Announcements

Assignments

- HW6 (written + programming)
  - Due Thu 3/26, 11:59 pm
- HW7 (online)
  - Out later tonight
  - Due Tue 3/31, 11:59 pm

$x \rightarrow \phi(x) \longrightarrow \text{SVM} \longrightarrow y = +/- 1$

$x \longrightarrow \text{Net} \longrightarrow x$

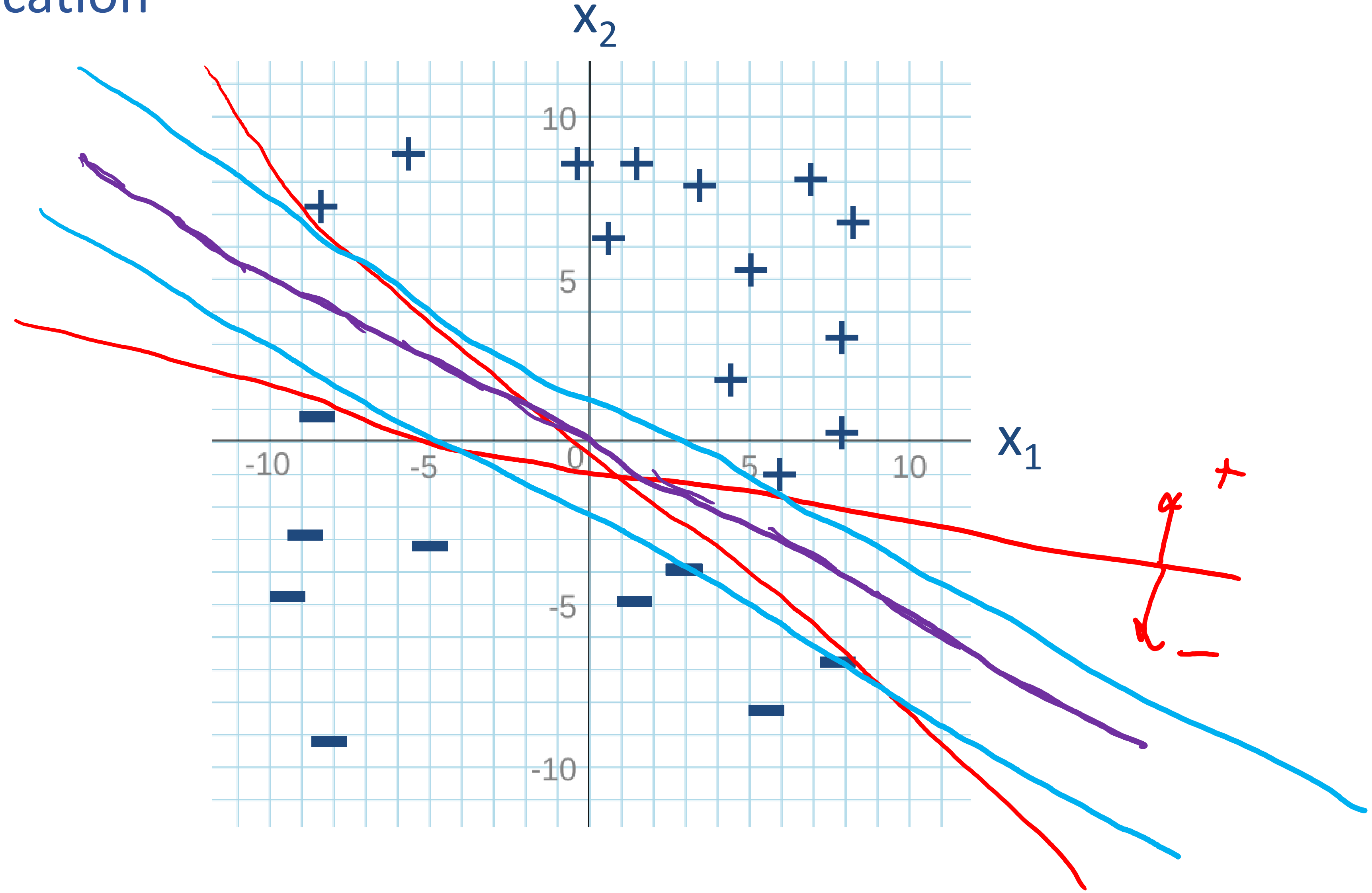
Introduction to Machine Learning

Support Vector Machines

Instructor: Pat Virtue

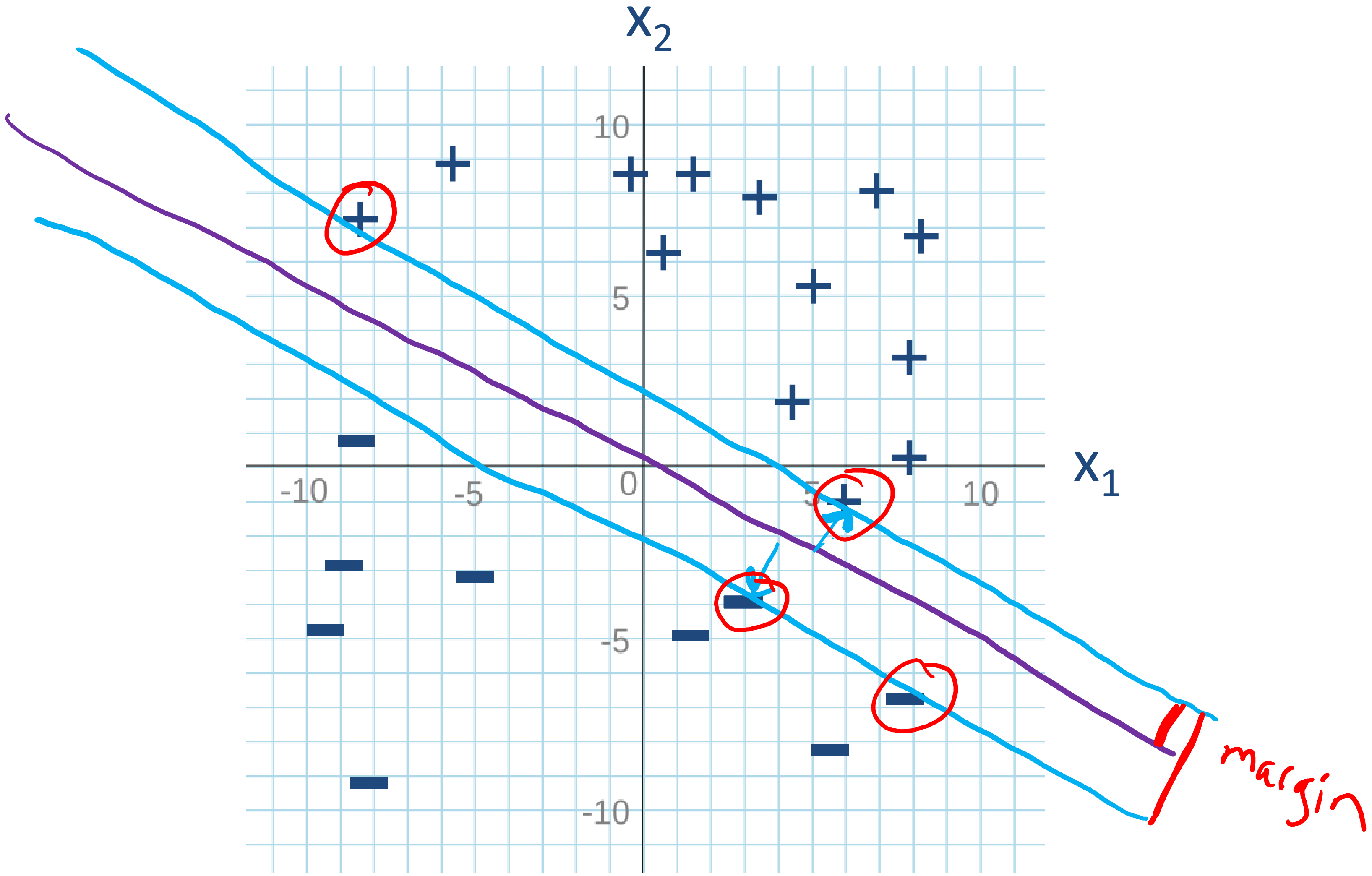
Support Vector Machines

Linear Classification



Support Vector Machines

Max Margin



Constrained Optimization

$$\min_x \quad f(x) \leftarrow \text{obj}$$

Linear Program

$$\min_{\boldsymbol{x}} \quad \boldsymbol{c}^T \boldsymbol{x}$$

$$\text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}$$

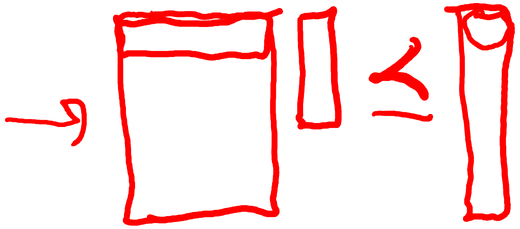
$$\underbrace{c_1 x_1 + c_2 x_2}_{} = y$$

