Announcements

Assignments

- HW8: due Thu, 12/3, 11:59 pm
- HW9
  - Out Friday
  - Due Wed, 12/9, 11:59 pm
  - The two slip days are free (last possible submission Fri, 12/11, 11:59 pm)

Final Exam

- Mon, 12/14
- Stay tuned to Piazza for more details

Wrap-up MDP/RL

RL slides

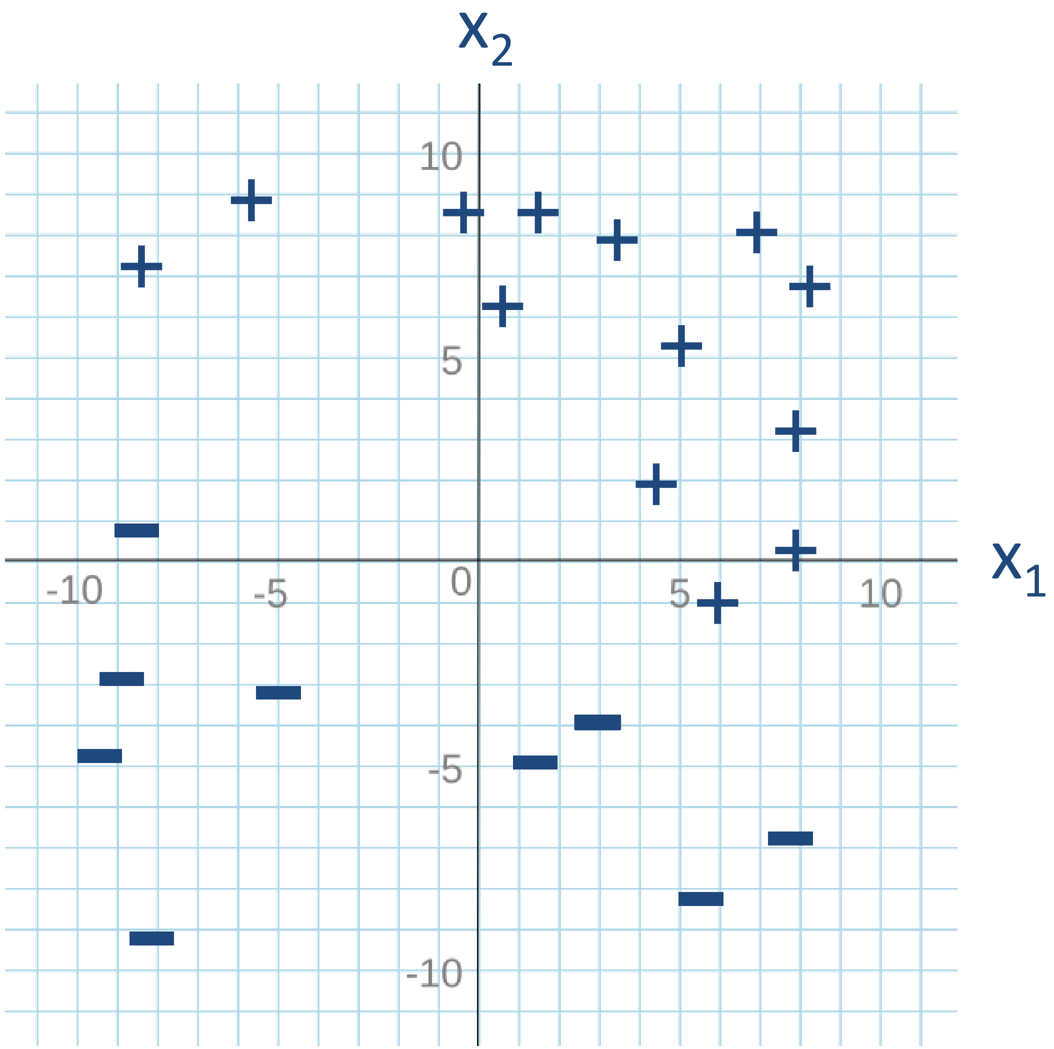
