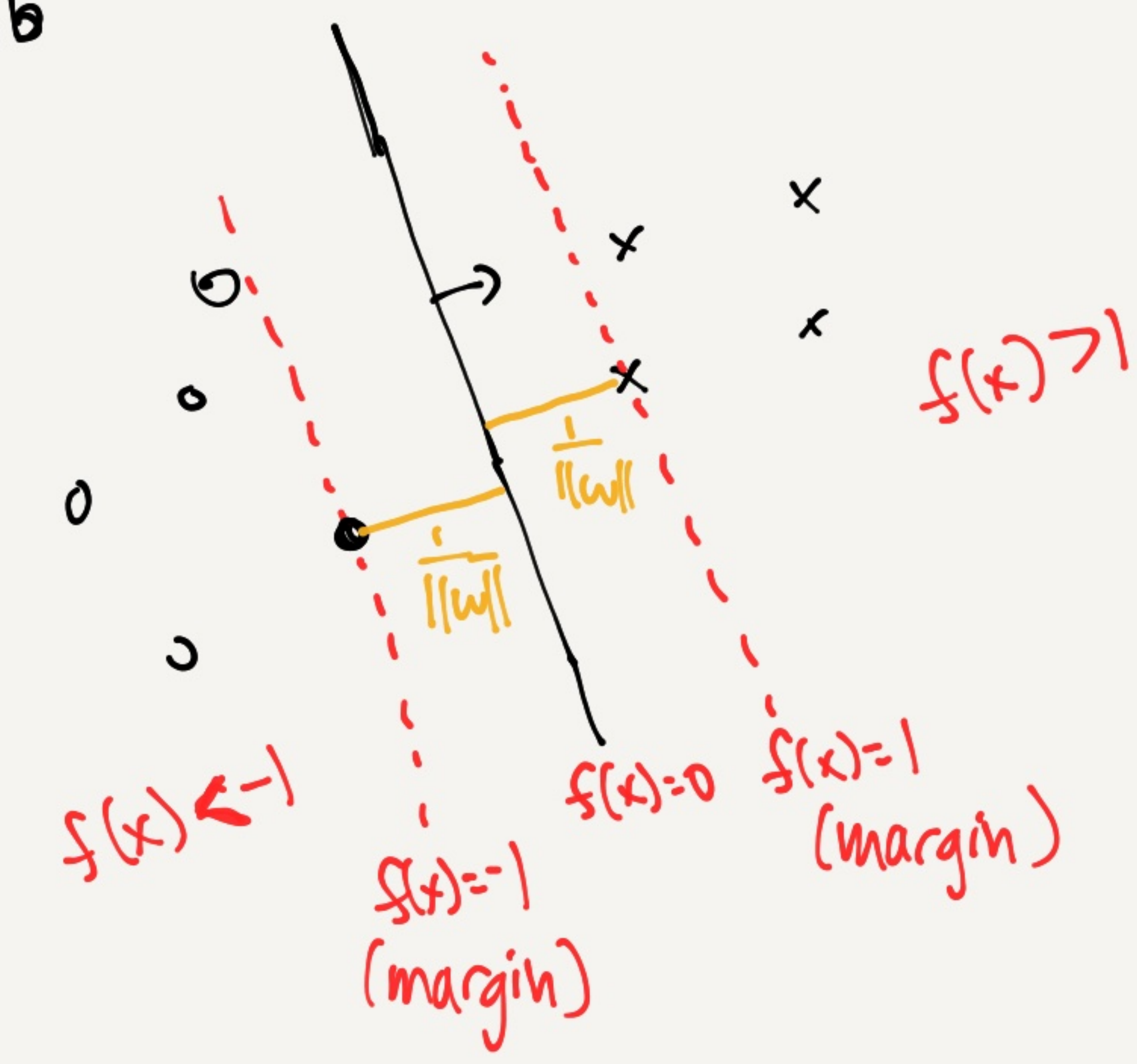
at the optimum, the smallest w will shrink $y_i(w^T x_i + b) = 1$ for some x_i.

$$\text{argmax}\ \frac{1}{\|w\|} = \text{argmax}\ \log\frac{1}{\|w\|} = \text{argmin}\ -\log\frac{1}{\|w\|} = \text{argmin}\ \log\|w\|$$

