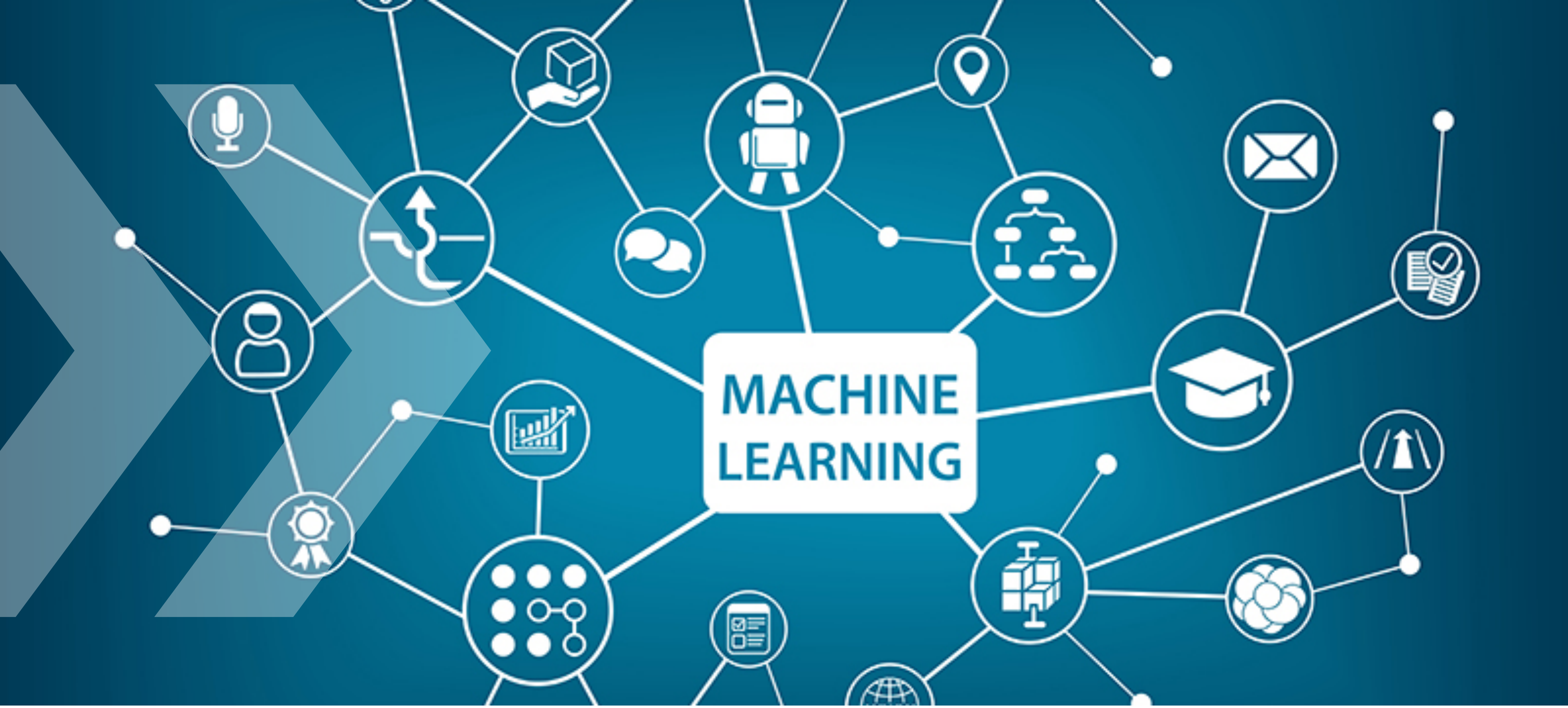


CS480/680: Intro to ML

Lecture 07: Hard-margin SVM

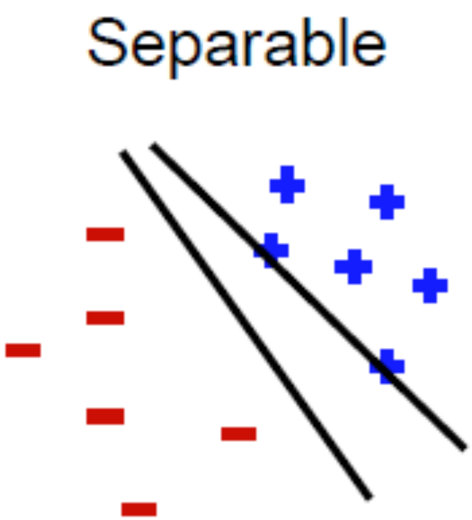
UNIVERSITY OF WATERLOO

Outline

- Maximum Margin
- Lagrangian Dual
- Alternative View

Perceptron revisited

- Two classes: y = 1 or y = -1



- Assuming linearly separable
  - exist **w** and b such that for all i,