Introduction to Machine Learning

Support Vector Machines

Instructor: Pat Virtue

Support Vector Machines

Linear Classification



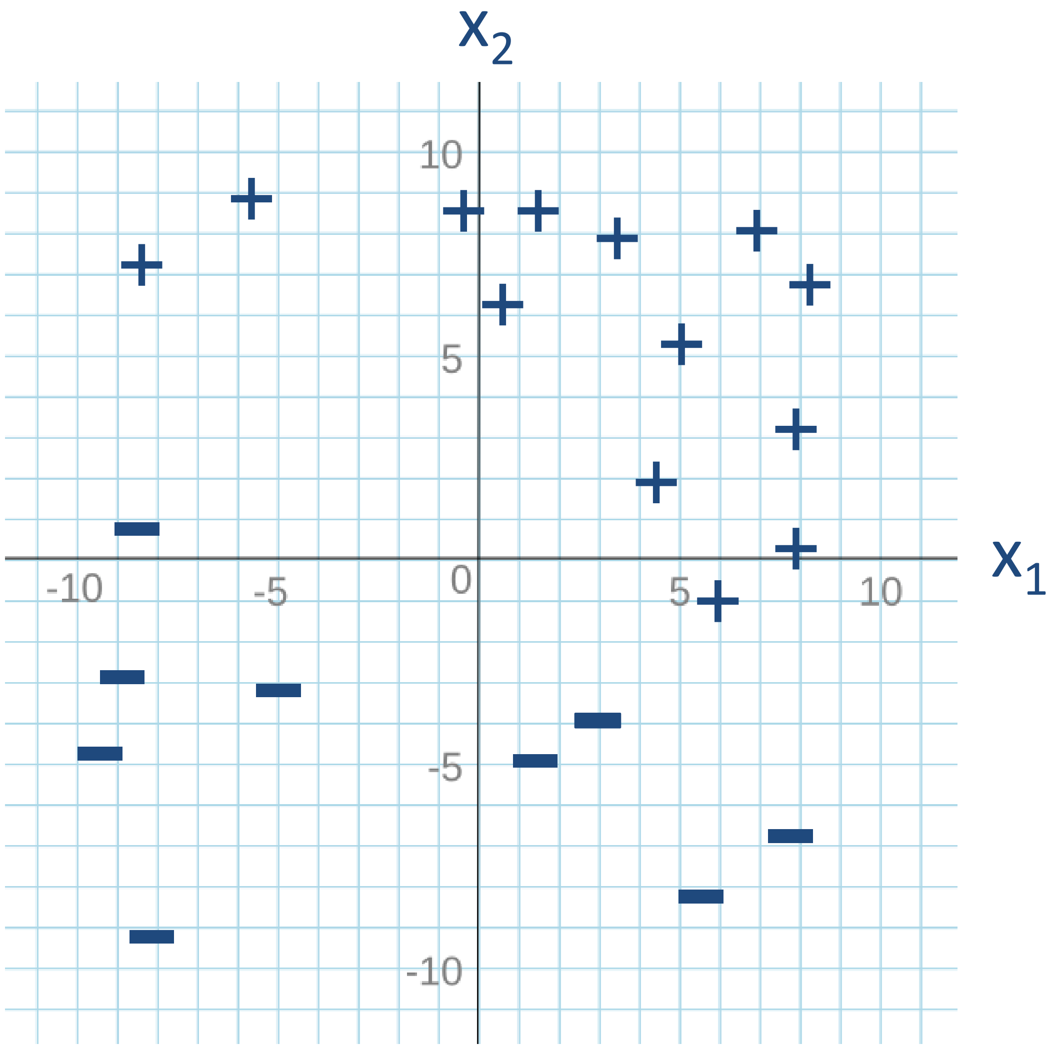
Support Vector Machines

Margin

Given a linearly separable dataset and a linear separator defined by the hyperplane $\boldsymbol{w}^T\boldsymbol{x} + b = 0$,