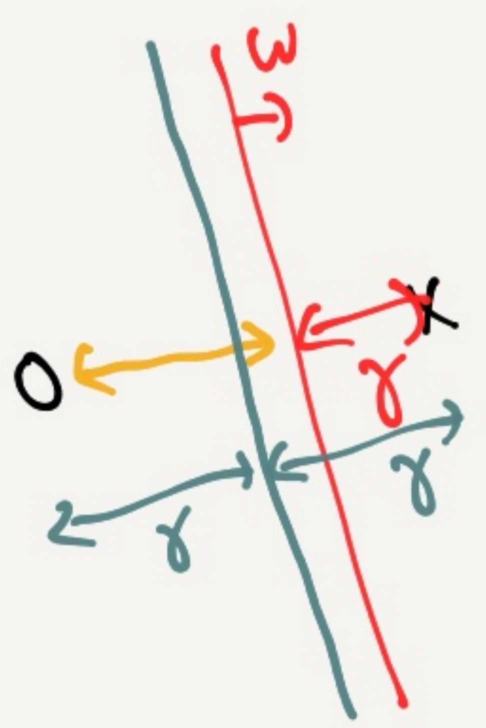
$$= \text{argmin}\ 2\log\|w\| = \text{argmin}\ \|w\|^2$$

SVM primal problem

$$\underset{w,b}{\operatorname{argmin}} \ \frac{1}{2}\|w\|^2 \quad \text{s.t.} \quad y_i(w^T x_i + b) \geq 1, \ \forall i$$



- by symmetry there should be at least 1 point on the margin on each side.



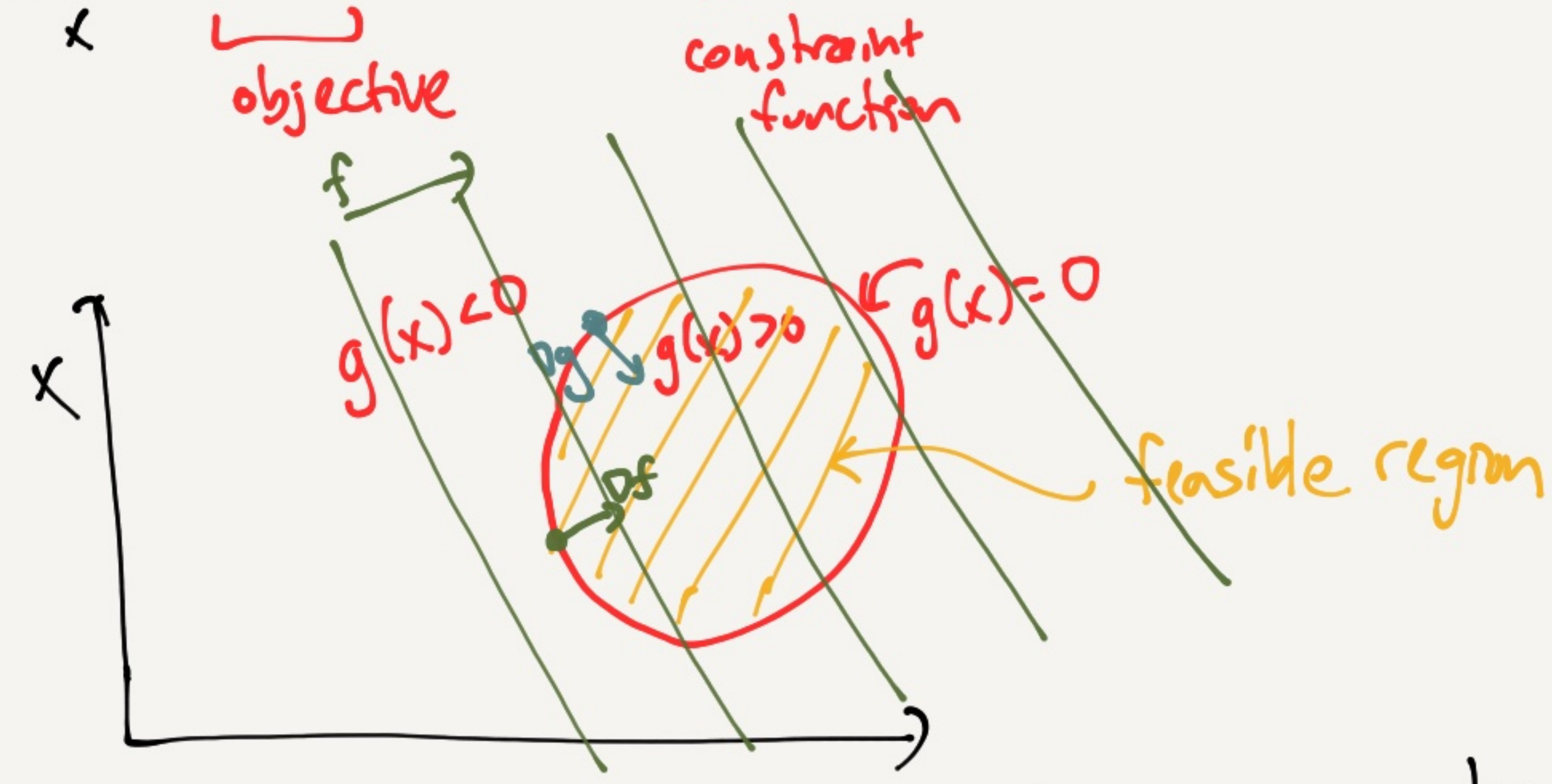
What are the support vectors???