$x \in \mathbb{R}^M$

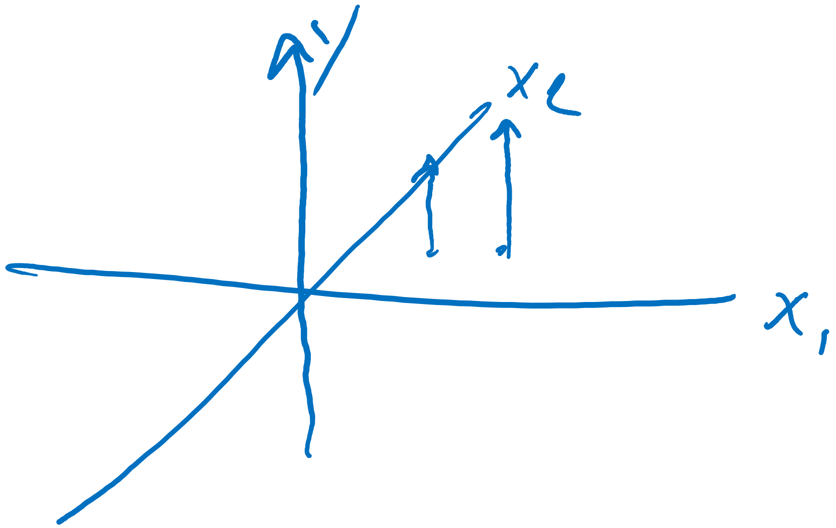
$c \in \mathbb{R}^M$

$b \in \mathbb{R}^N$

$A \in \mathbb{R}^{N \times M}$



$$a_{11} x_1 + a_{12} x_2 \le b_1$$



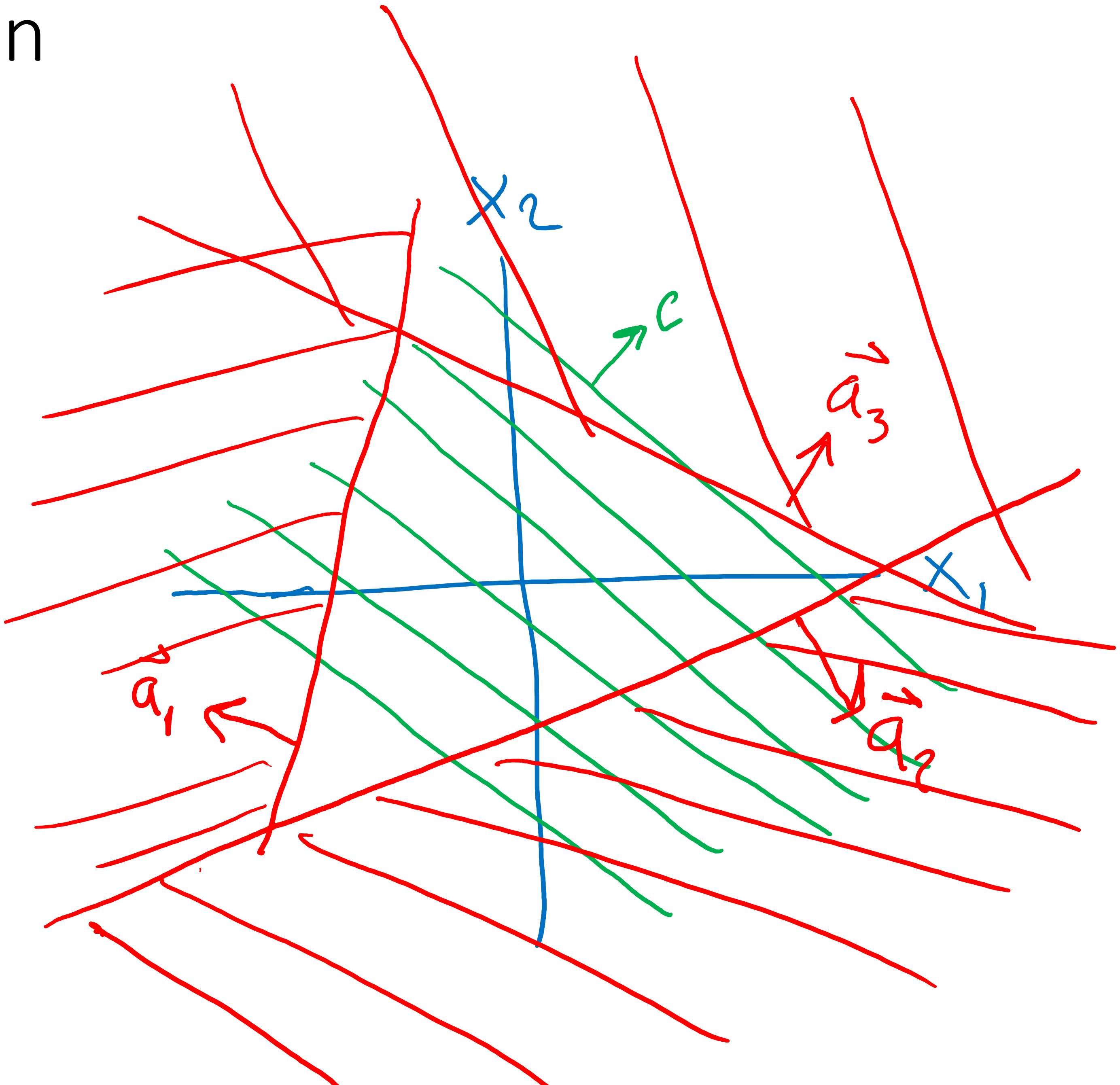
Constrained Optimization

Linear Program

$$\min_{\boldsymbol{x}} \quad \boldsymbol{c}^T \boldsymbol{x}$$
$$\text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} \preceq \boldsymbol{b}$$

$a_{11} x_1 + a_{12} x_2 \leq b_1$

$a_{21} x_1 + a_{22} x_2 \leq b_2$



Constrained Optimization

Linear Program

$$\min_{\boldsymbol{x}} \quad \boldsymbol{c}^T \boldsymbol{x}$$
$$\text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} \preceq \boldsymbol{b}$$

Solvers

- Simplex
- Interior point methods

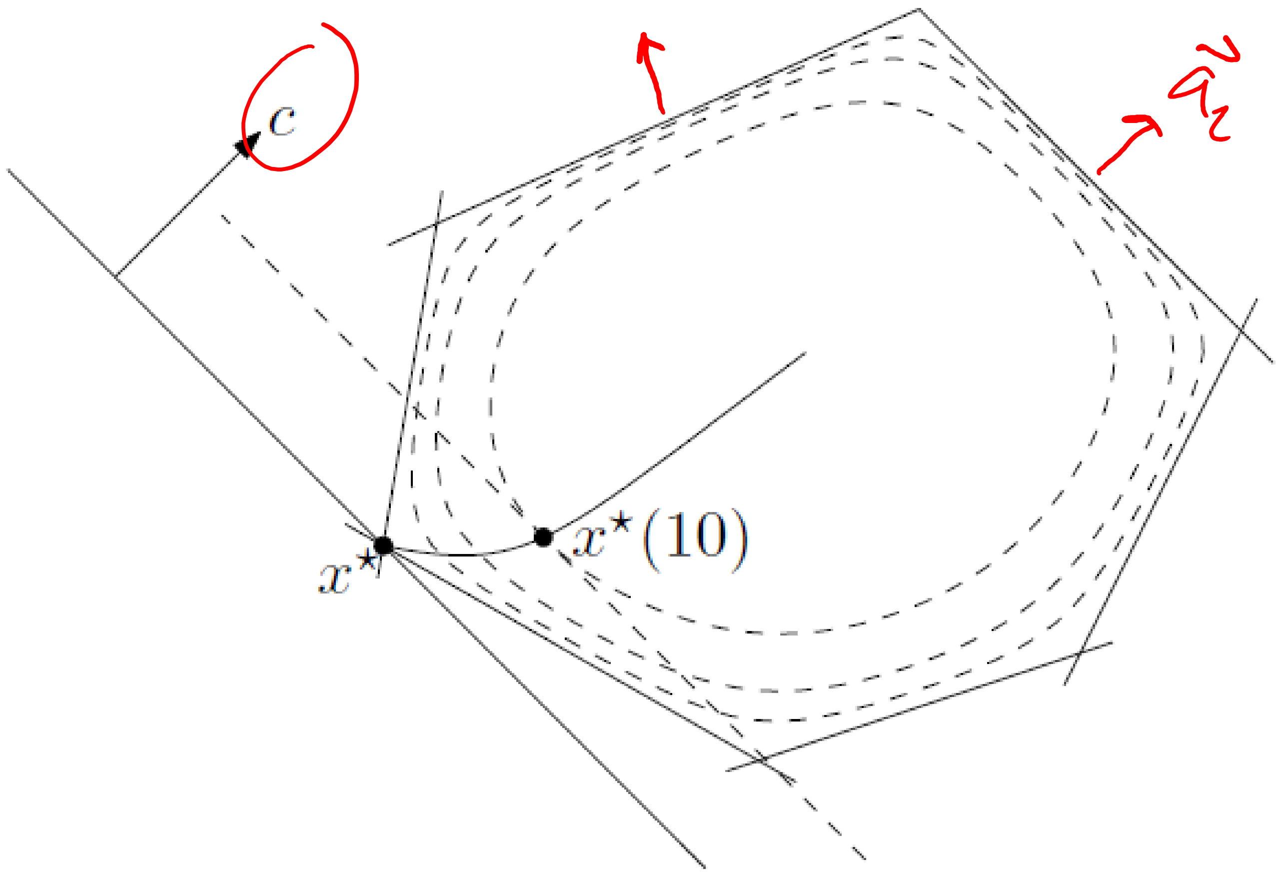


Figure: Fig 11.2 from Boyd and Vandenberghe, *Convex Optimization*

Constrained Optimization

$x \in \mathbb{R}^{M}$

$Q \in \mathbb{R}^{M \times M}$

Linear Program

$$\min_{\boldsymbol{x}} \quad \boldsymbol{c}^T \boldsymbol{x}$$
$$\text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} \preceq \boldsymbol{b}$$

Quadratic Program

$$\min_{\boldsymbol{x}} \quad \underbrace{\boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x}} + \underline{\boldsymbol{c}^T \boldsymbol{x}}$$
$$\text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} \preceq \boldsymbol{b}$$

$M = 1$

Solvers

- Simplex
- Interior point methods

$\|y - Xw\|_2^2$

$\rightarrow (y - Xw)^T (y - Xw)$

$\vdots$

$\underbrace{w^T X^T X w}$

Constrained Optimization

Linear Program

$$\min_{\boldsymbol{x}} \quad \boldsymbol{c}^T \boldsymbol{x}$$
$$\text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} \preceq \boldsymbol{b}$$

Quadratic Program

$$\min_{\boldsymbol{x}} \quad \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} + \boldsymbol{c}^T \boldsymbol{x}$$
$$\text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} \preceq \boldsymbol{b}$$