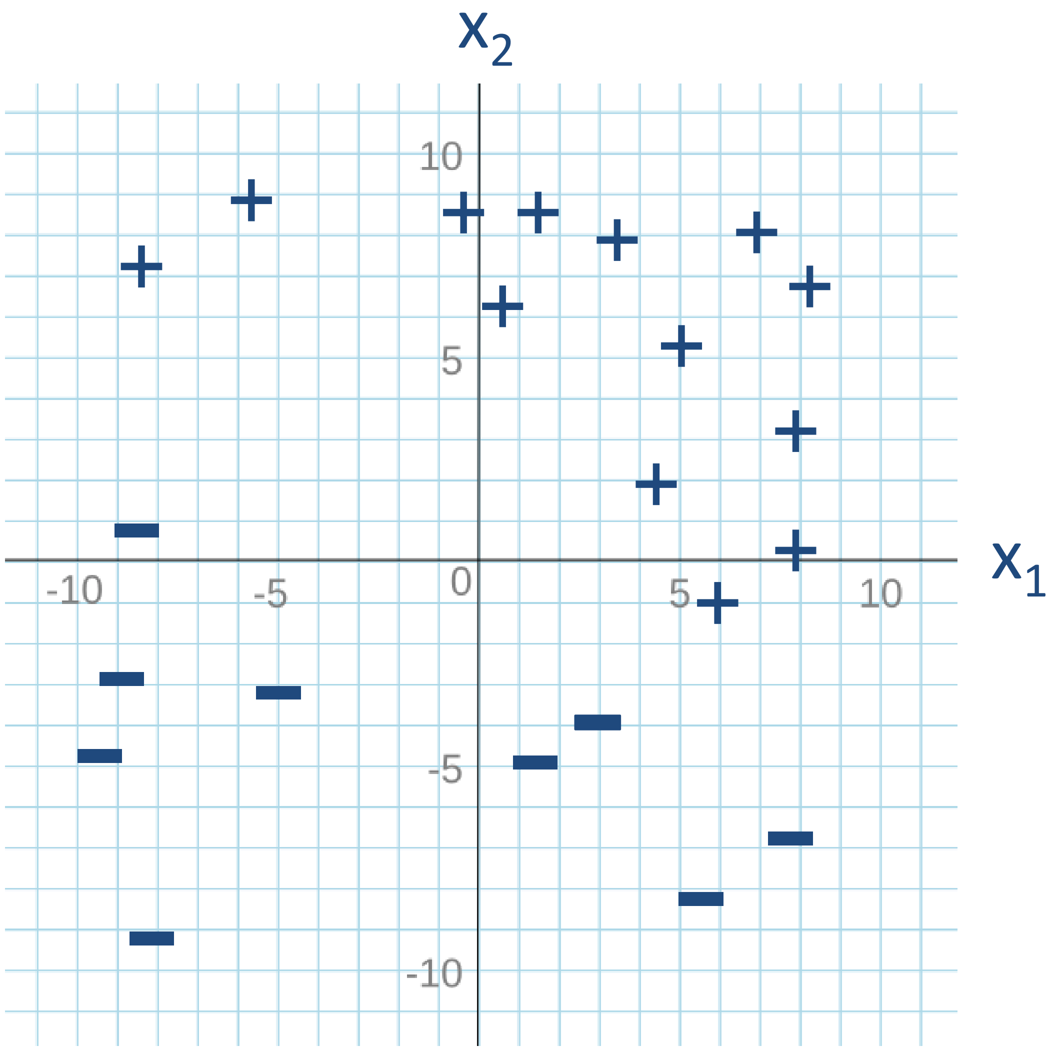
the margin, γ, is the distance from this hyperplane to the closest point, $\boldsymbol{x}^{(i)}$, in a dataset.

The closest point may be on either side of the hyperplane.



Support Vector Machines

Max Margin



Support Vector Machines

SVM were super popular right before the current deep learning craze

Support Vector Machines

Important concepts withing SVMs

- Max-margin classification
- Optimization
  - Constrained optimization
  - Quadratic program
  - Primal → dual
  - Lagrange Multipliers
- Support non-linear classification
  - Feature maps
  - Kernel trick