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) > 0$$

- Find any such **w** and b

$$\min_{\mathbf{w},b} \; 0$$

$$\text{s.t. } \forall i, \; y_i(\mathbf{w}^\top \mathbf{x}_i + b) > 0$$

→ feasibility problem

Margin

- Take any linear separating hyperplane H

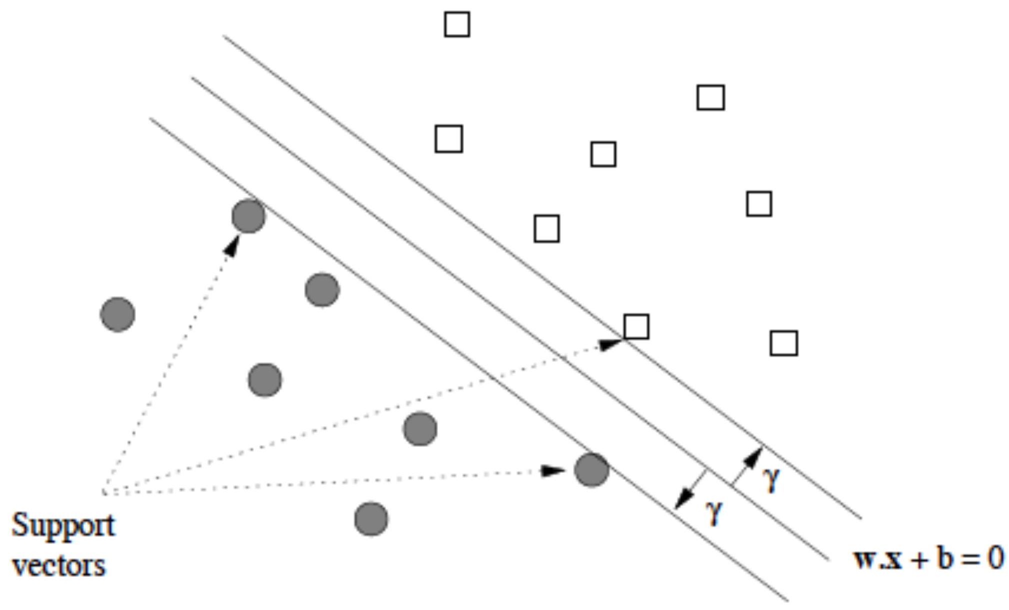
  for all i, $y_i(\mathbf{w}^T\mathbf{x}_i + b) > 0$