Solvers

- Simplex
- Interior point methods

Solvers

- Conjugate gradient
- Ellipsoid method
- Interior point methods

Constrained Optimization

Linear Program

$$\min_{\boldsymbol{x}} \quad \boldsymbol{c}^T \boldsymbol{x}$$
$$\text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} \preceq \boldsymbol{b}$$

Solvers

- Simplex
- Interior point methods

Quadratic Program

$$\min_{\boldsymbol{x}} \quad \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} + \boldsymbol{c}^T \boldsymbol{x}$$
$$\text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} \preceq \boldsymbol{b}$$

Special Case

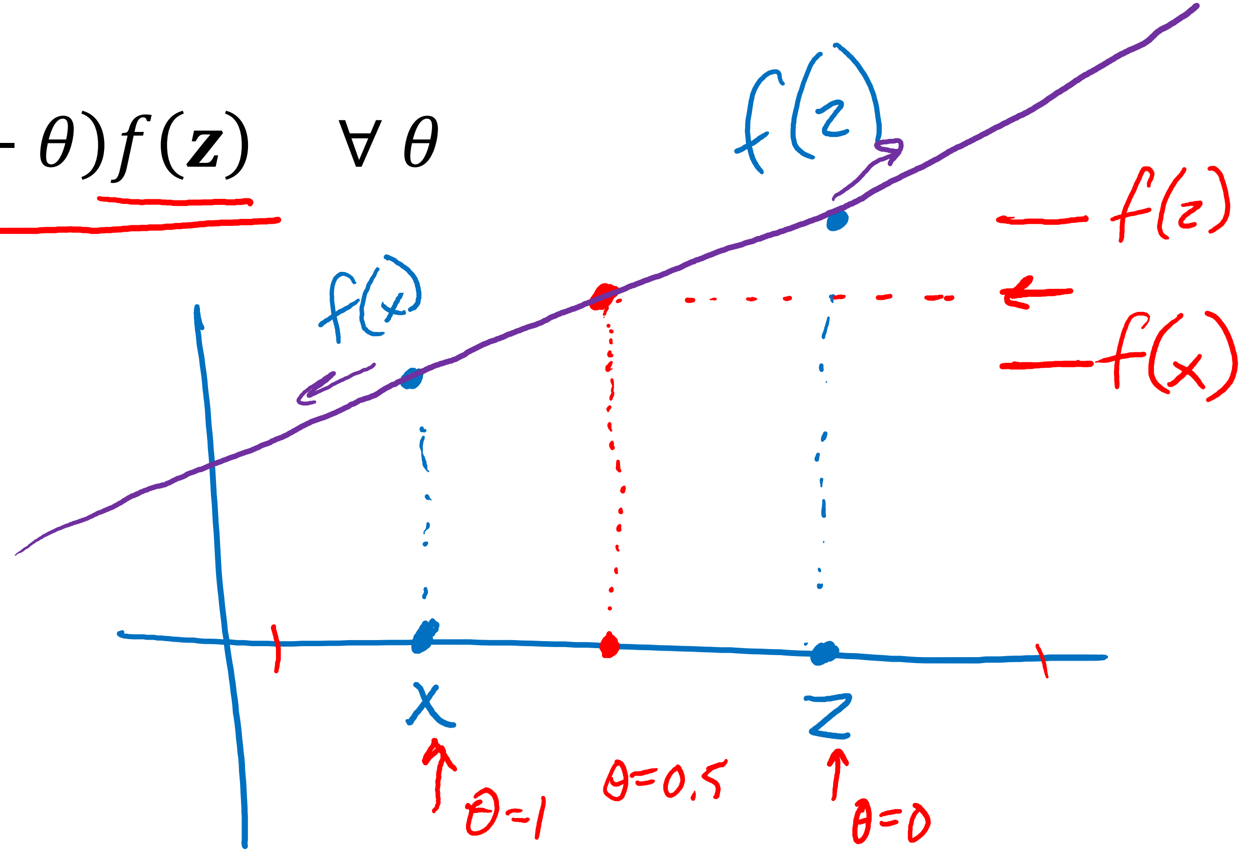
- If $\boldsymbol{Q}$ is positive-definite, the problem is convex
- $\boldsymbol{Q}$ is positive-definite if:
  $\boldsymbol{v}^T \boldsymbol{Q} \boldsymbol{v} > 0 \quad \forall\, \boldsymbol{v} \in \mathbb{R}^M \setminus \boldsymbol{0}$

Convex Optimization

Linear function

If $f(\boldsymbol{x})$ is linear, then:

- $f(\boldsymbol{x}+\boldsymbol{z}) = f(\boldsymbol{x}) + f(\boldsymbol{z})$
- $f(\theta\boldsymbol{x}) = \theta f(\boldsymbol{x}) \quad \forall\, \theta$
- $f(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{z}) = \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{z}) \quad \forall\, \theta$

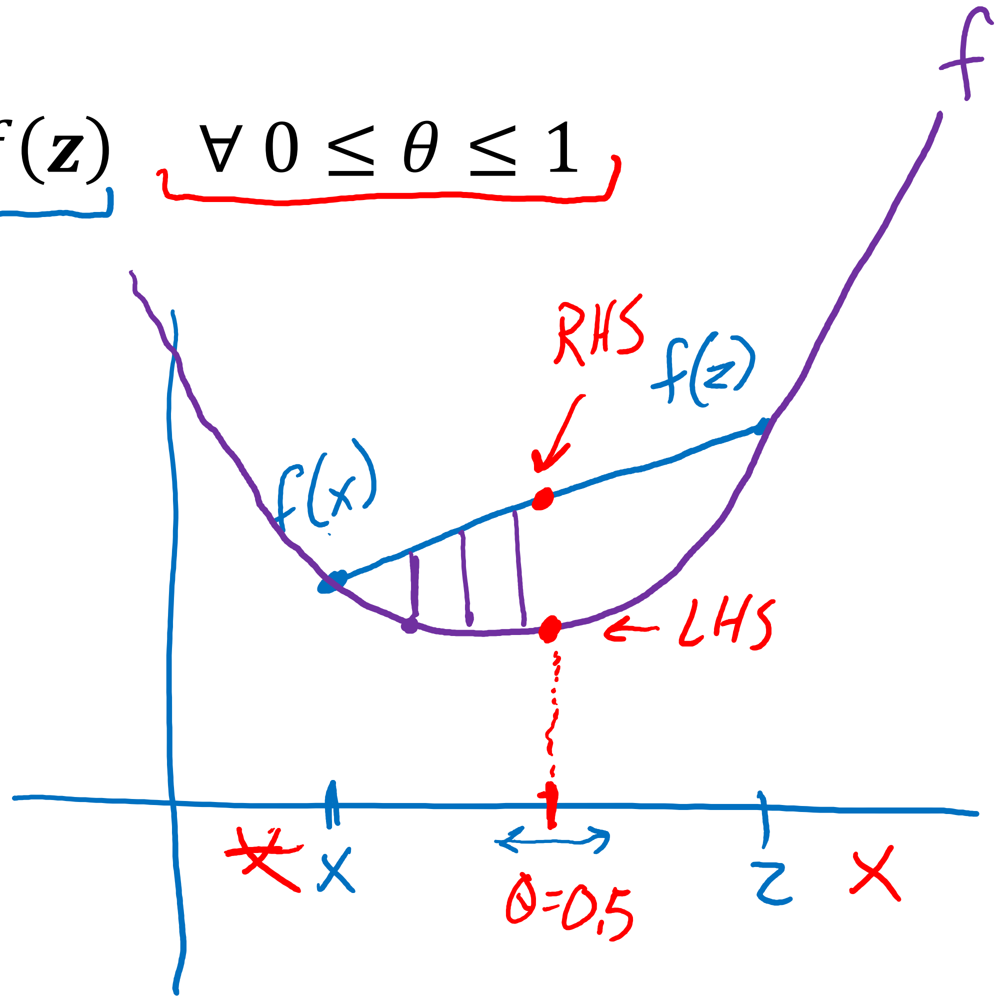


Convex Optimization

Convex function

If $f(\boldsymbol{x})$ is convex, then:

- $f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{z}) \leq \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{z}) \quad \forall\ 0 \leq \theta \leq 1$



Convex optimization

If $f(\boldsymbol{x})$ is convex, then:

- Every local minimum is also a global minimum ☺

Constrained Optimization

Linear Program

$$\min_{\boldsymbol{x}} \quad \boldsymbol{c}^T \boldsymbol{x}$$
$$\text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} \preceq \boldsymbol{b}$$

Quadratic Program

$$\min_{\boldsymbol{x}} \quad \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} + \boldsymbol{c}^T \boldsymbol{x}$$
$$\text{s.t.} \quad A\boldsymbol{x} \preceq \boldsymbol{b}$$