Piazza Poll 1

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

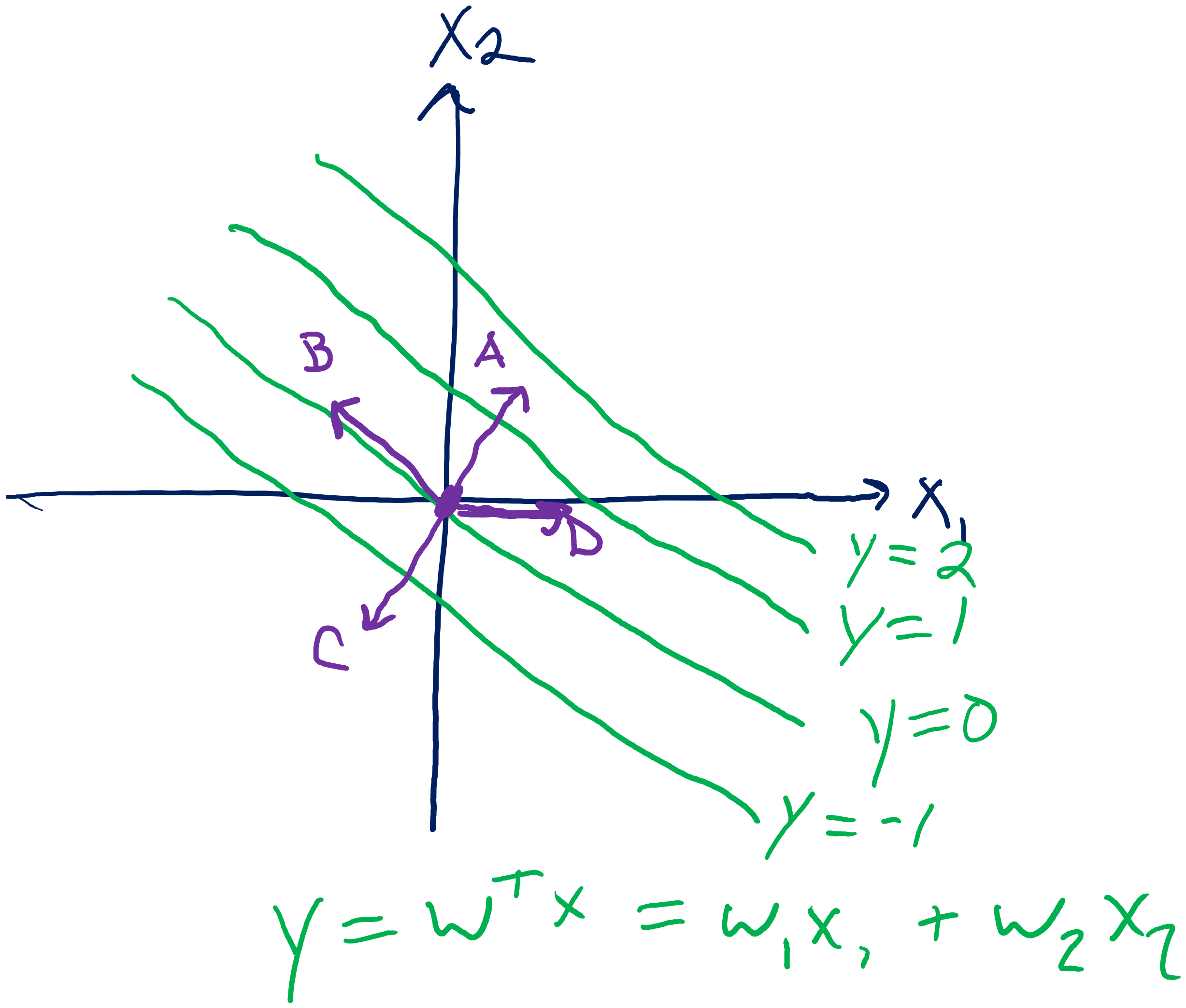
A.

B.

C.

D.