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Optimization w/ inequality Constraints

Goal:

$$\min_x \ \underbrace{f(x)}_{\text{objective}} \quad \text{s.t.} \quad \underbrace{g(x)}_{\text{constraint function}} \geq 0$$



Note: $\nabla g(x)$ points into the feasible region, when x is on the boundary.

Find 2 solutions

1) x^* is on boundary $\Rightarrow g(x^*) = 0$ (active equality)

$\Rightarrow$ min when $\nabla f(x) = \lambda \nabla g(x)$, $\lambda > 0$

f cannot decrease except by leaving the feasible region.

2) x^* is inside feasible region $\Rightarrow g(x^*) > 0$ (inactive)

min when $\nabla f(x) = 0$, i.e. $\lambda = 0$

Combine 2 cases:

$$L(x,\lambda) = f(x) - \lambda g(x)$$

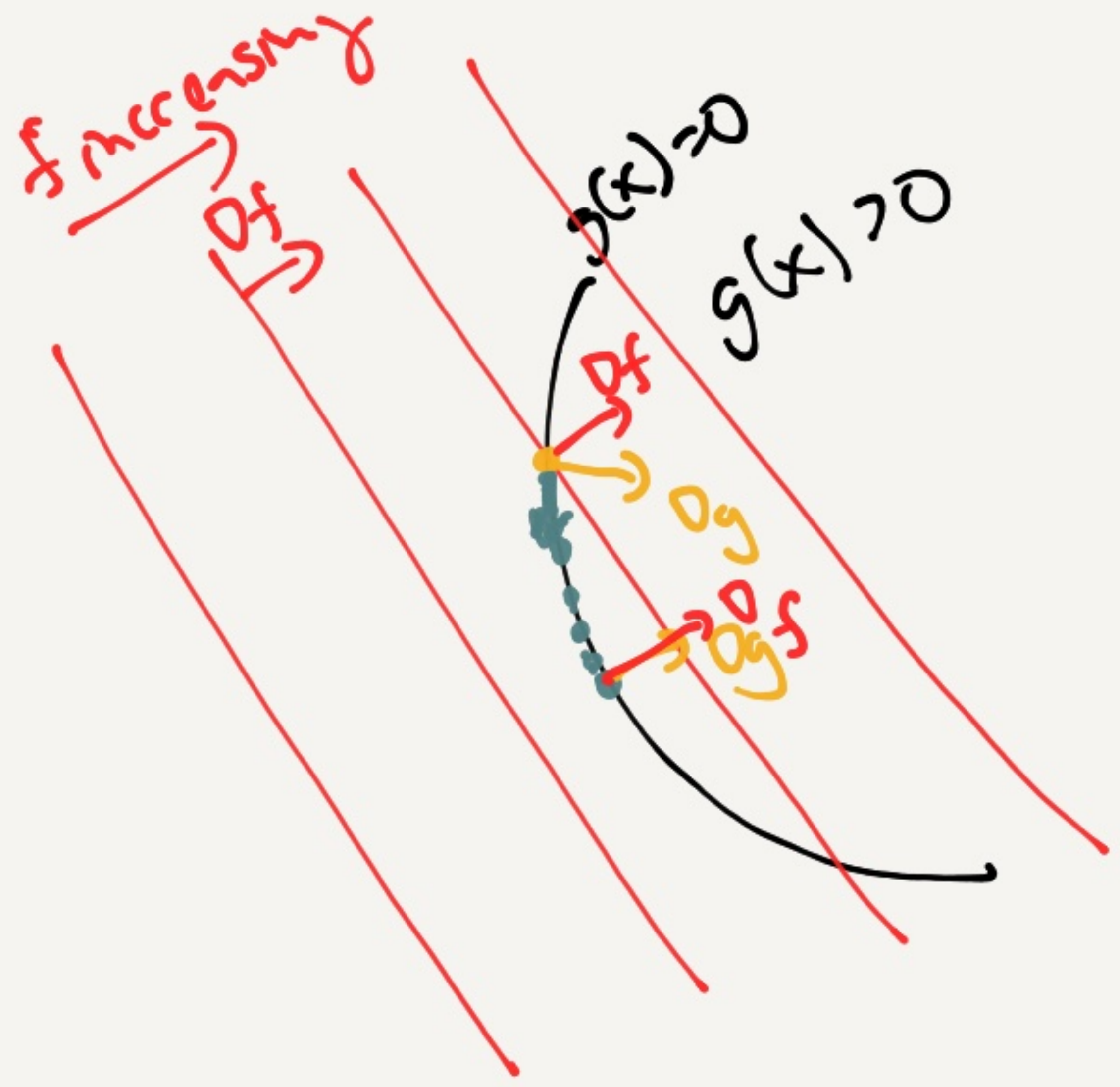
Find stationary point of L

$$\nabla f(x) - \lambda \nabla g(x) = 0$$

s.t.

$$\begin{cases} g(x) \geq 0 \\ \lambda \geq 0 \\ \lambda g(x) = 0 \quad \text{(combine 2 cases)} \end{cases}$$

KKT conditions



Duality

Suppose optimal λ^* is known, then consider minimizing $L(x, \lambda^*)$:

$$L^* = \min_x L(x, \lambda^*) = \min_x [f(x) - \underbrace{\lambda^* g(x)}_{=0 \text{ at the optimum}}]$$

$$= \min_x f(x) = f(x^*) = f^*$$

(the minimum)

Define $g(\lambda) = \min_x L(x,\lambda) = \min_x \underbrace{[f(x) - \lambda g(x)]}_{\text{for every } \lambda \text{, find the min w.r.t. } x.}$