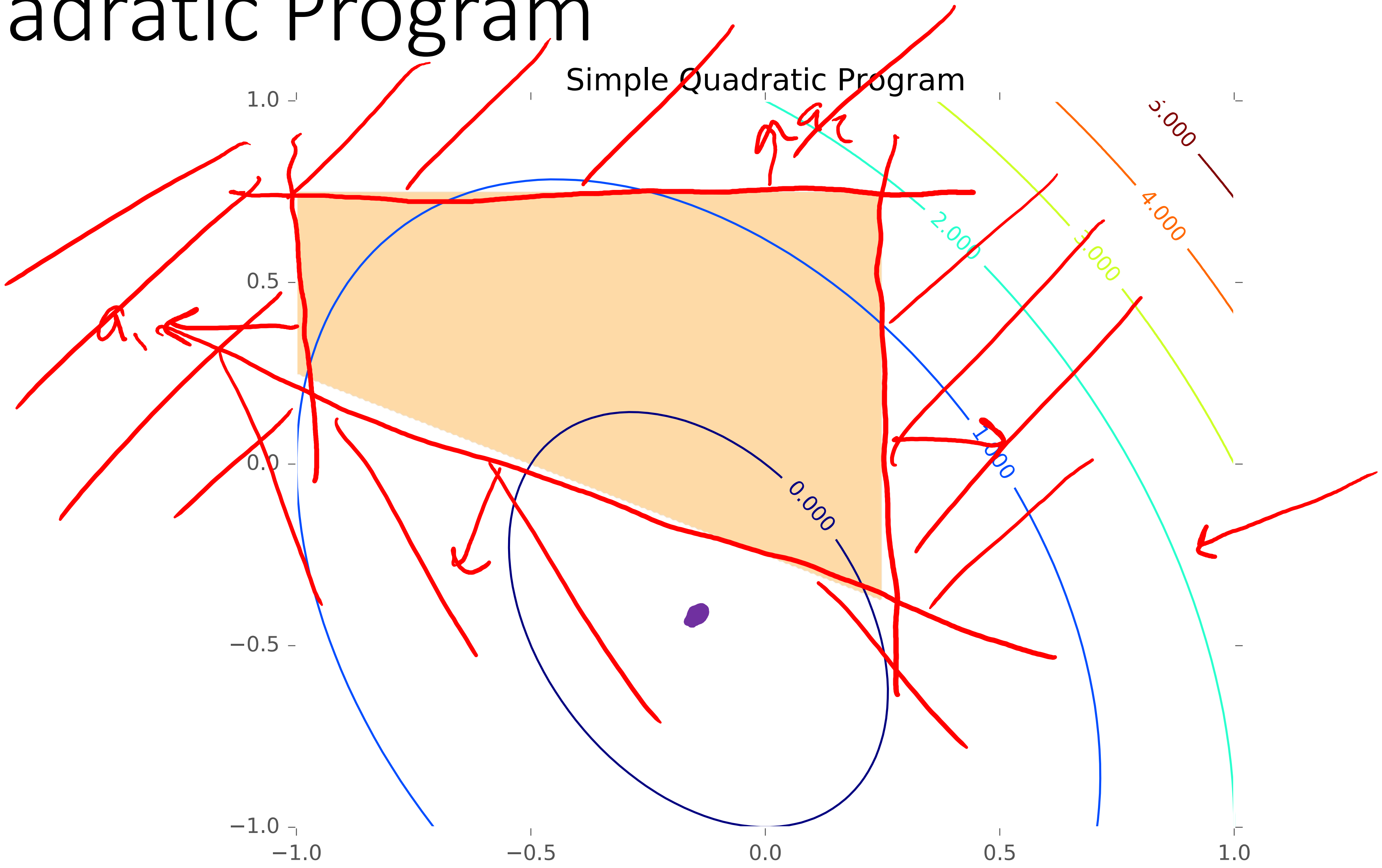
Solvers

- Simplex
- Interior point methods

Special Case

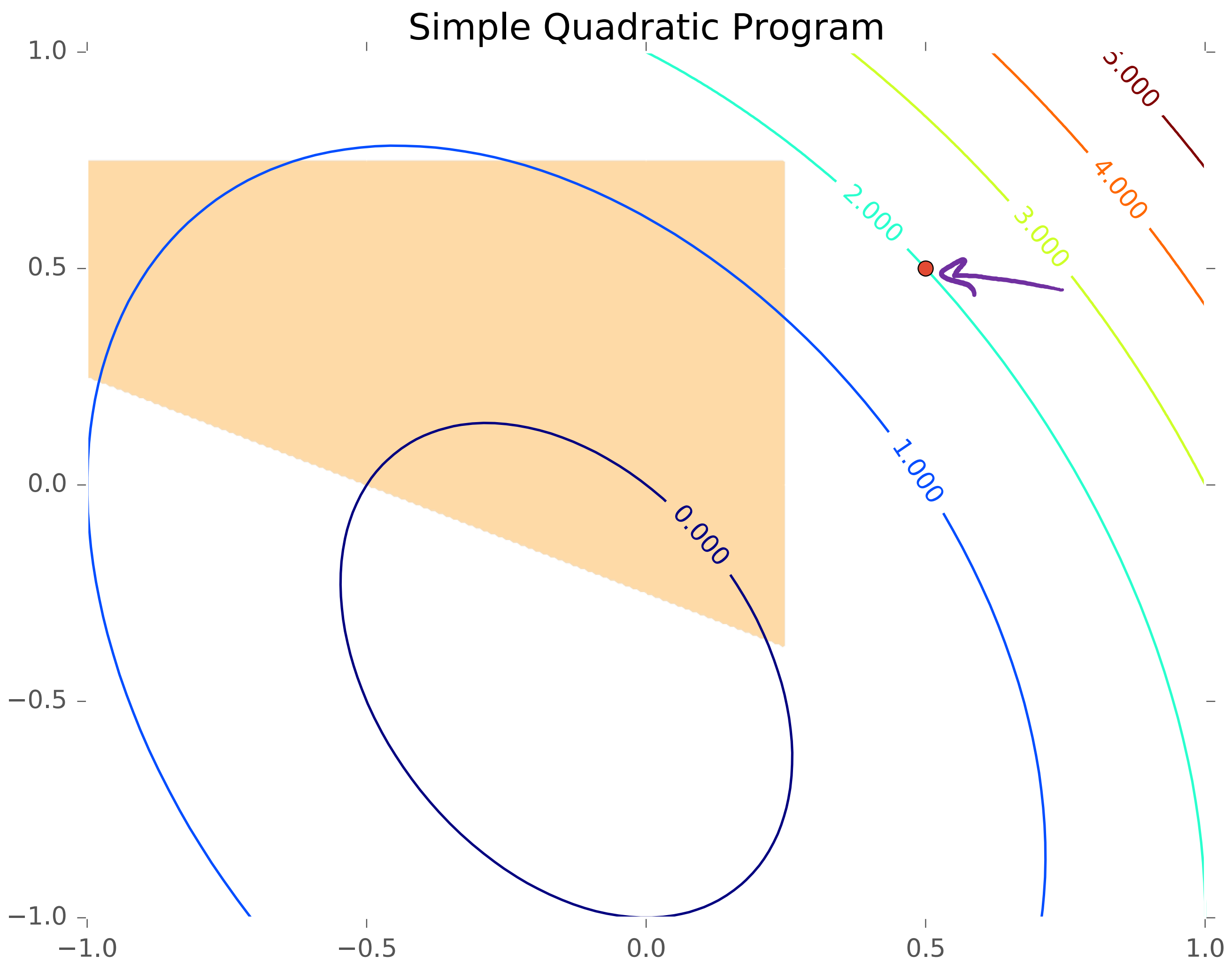
- If $\boldsymbol{Q}$ is positive-definite, the problem is convex
- $\boldsymbol{Q}$ is positive-definite if:
  $$\boldsymbol{v}^T \boldsymbol{Q} \boldsymbol{v} > 0 \quad \forall\, \boldsymbol{v} \in \mathbb{R}^M \backslash \boldsymbol{0}$$