E. I don't know



Constrained Optimization

Linear Program

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

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

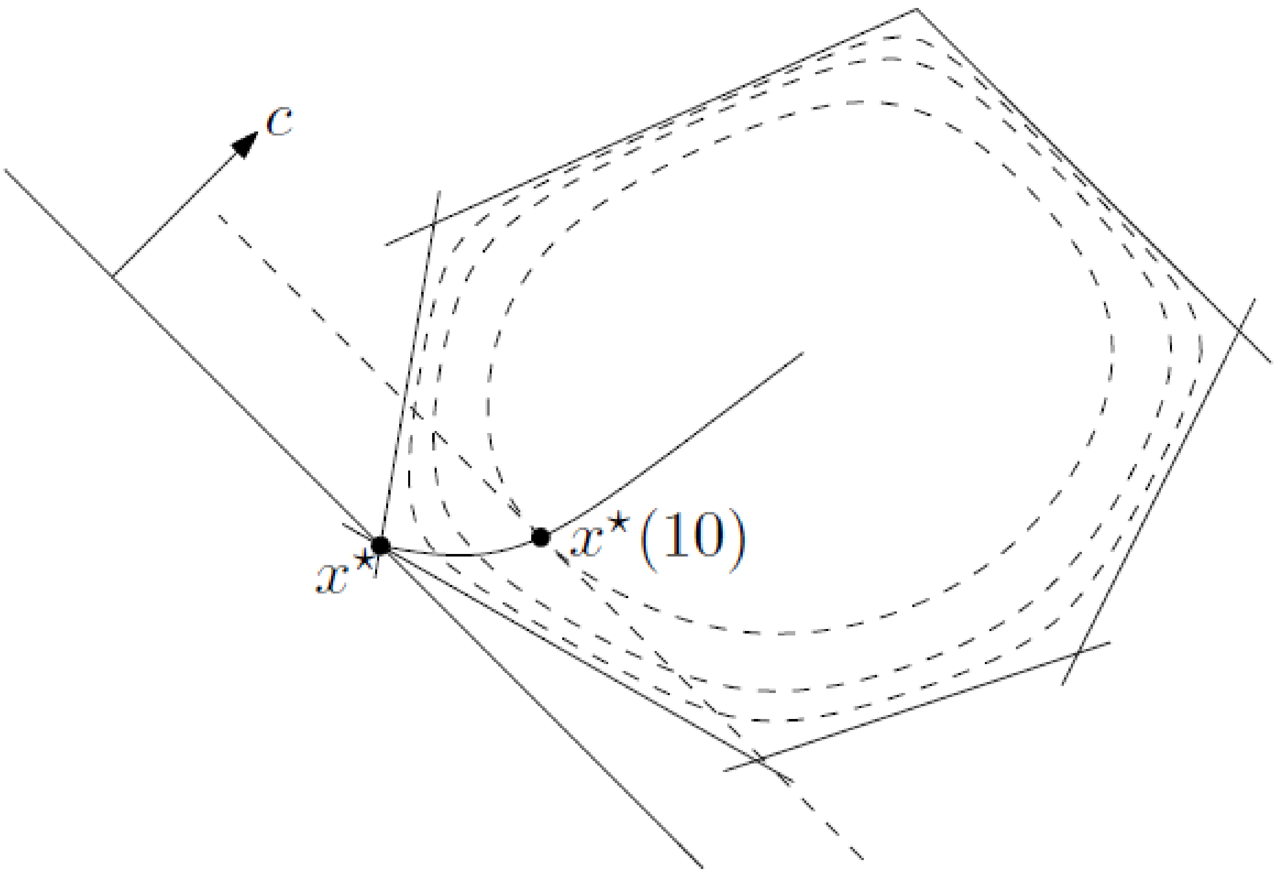


Figure: Fig 11.2 from Boyd and Vandenberghe, *Convex Optimization*

Constrained Optimization

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

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

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:
  $\boldsymbol{v}^T \boldsymbol{Q} \boldsymbol{v} > 0 \quad \forall\, \boldsymbol{v} \in \mathbb{R}^M \backslash \boldsymbol{0}$
- A symmetric $\boldsymbol{Q}$ is positive-definite if all of its eigenvalues are positive

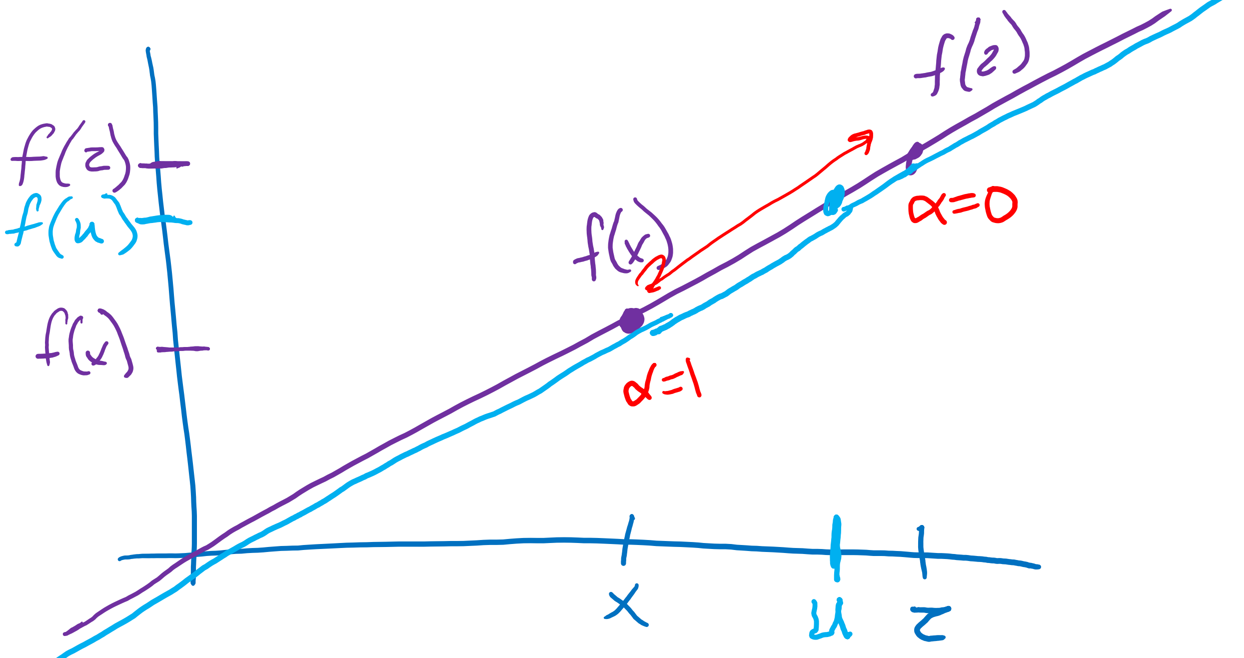
Optimization (from Lecture 7)