- Move H until it touches some positive point, H_1

  increase b say

- Move H until it touches some negative point, H_{-1}

  decrease b say



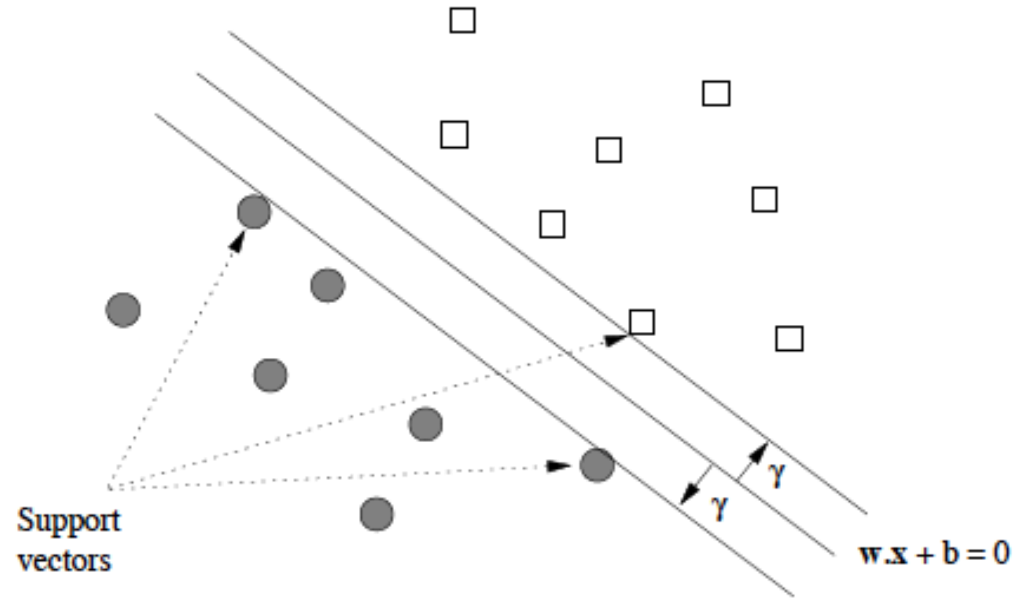
margin = $\mathrm{dist}(H_1,H) \wedge \mathrm{dist}(H_{-1},H)$

Put into formula

H : $\mathbf{w}^T\mathbf{x} + b = 0$

H_1 : $\mathbf{w}^T\mathbf{x} + b = t$

H_{-1} : $\mathbf{w}^T\mathbf{x} + b = -s$



What is the distance between H_1 and H ?

$$\min_{\mathbf{x},\mathbf{z}} \|\mathbf{x} - \mathbf{z}\|_2 \geq \left\| \frac{\mathbf{w}^\top}{\|\mathbf{w}\|_2}(\mathbf{x} - \mathbf{z}) \right\|_2 = \frac{|t|}{\|\mathbf{w}\|_2}$$

$$\text{s.t. } \mathbf{w}^\top \mathbf{x} + b = t$$

$$\mathbf{w}^\top \mathbf{z} + b = 0$$

equality is attained

$$\mathbf{x} = \frac{\mathbf{w}}{\|\mathbf{w}\|_2^2}(t - b)$$

$$\mathbf{z} = \frac{\mathbf{w}}{\|\mathbf{w}\|_2^2}(-b)$$

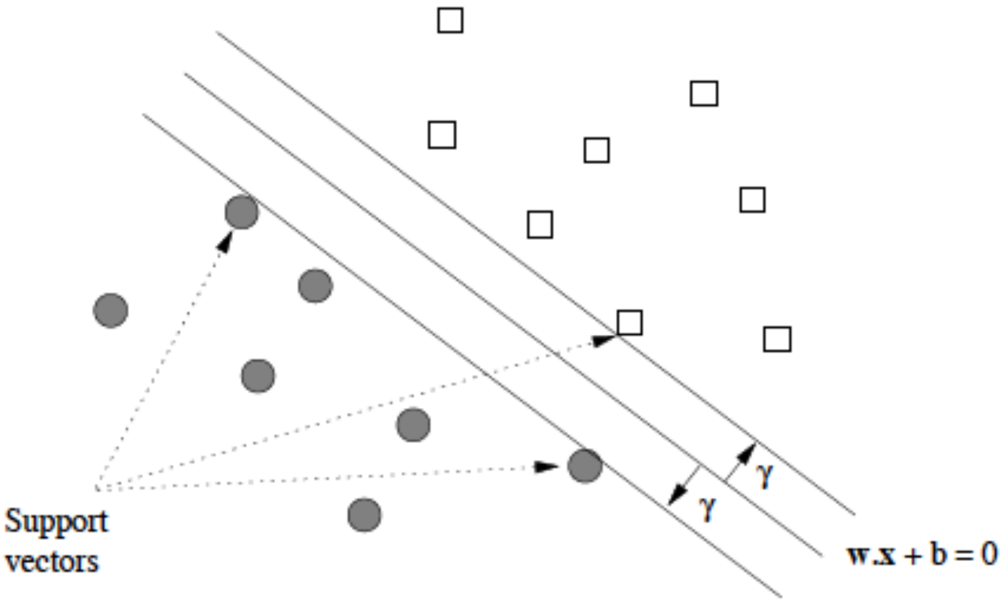
Put into formula

$H \;:\; \mathbf{w}^T\mathbf{x} + b = 0$

$H_1 :\; \mathbf{w}^T\mathbf{x} + b = t$

$H_{-1} :\; \mathbf{w}^T\mathbf{x} + b = -t$

What is the distance between H_1 and H ?