Quadratic Program



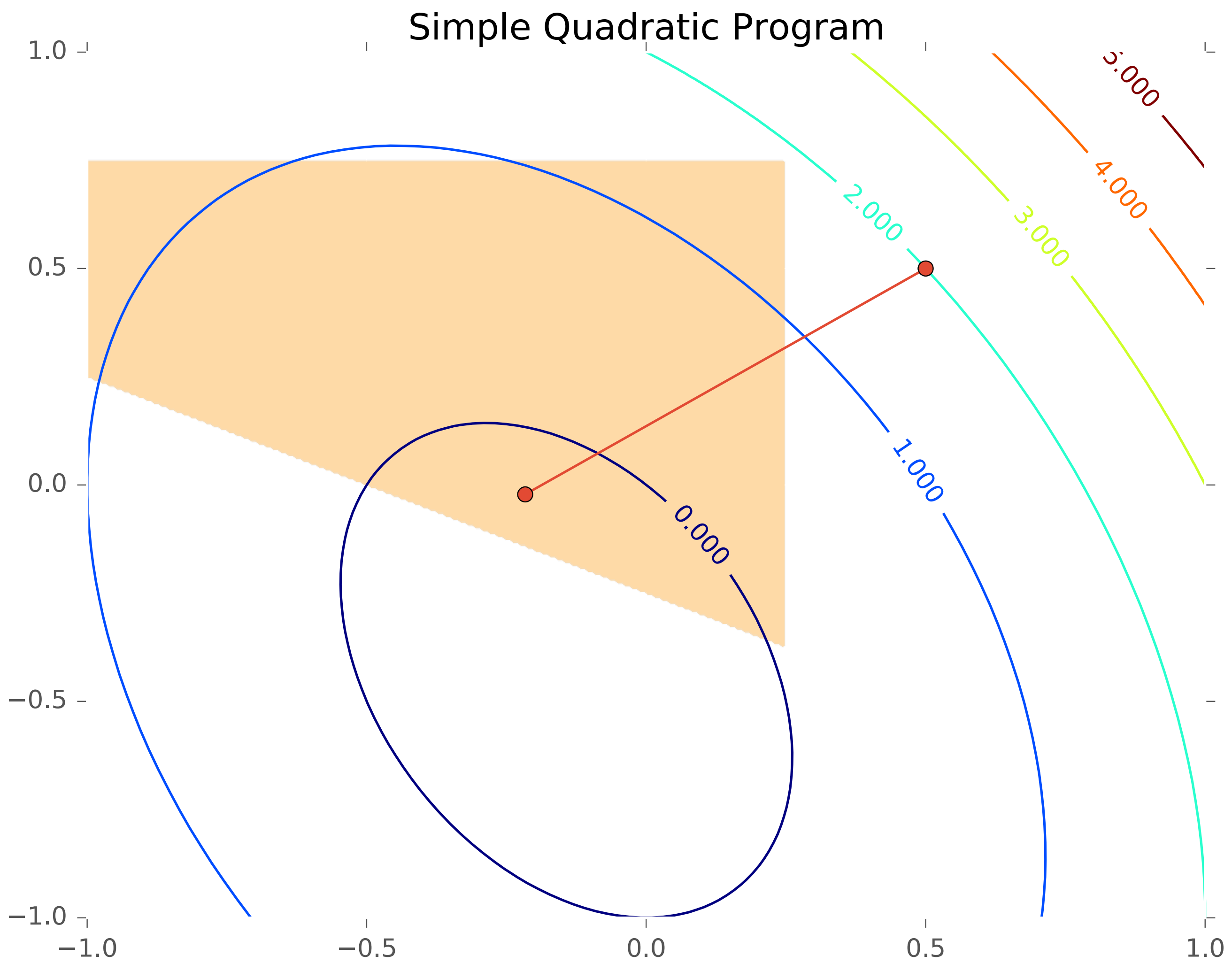
Slide credit: CMU MLD Matt Gormley

Quadratic Program

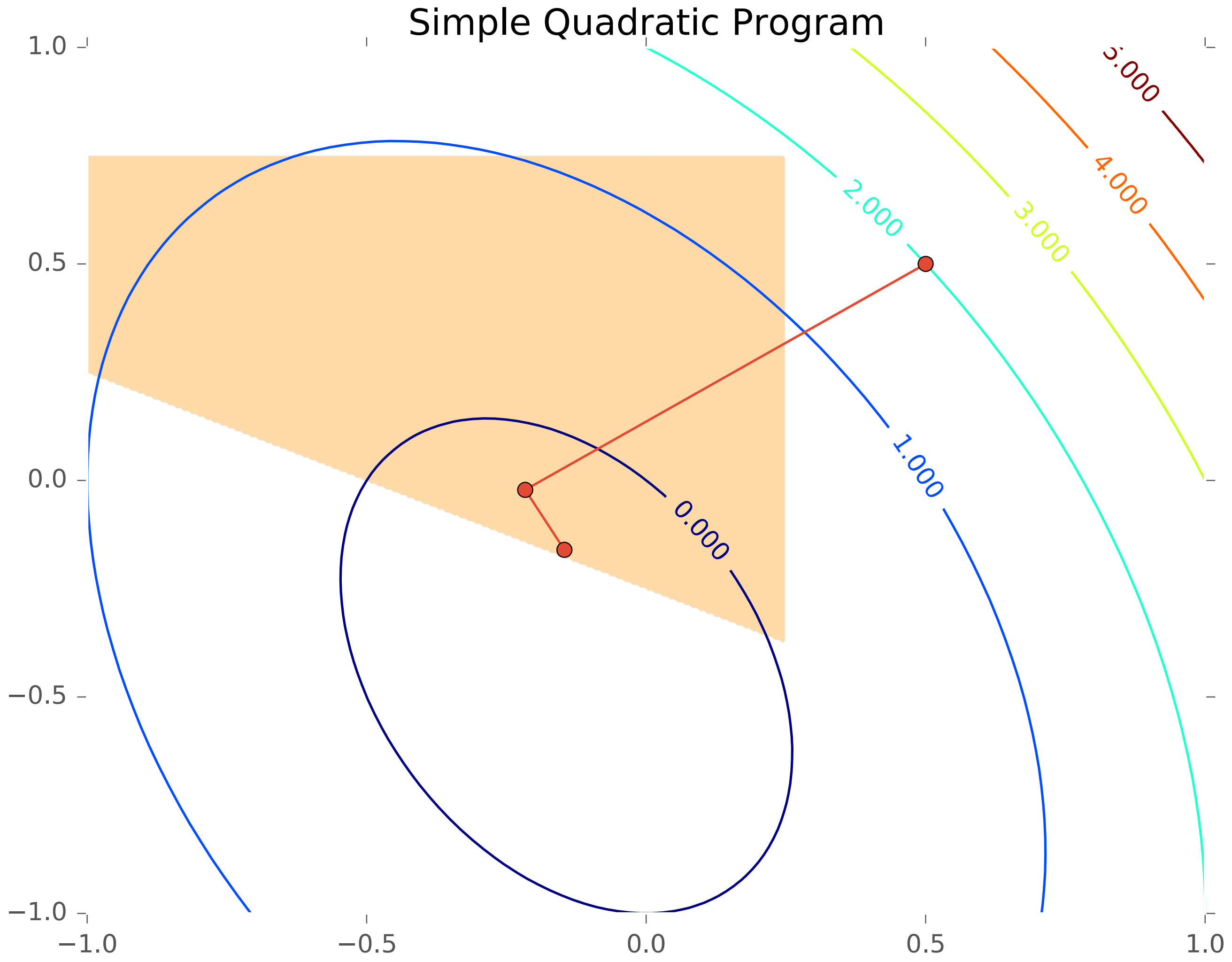


Slide credit: CMU MLD Matt Gormley

Quadratic Program

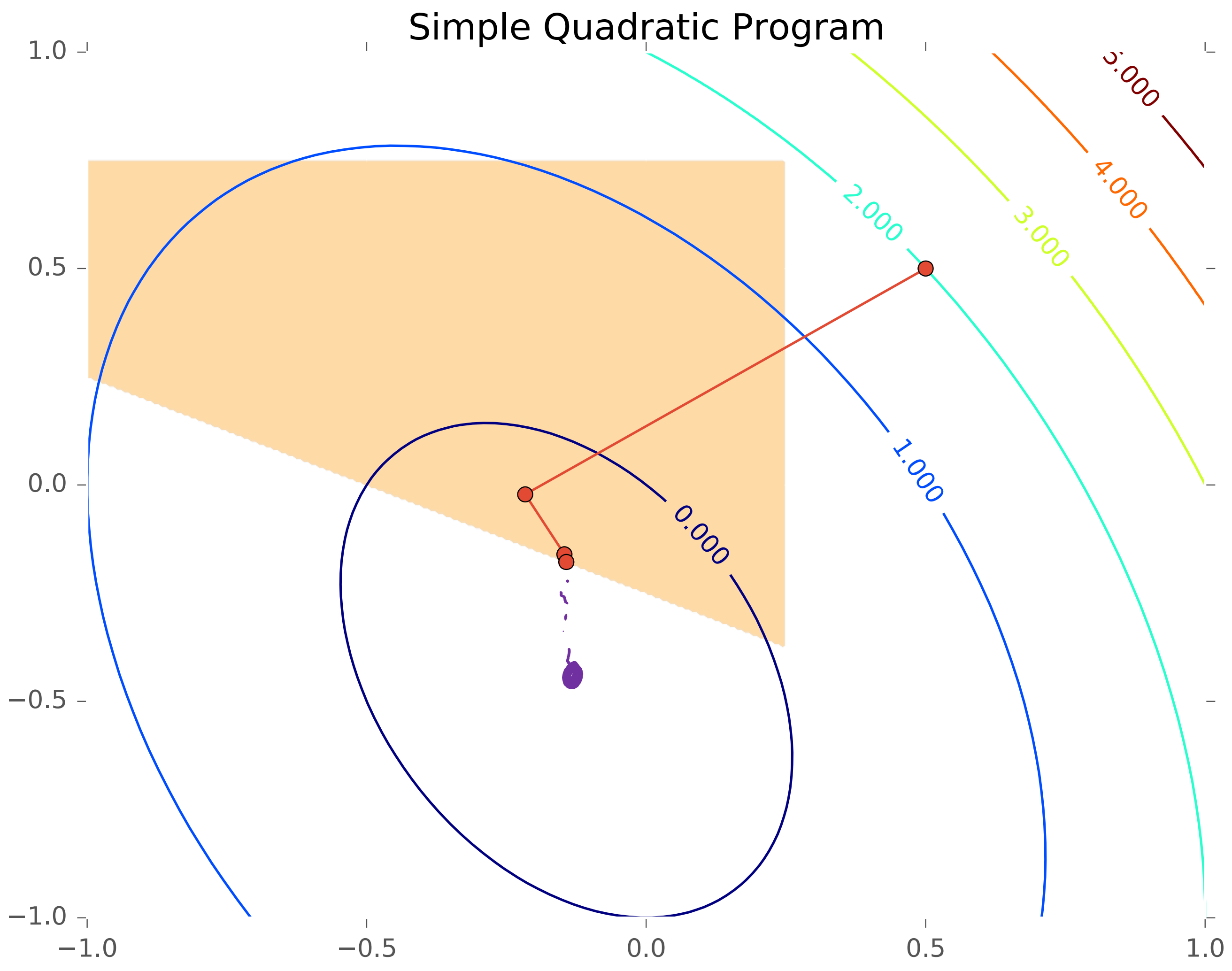


Quadratic Program



Slide credit: CMU MLD Matt Gormley

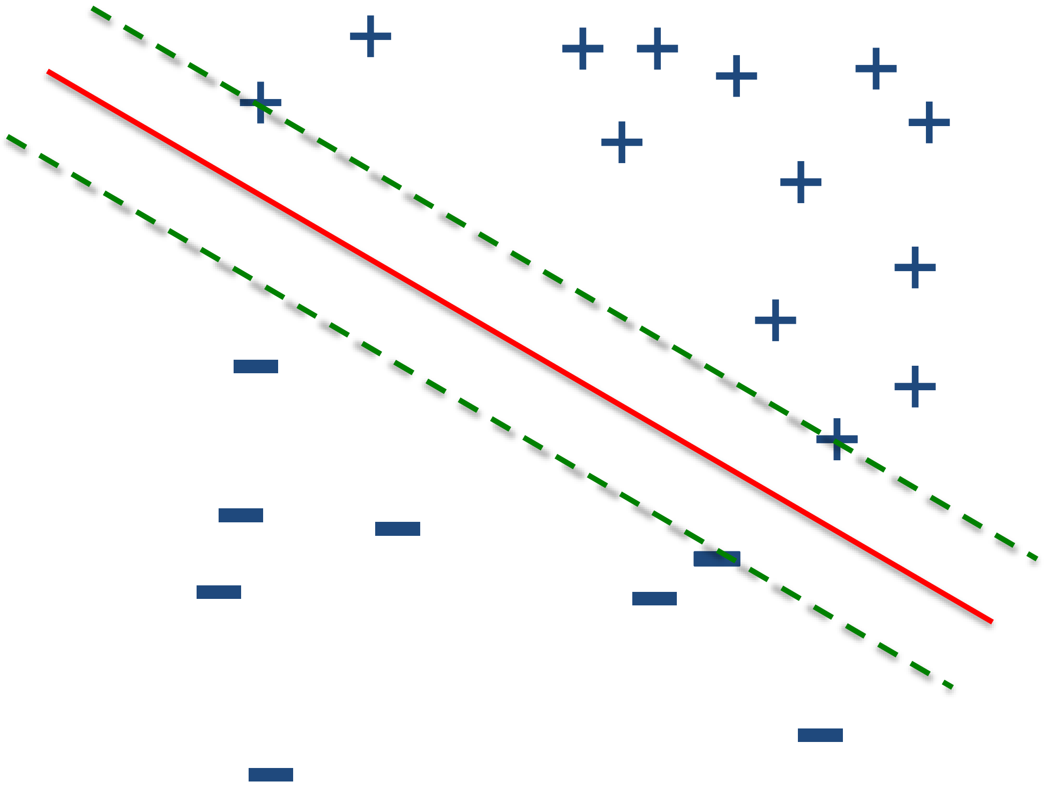
Quadratic Program



Slide credit: CMU MLD Matt Gormley

Support Vector Machines