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Note: $\lambda \geq 0$, $g(x) \geq 0$

$$\Rightarrow g(\lambda) \leq \min_{\substack{x \\ g(x)\geq 0}} f(x) = f^*$$

i.e. $g(\lambda)$ is a lower bound to $f^* = f(x^*)$

Thus maximizing $g(\lambda)$ could yield $f(x^*)$ under some conditions.

The dual problem: $g^* = \max_{\lambda \geq 0} g(\lambda)$

Weak duality Thm: $g^* \leq f^*$ ($g^* \neq f^*$ we call it a "duality gap")

Strong Duality Thm: if 1) $f(x)$ is convex; 2) the feasible region is convex (and not degenerate)

$\Rightarrow$ then $g^* = f^*$. i.e. solving the dual is equivalent to solving the primal.

SVM dual

primal: $\underset{w,b}{\operatorname{argmin}} \frac{1}{2}\|w\|^2$ s.t. $y_i(w^T x_i + b) \geq 1$

1) let $\alpha_i \geq 0$ be the Lagrange multiplier for $y_i(w^T x_i + b) \geq 1$

2) Lagrangian

$$L(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \alpha_i \left(y_i(w^T x_i + b) - 1\right)$$

3) Dual function:

$$L(\alpha) = \underbrace{\min_{w,b} L(w,b,\alpha)}$$

$$\frac{\partial L}{\partial w} = w - \sum_i \alpha_i y_i x_i = 0 \Rightarrow w^* = \sum_{i=1}^{N} \alpha_i y_i x_i$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{N} \alpha_i y_i = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0$$