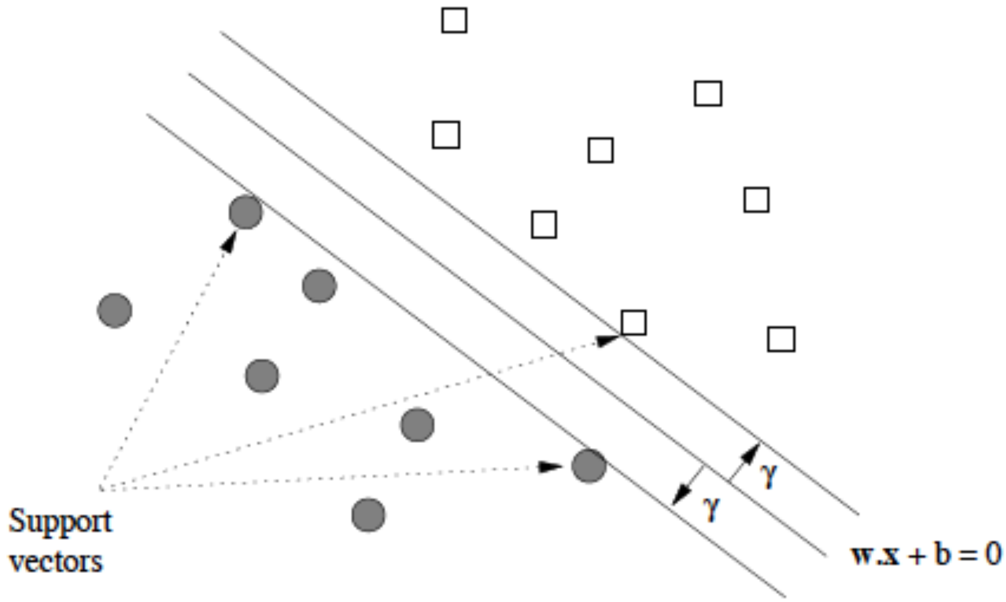
Put into formula

$$H \;:\; \mathbf{w}^T\mathbf{x} + b = 0$$

$$H_1 \;:\; \mathbf{w}^T\mathbf{x} + b = 1$$

$$H_{-1} \;:\; \mathbf{w}^T\mathbf{x} + b = -1$$

What is the distance between H_1 and H ?



Maximum Margin

$$\max_{\mathbf{w},b} \frac{1}{\|\mathbf{w}\|_2}$$

$$\text{s.t. } \forall i, y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1$$

Important facts.

- For any *f*, $\max_{\mathbf{w}} f(\mathbf{w}) = -\min_{\mathbf{w}} -f(\mathbf{w})$
- For positive *f*, $\max_{\mathbf{w}} \frac{1}{f(\mathbf{w})} = \frac{1}{\min_{\mathbf{w}} f(\mathbf{w})}$
- For s. monotone *g*, $\min_{\mathbf{w}} f(\mathbf{w}) \equiv \min_{\mathbf{w}} g(f(\mathbf{w}))$

Hard-margin Support Vector Machines

$$\max_{\mathbf{w},b} \frac{1}{\|\mathbf{w}\|_2}$$

$$\text{s.t. } \forall i, y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1$$

Quadratic programming

margin

$$\frac{1}{\|\mathbf{w}\|_2}$$

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|_2^2$$

$$\text{s.t. } \forall i, y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1$$