Find linear separator with maximum margin



w, b

$\hat{y} = h(x)$

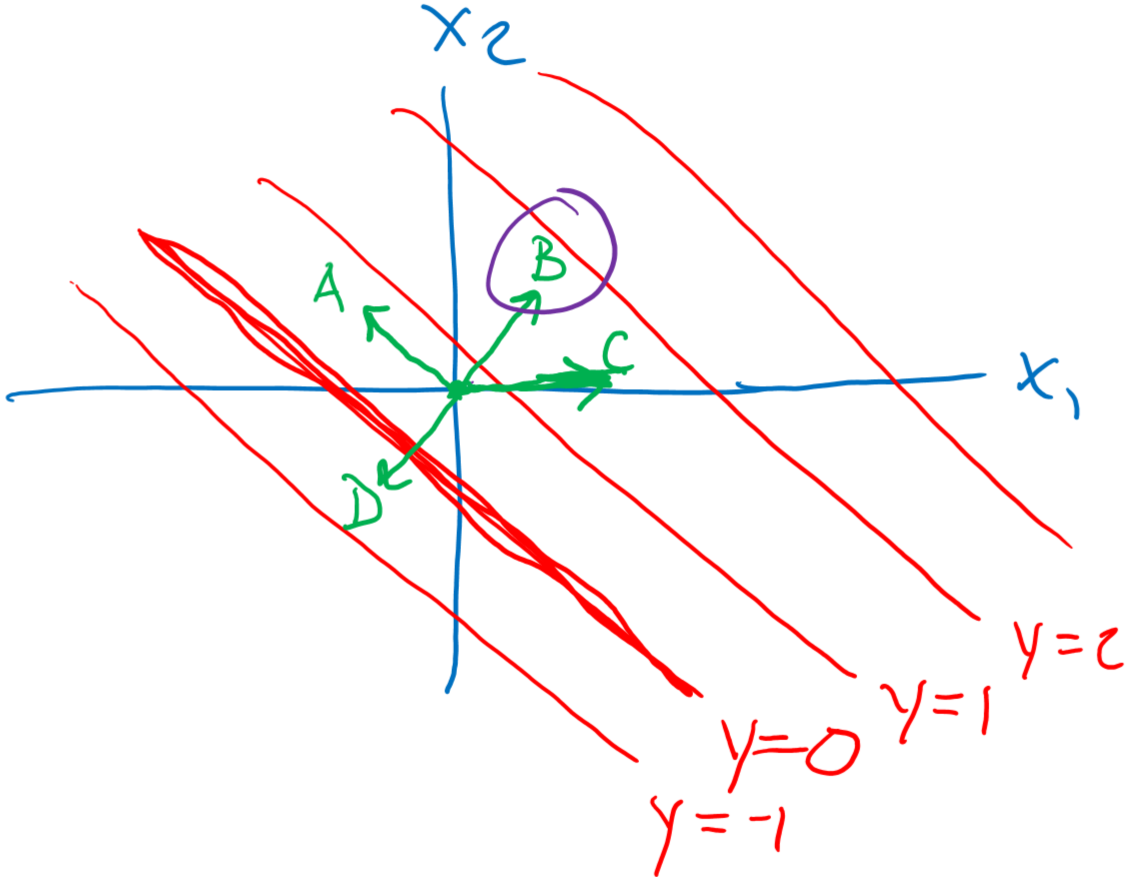
$\hat{y} = [\vec{w}^T \vec{x} + b]$

(Lecture 5) Poll

Which vector is the correct $\boldsymbol{w}$?

$y = \boldsymbol{w}^T \boldsymbol{x} + b$

$\vec{w} = \vec{B}$



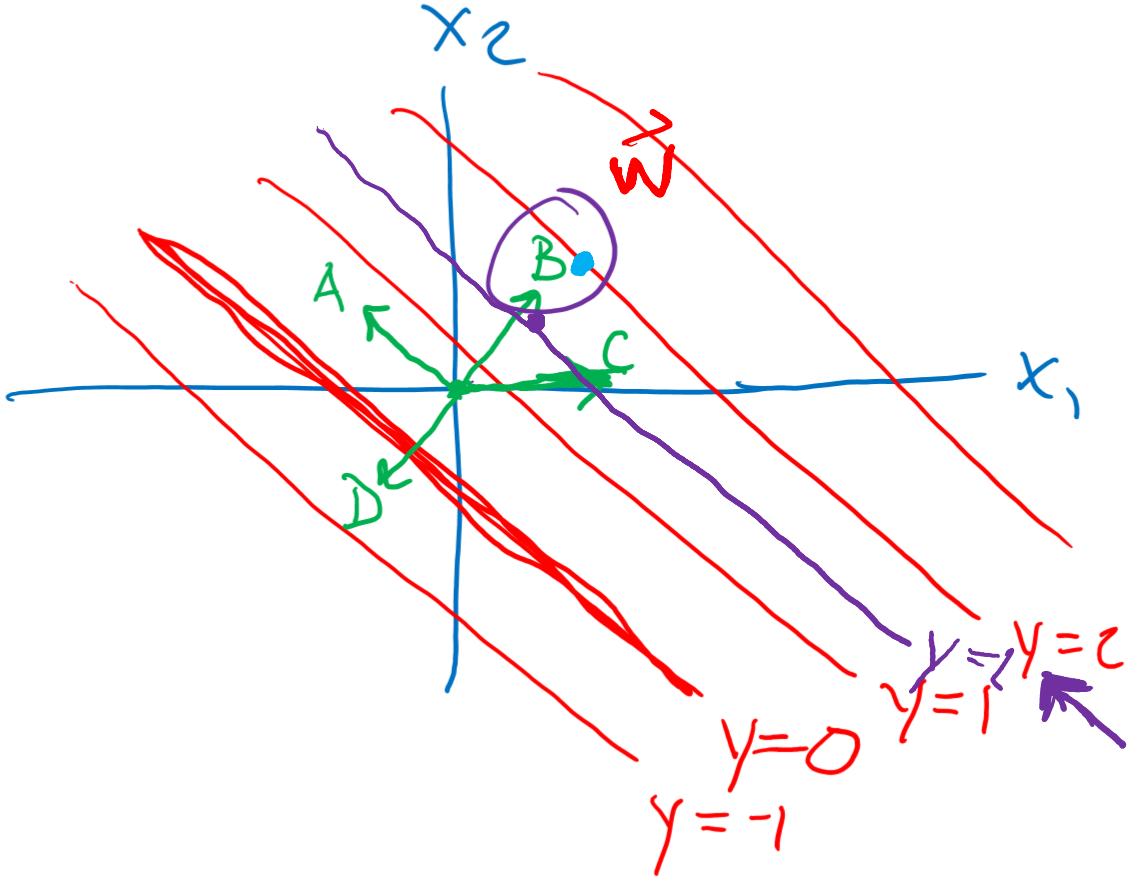
Piazza Poll 1

As the magnitude of w increases, will the distance between the contour lines of $y = \boldsymbol{w}^T \boldsymbol{x} + b$ increase or decrease?