plug into $L(w,b,\alpha)$

$$\boxed{L(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j}$$

4) SVM dual problem

$$\max_{\alpha} L(\alpha)$$

s.t. $\alpha_i \geq 0 \quad \forall i$

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

5) Given α^*,

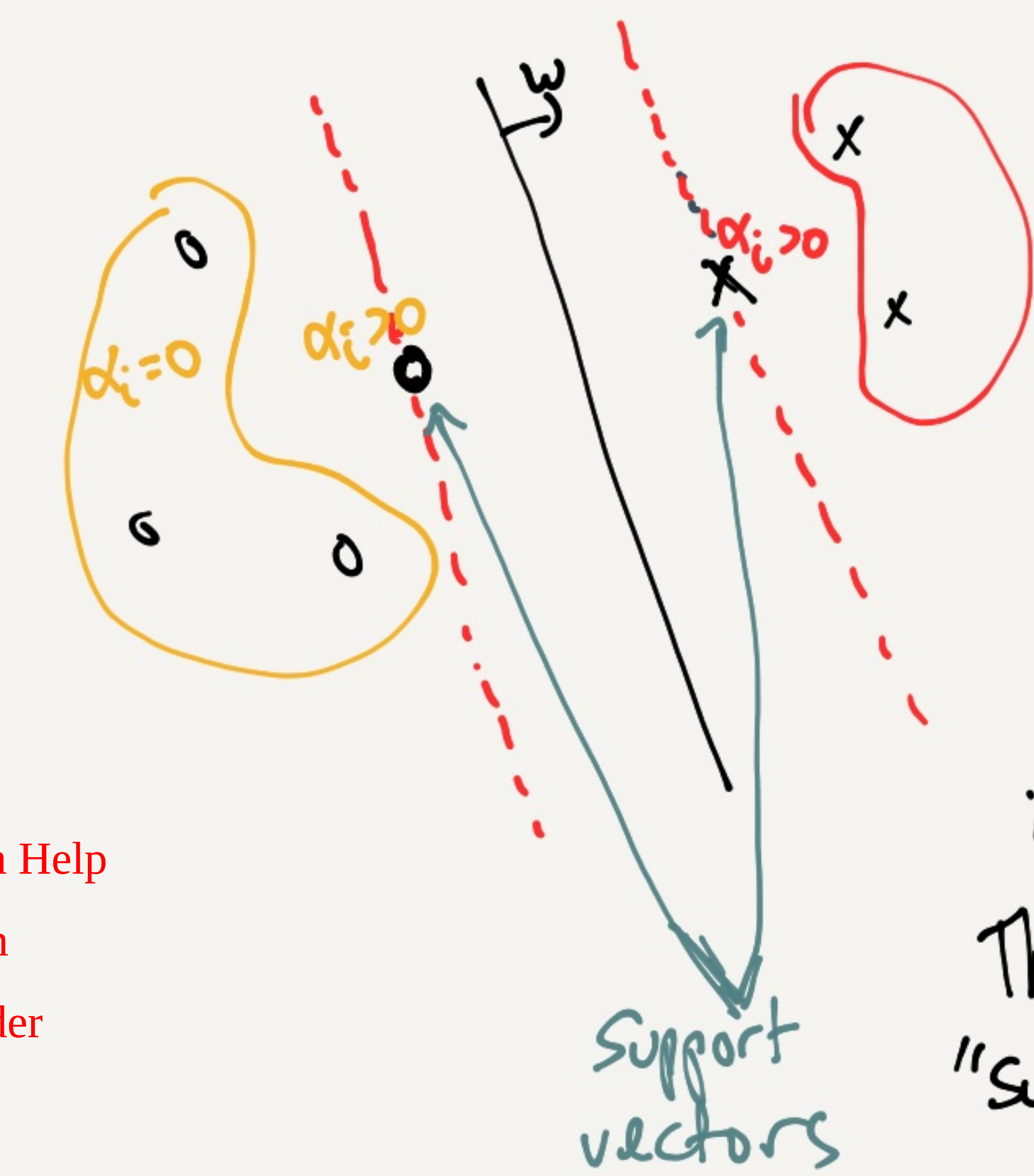
$$\boxed{w^* = \sum_{i=1}^{N} \alpha_i^* y_i x_i}$$

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Recall KKT conditions:

1) $g(x) = 0$ (active): $\alpha_i > 0$, $y_i(w^T x_i + b) - 1 = 0 \Rightarrow$
$y_i(w^T x_i + b) = 1 \Rightarrow x_i$ is on margin

2) $g(x) > 0$ (inactive): $\alpha_i = 0$, $y_i(w^T x_i + b) - 1 > 0$
$\Rightarrow y_i(w^T x_i + b) > 1 \Rightarrow x_i$ is not on margin.



Note: w^* only depends on the x_i's with $\alpha_i > 0$, i.e. on the margin.
These points are the "support vectors".