Comparing to perceptron

Linear programming

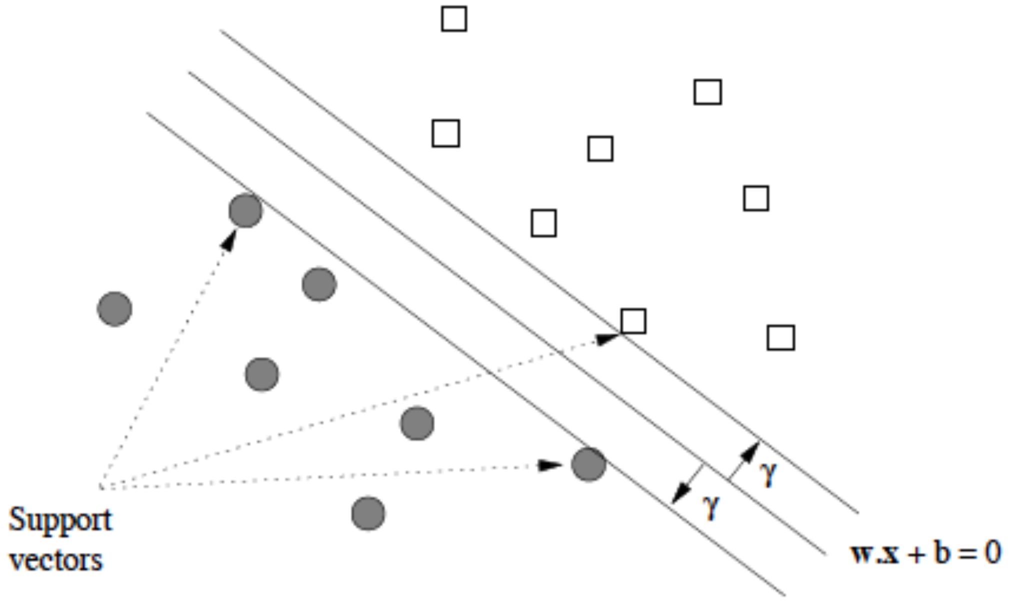
↑

$$\min_{\mathbf{w},b} \; 0$$

$$\text{s.t. } \forall i, \;\; y_i(\mathbf{w}^\top \mathbf{x}_i + b) > 0$$

Support Vectors

- Those touch the parallel hyperplanes H_1 and H_{-1}
- Usually only a handful
- Entirely determine the hyperplanes!



Existence and uniqueness

$$\min_{\mathbf{w},b} \ \frac{1}{2}\|\mathbf{w}\|_2^2$$
$$\text{s.t. } \forall i, y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1$$

- Always exists a minimizer **w** and b (if linearly separable)
- The minimizer **w** is unique (strict convexity of $\frac{1}{2}\|\mathbf{w}\|_2^2$)
- The minimizer b is also unique (why?)

Outline

- Maximum Margin
- Lagrangian Dual
- Alternative View

Lagrangian

Primal

$$\min_{\mathbf{w},b} \ \frac{1}{2}\|\mathbf{w}\|_2^2$$

$$\text{s.t. } \forall i, y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1$$

Lagrangian

$$\min_{\mathbf{w},b} \max_{\alpha \geq 0} \frac{1}{2}\|\mathbf{w}\|_2^2 - \sum_i \alpha_i [y_i(\mathbf{w}^\top \mathbf{x}_i + b) - 1]$$

[primal variable]

Lagrangian multiplier
[dual variable]

Deriving the dual

$$\min_{\mathbf{w},b} \max_{\alpha \geq 0} \frac{1}{2}\|\mathbf{w}\|_2^2 - \sum_i \alpha_i [y_i(\mathbf{w}^\top \mathbf{x}_i + b) - 1]$$

$$\max_{\alpha \geq 0} \min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|_2^2 - \sum_i \alpha_i [y_i(\mathbf{w}^\top \mathbf{x}_i + b) - 1]$$

$$\frac{\partial}{\partial b} = \sum_i \alpha_i y_i = 0$$

$$\frac{\partial}{\partial \mathbf{w}} = \mathbf{w} - \sum_i \alpha_i y_i \mathbf{x}_i = 0$$

The dual problem

$$\max_{\alpha \geq 0} \sum_i \alpha_i - \frac{1}{2}\Big\| \sum_i \alpha_i y_i \mathbf{x}_i \Big\|_2^2$$

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

Only need dot product in the dual !