A) Increase

B) Decrease

X C) Calamity



$x = [1, 1]^T$

$w_1 = [1, 1]$

$y = 2$

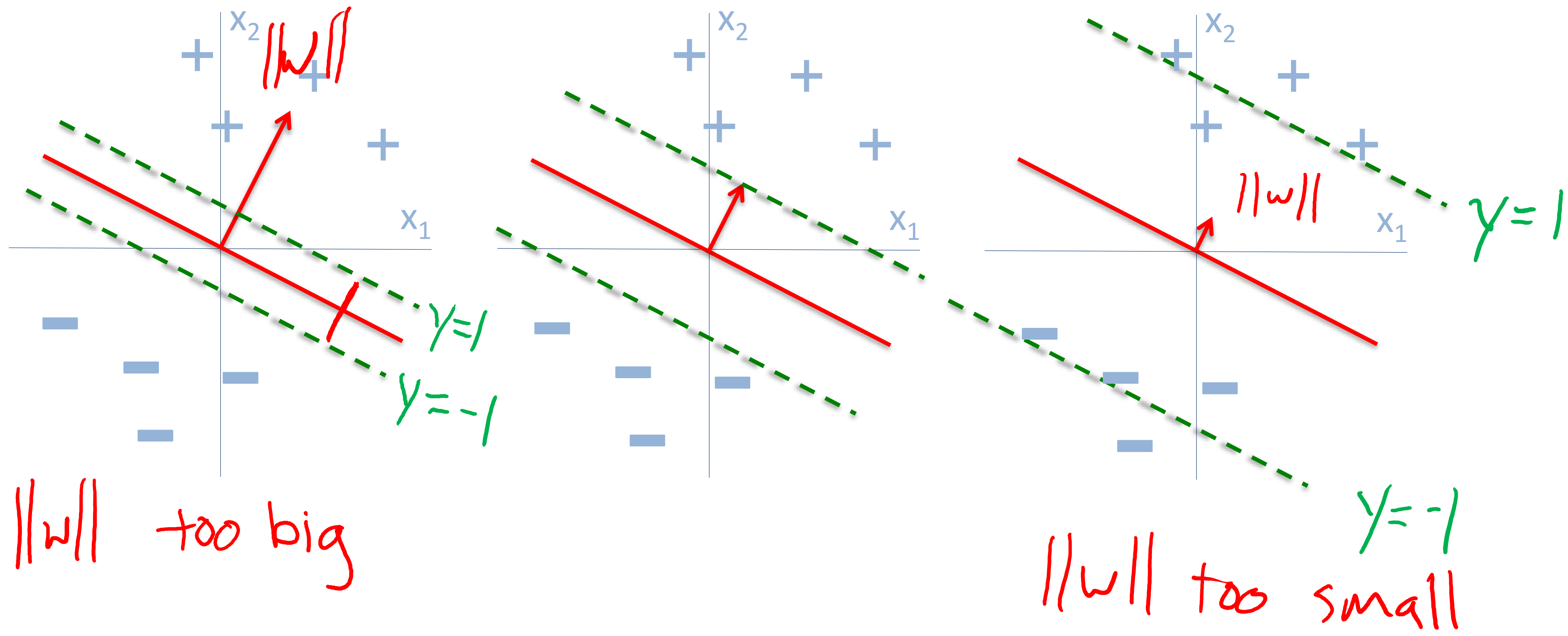
$w_2 = [2, 2]$

$x = [1/2, 1/2]$

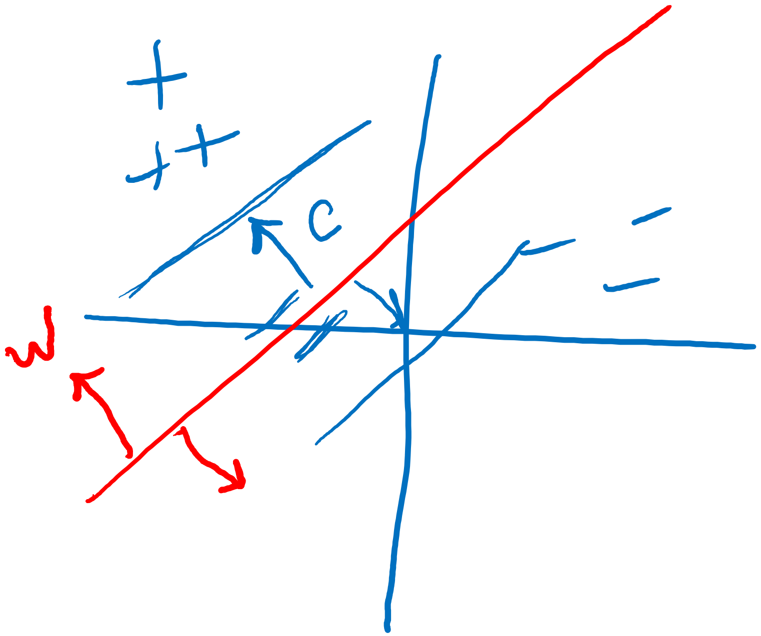
$y = 2$

Support Vector Machines

Find linear separator with maximum margin



Linear Separability



Data

$$\mathcal{D} = \{\boldsymbol{x}^{(i)}, y^{(i)}\}_{i=1}^{N} \quad \boldsymbol{x} \in \mathbb{R}^M, \quad y \in \{-1, +1\}$$

Linearly separable iff:

$$\exists\, \boldsymbol{w}, b \qquad s.t. \quad \boldsymbol{w}^T \boldsymbol{x}^{(i)} + b > 0 \quad \text{if } \; y^{(i)} = +1, \quad \text{and}$$

$$\boldsymbol{w}^T \boldsymbol{x}^{(i)} + b < 0 \quad \text{if } \; y^{(i)} = -1$$

$\Leftrightarrow \quad \exists w, b \quad s.t \; y^i(w^T x^i + b) > 0$

$\Leftrightarrow \quad \exists w, b, c \quad s.t \; y^i(w^T x^i + b) \geq c \quad \text{and} \quad c > 0$

$\Leftrightarrow \quad \exists w, b \quad s.t \; y^i(w^T x^i + b) \geq 1$

Linear Separability

Data

$$\mathcal{D} = \{\boldsymbol{x}^{(i)}, y^{(i)}\}_{i=1}^{N} \quad \boldsymbol{x} \in \mathbb{R}^M, \quad y \in \{-1, +1\}$$

Linearly separable iff:

$$\exists\, \boldsymbol{w}, b \quad s.t. \quad \boldsymbol{w}^T \boldsymbol{x}^{(i)} + b > 0 \;\text{ if }\; y^{(i)} = +1 \quad \text{and}$$

$$\boldsymbol{w}^T \boldsymbol{x}^{(i)} + b < 0 \;\text{ if }\; y^{(i)} = -1$$

$$\Leftrightarrow \exists\, \boldsymbol{w}, b \quad s.t. \quad y^{(i)}\left(\boldsymbol{w}^T \boldsymbol{x}^{(i)} + b\right) > 0$$

$$\Leftrightarrow \exists\, \boldsymbol{w}, b, c \quad s.t. \quad y^{(i)}\left(\boldsymbol{w}^T \boldsymbol{x}^{(i)} + b\right) \geq c \quad \text{and} \quad c > 0$$

$$\Leftrightarrow \exists\, \boldsymbol{w}, b \quad s.t. \quad y^{(i)}\left(\boldsymbol{w}^T \boldsymbol{x}^{(i)} + b\right) \geq 1$$

Piazza Poll 2

$\min_{w,b}$ width

s.t. ______

Are these to statements equivalent?

$\exists\, \boldsymbol{w}, b, c \quad s.t. \quad y^{(i)}\left(\boldsymbol{w}^T \boldsymbol{x}^{(i)} + b\right) \geq c \quad \text{and} \quad c > 0$

$\exists\, \boldsymbol{w}, b \quad s.t. \quad y^{(i)}\left(\tilde{\boldsymbol{w}}^T \boldsymbol{x}^{(i)} + \tilde{b}\right) \geq 1$

A) Yes

B) No

C) Calamity

y^i

$\tilde{w} = \frac{w}{c}$

$\tilde{b} = \frac{w}{c}$

Linear Separability

Data

$$\mathcal{D} = \{\boldsymbol{x}^{(i)}, y^{(i)}\}_{i=1}^{N} \quad \boldsymbol{x} \in \mathbb{R}^M, \quad y \in \{-1, +1\}$$

Linearly separable iff:

$$\begin{aligned}
& \exists\, \boldsymbol{w}, b \quad s.t. \quad \boldsymbol{w}^T \boldsymbol{x}^{(i)} + b > 0 \quad \text{if} \quad y^{(i)} = +1 \quad \text{and} \\
& \qquad\qquad\qquad\quad \boldsymbol{w}^T \boldsymbol{x}^{(i)} + b < 0 \quad \text{if} \quad y^{(i)} = -1 \\
\Leftrightarrow\ & \exists\, \boldsymbol{w}, b \quad s.t. \quad y^{(i)}\left(\boldsymbol{w}^T \boldsymbol{x}^{(i)} + b\right) > 0 \\
\Leftrightarrow\ & \exists\, \boldsymbol{w}, b, c \quad s.t. \quad y^{(i)}\left(\boldsymbol{w}^T \boldsymbol{x}^{(i)} + b\right) \geq c \quad \text{and} \quad c > 0 \\
\Leftrightarrow\ & \exists\, \boldsymbol{w}, b \quad s.t. \quad y^{(i)}\left(\boldsymbol{w}^T \boldsymbol{x}^{(i)} + b\right) \geq 1
\end{aligned}$$

Support Vector Machines

Find linear separator with maximum margin

Let $\boldsymbol{x}_+$ and $\boldsymbol{x}_-$ be hypothetical points on the +/- margin from the decision boundary

$$\exists\, \boldsymbol{w}, b \quad s.t. \quad y^{(i)}\left(\boldsymbol{w}^T\boldsymbol{x}^{(i)} + b\right) \geq 1$$

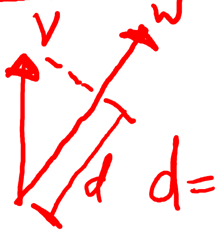
$$\Leftrightarrow \exists\, \boldsymbol{w}, b \quad s.t. \quad \boldsymbol{w}^T\boldsymbol{x}_+ + b \geq +1 \quad \text{and}$$