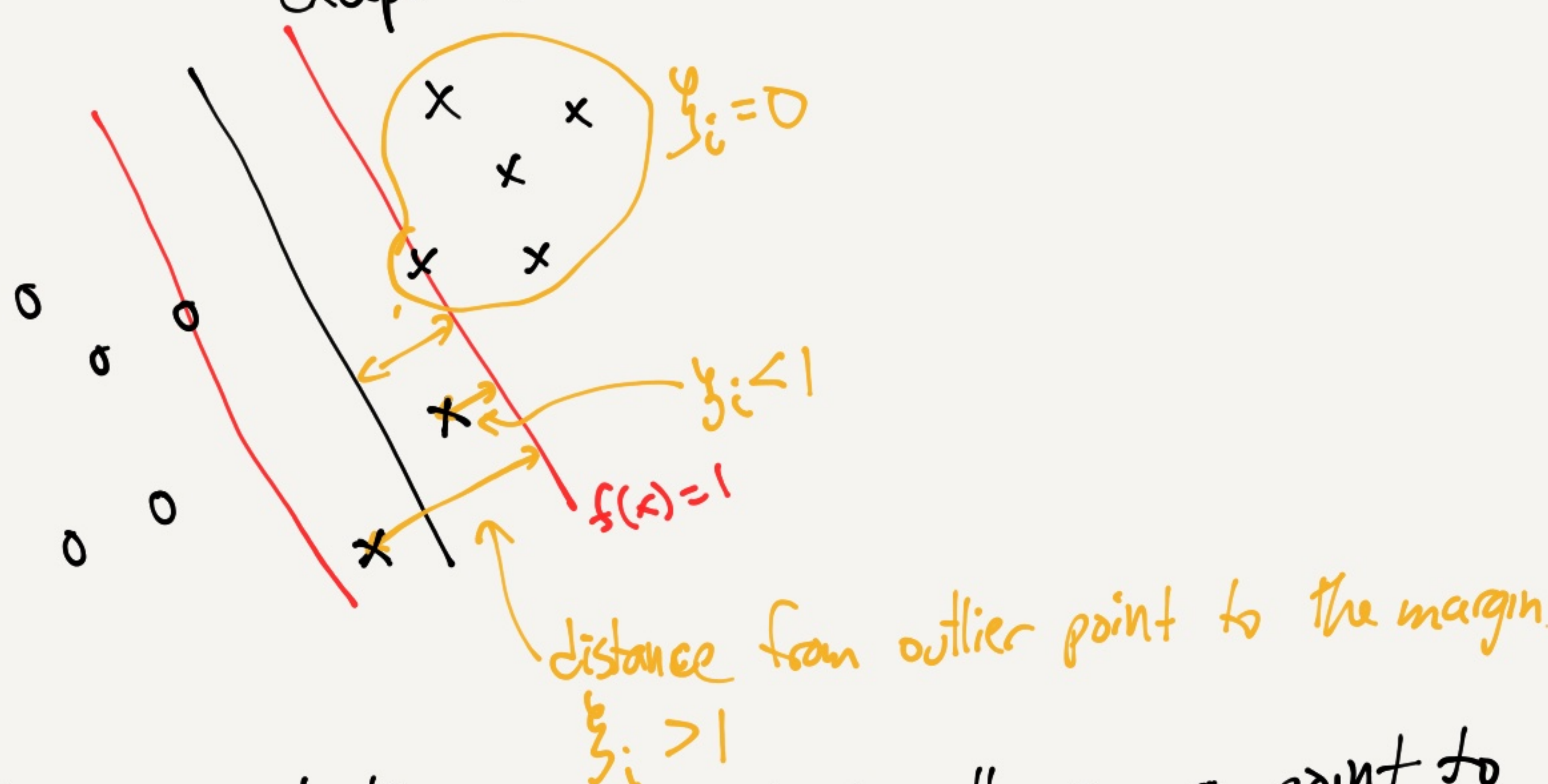
Note: the primal solves $w \in \mathbb{R}^D$
the dual solves $\alpha \in \mathbb{R}^N$

primal more efficient when $D < N$, and vice versa.

What about when the data is not linearly separable?

Soft-SVM

"soft-margin" – most points are on the margin or beyond, except a few outliers.



- ξ_i is a slack variable, which allows a point to violate the margin. $\xi_i \geq 0$

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

- New margin constraint: $y_i(w^T x_i + b) \geq (1 - \xi_i)$

  $\xi_i \geq 0$ margin becomes <1 when $\xi_i > 0$

- New objective & problem

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

penalty on large ξ_i, i.e. too many points violating margin.

$$\text{s.t. } y_i(w^T x_i + b) \geq 1 - \xi_i, \ \forall i$$

$$\xi_i \geq 0, \ \forall i$$

Dual SVM

$$\max_{\alpha \geq 0} \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$\text{s.t. } \sum_i \alpha_i y_i = 0$$

$$0 \leq \alpha_i \underbrace{\leq C}, \ \forall i$$

new constraint on max value of α_i.