Dual

$$\min_{\alpha \geq 0} \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j - \sum_k \alpha_k$$

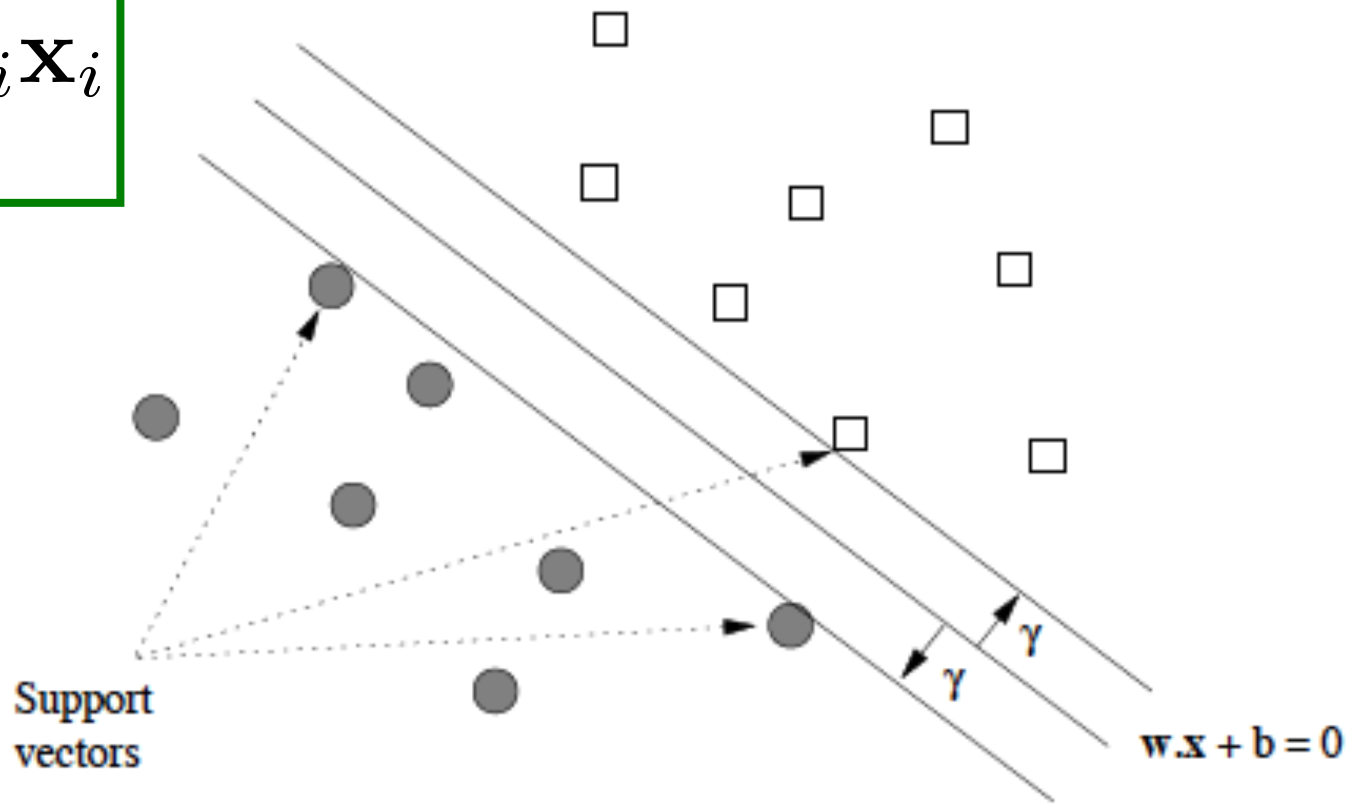
$$\text{s.t.} \sum_i \alpha_i y_i = 0$$

$\mathbf{R}^n$

Support Vectors

$$\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i$$

$$\alpha_i > 0$$



Outline

- Maximum Margin

- Dual

- Alternative View

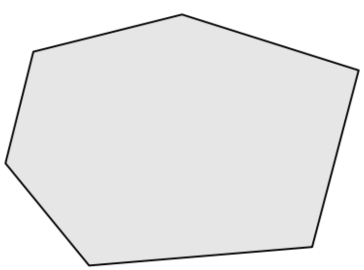
An dual view

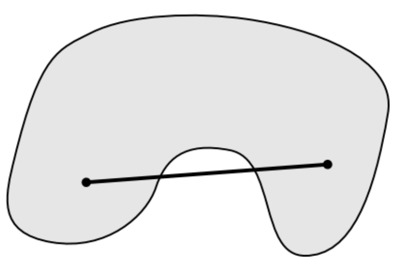
Convex sets and Convex hull

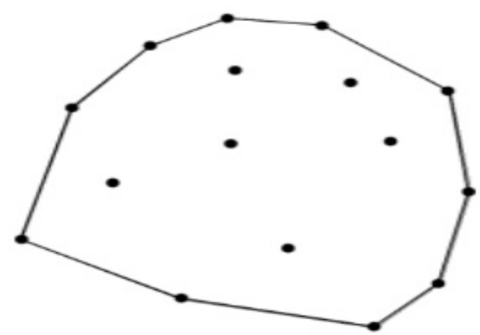
Convex set. A point set $C \in \mathbf{R}^d$ is convex if the line segment [x,y] connecting any two points x and y in C lies entirely in C.

Convex hull. Smallest convex set containing C.

$$\mathrm{ch}(C) := \left\{ \sum_i \alpha_i \mathbf{x}_i : \mathbf{x}_i \in C, \alpha_i \geq 0, \sum_i \alpha_i = 1 \right\}.$$





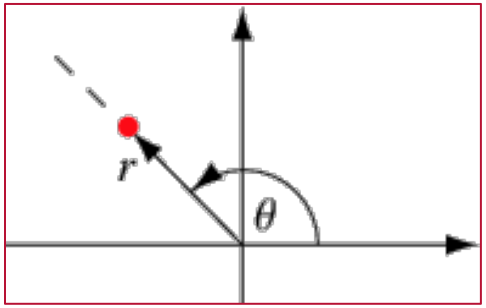


Separating scale from direction