$$\boldsymbol{w}^T\boldsymbol{x}_- + b \leq -1$$

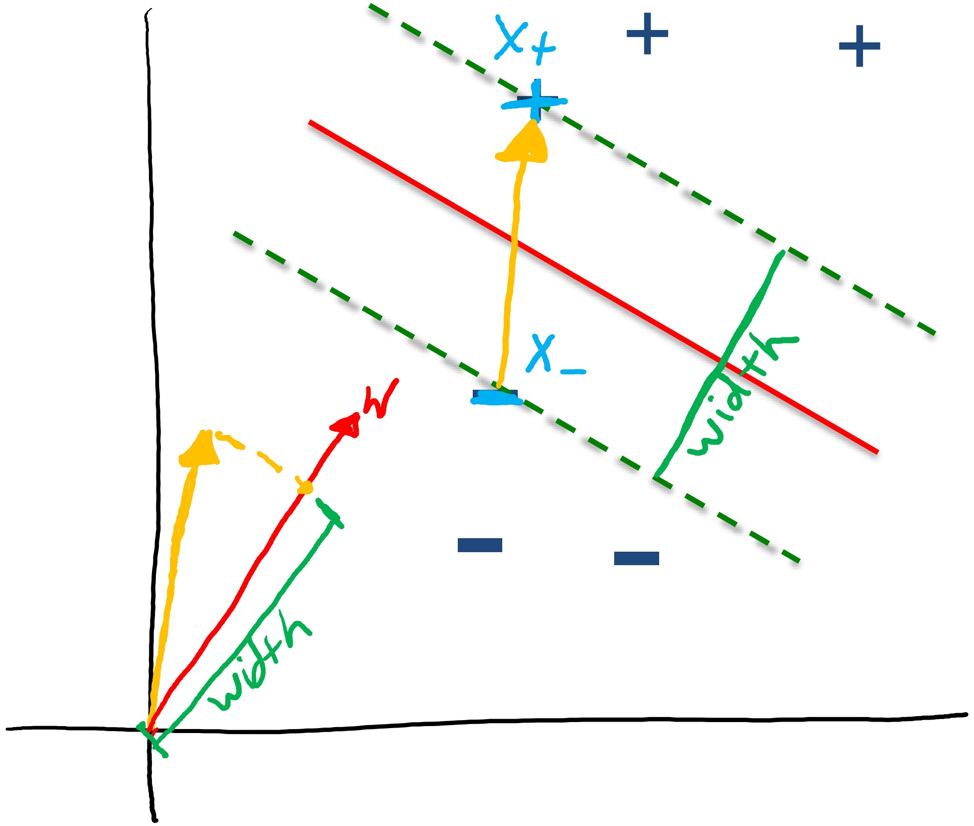
Consider the vector from $\boldsymbol{x}_-$ to $\boldsymbol{x}_+$ and its projection onto the vector $\boldsymbol{w}$:

$$\text{width} = \frac{w^T}{\|w\|_2}(x_+ - x_-)$$

Projection



$$d = \frac{w^T v}{\|w\|_2}$$



Support Vector Machines

$$width = \frac{\boldsymbol{w}^T}{\|\boldsymbol{w}\|_2}(\boldsymbol{x}_+ - \boldsymbol{x}_-)$$

Find linear separator with maximum margin

$$\max_{\boldsymbol{w},\boldsymbol{b}} \quad \underbrace{\text{"width"}}$$

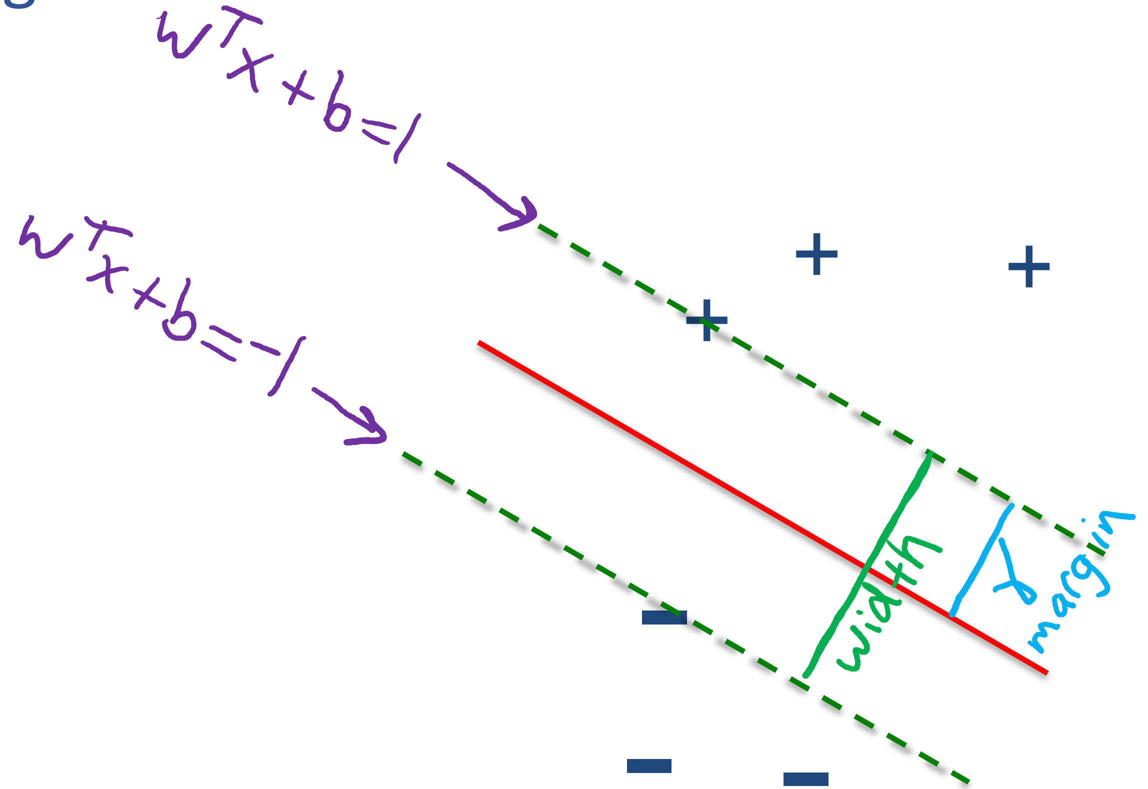
$$\text{s.t.} \quad y^{(i)}\left(\boldsymbol{w}^T\boldsymbol{x}^{(i)} + b\right) \geq 1 \quad \forall\, i$$

$$width = \frac{w^T}{\|w\|_2}(x_+ - x_-)$$

$$= \frac{1}{\|w\|_2}(w^T x_+ - w^T x_-)$$

$$= \frac{1}{\|w\|_2}\left((1-b) - (-1-b)\right)$$

$$= \frac{2}{\|w\|_2}$$



$$w^T x_+ + b = 1 \qquad w^T x_- + b = -1$$

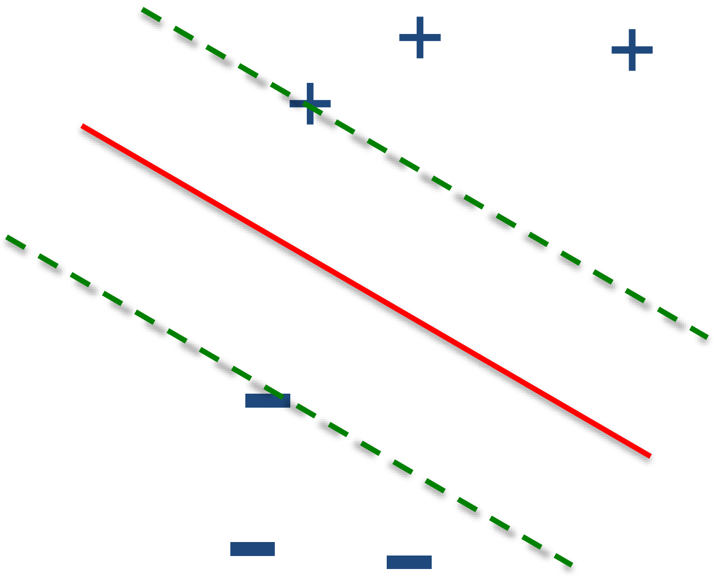
$$w^T x_+ = 1 - b \qquad w^T x_- = -1 - b$$

Support Vector Machines

$$\text{width} = \frac{2}{\|w\|_2}$$

Find linear separator with maximum margin

$$\underset{\boldsymbol{w},\boldsymbol{b}}{\operatorname{argmax}} \quad \text{width}$$



Support Vector Machines

$$\text{width} = \frac{2}{\|\boldsymbol{w}\|_2}$$

Find linear separator with maximum margin

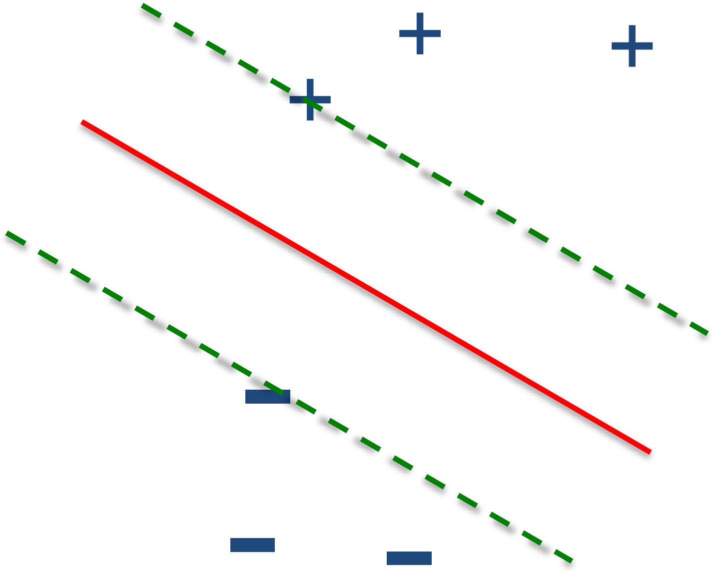
$$\underset{\boldsymbol{w},\boldsymbol{b}}{\operatorname{argmax}} \quad \text{width}$$

$$\Leftrightarrow \underset{\boldsymbol{w},\boldsymbol{b}}{\operatorname{argmax}} \quad \frac{2}{\|\boldsymbol{w}\|_2}$$

$$\Leftrightarrow \underset{\boldsymbol{w},\boldsymbol{b}}{\operatorname{argmin}} \quad \frac{1}{2}\|\boldsymbol{w}\|_2$$

$$\Leftrightarrow \underset{\boldsymbol{w},\boldsymbol{b}}{\operatorname{argmin}} \quad \frac{1}{2}\|\boldsymbol{w}\|_2^2$$

$$\Leftrightarrow \underset{\boldsymbol{w},\boldsymbol{b}}{\operatorname{argmin}} \quad \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}$$



SVM Optimization

Quadratic program!

$$\min_{\boldsymbol{w},b} \quad \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}$$

$$\text{s.t.} \quad y^{(i)}\left(\boldsymbol{w}^T\boldsymbol{x}^{(i)} + b\right) \geq 1 \quad \forall\, i$$

Quadratic Program

$$\min_{\boldsymbol{x}} \quad \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x}$$

$$\text{s.t.} \quad A\boldsymbol{x} \preceq \boldsymbol{b}$$