$$\min_{\alpha \geq 0} \frac{1}{2} \left\| \sum_i \alpha_i y_i \mathbf{x}_i \right\|_2^2 - \sum_i \alpha_i$$

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

$$\alpha = r\bar{\alpha}$$



$$\min_{r \geq 0} \min_{\bar{\alpha} \in 2\Delta} \frac{r^2}{2} \left\| \sum_i \bar{\alpha}_i y_i \mathbf{x}_i \right\|_2^2 - 2r$$

$$\text{s.t.} \sum_i \bar{\alpha}_i y_i = 0$$

$$r = \frac{2}{\| \sum_i \bar{\alpha}_i y_i \mathbf{x}_i \|_2^2}$$

Scale r does not matter

$$\min_{\bar{\alpha} \in 2\Delta} \frac{\cancel{r^2}}{2} \Big\| \sum_i \bar{\alpha}_i y_i \mathbf{x}_i \Big\|_2^2 \cancel{- 2r}$$

$$\text{s.t.} \sum_i \bar{\alpha}_i y_i = 0$$

$$\min_{\mu \in \Delta, \nu \in \Delta} \frac{1}{\cancel{2}} \Big\| \sum_{i \in P} \mu_i \mathbf{x}_i - \sum_{j \in N} \nu_j \mathbf{x}_j \Big\|_2^{\cancel{2}}$$

$\text{conv}(\{\mathbf{x}_i : i \in P\})$ $\text{conv}(\{\mathbf{x}_i : i \in N\})$

$P := \{i : y_i = 1\}$

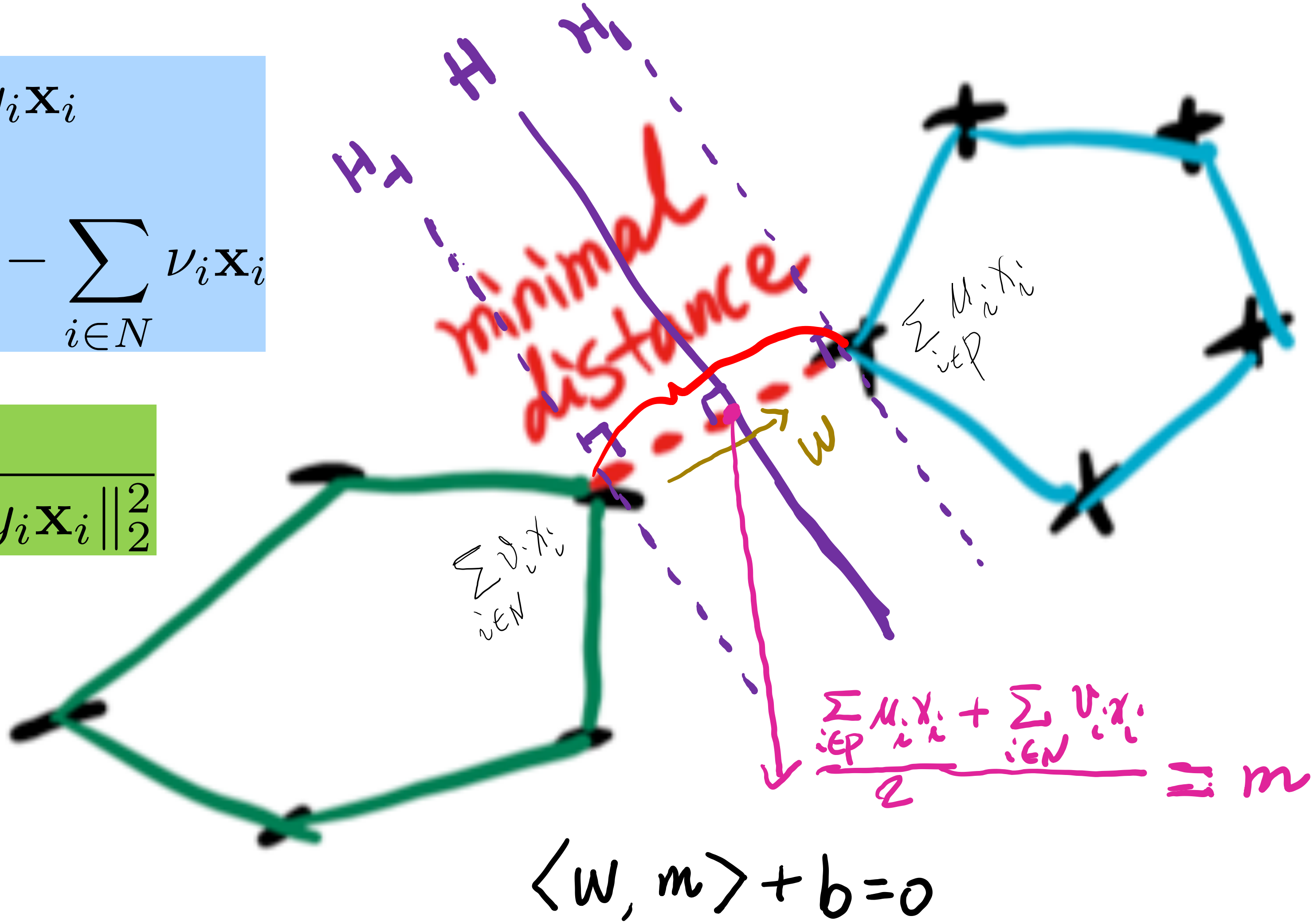
$N := \{i : y_i = -1\}$

$\bar{\alpha} = [\mu; \nu]$

NOW this

$$\mathbf{w} = r \sum_i \bar{\alpha}_i y_i \mathbf{x}_i \propto \sum_{i \in P} \mu_i \mathbf{x}_i - \sum_{i \in N} \nu_i \mathbf{x}_i$$

$$r = \frac{2}{\| \sum_i \bar{\alpha}_i y_i \mathbf{x}_i \|_2^2}$$



Questions?

UNIVERSITY OF WATERLOO