Linear function

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

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

$\alpha = 0.25$



Optimization (from Lecture 7)

Convex function

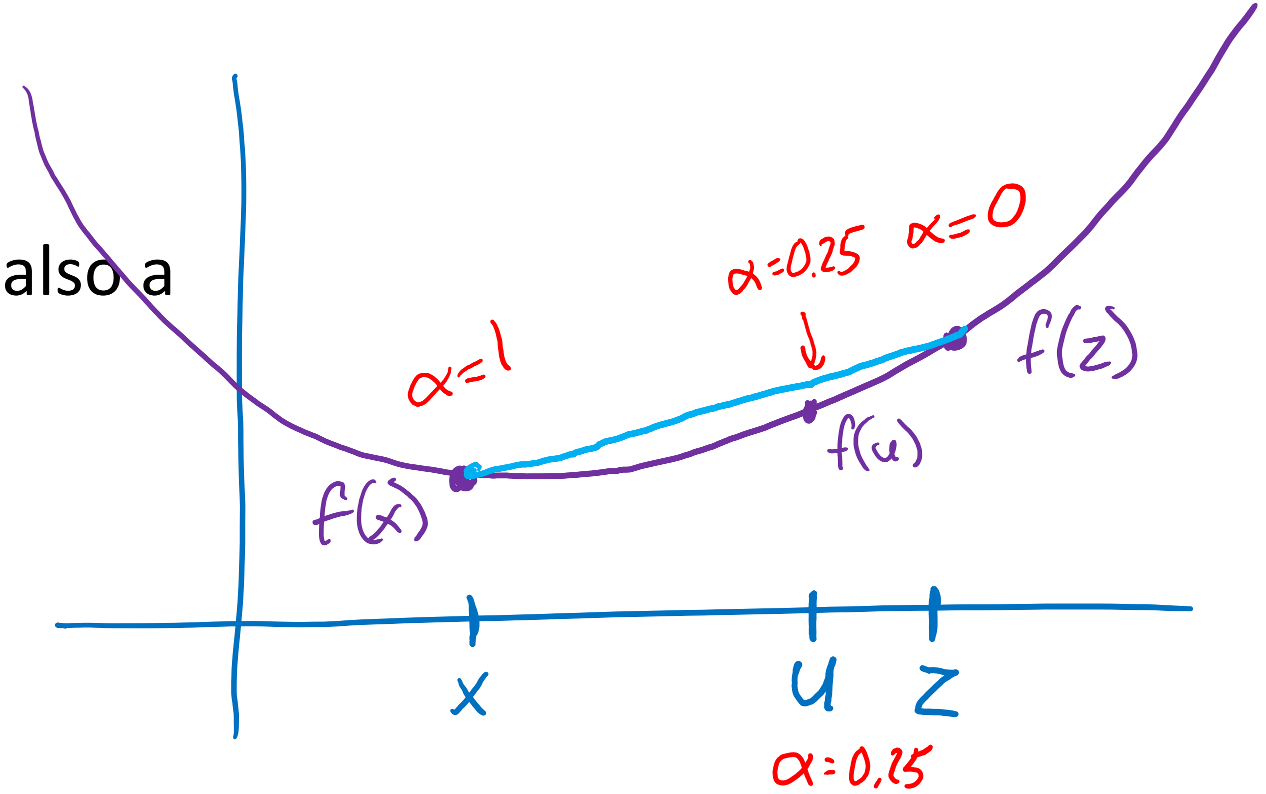
If $f(\boldsymbol{x})$ is convex, then:

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

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

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

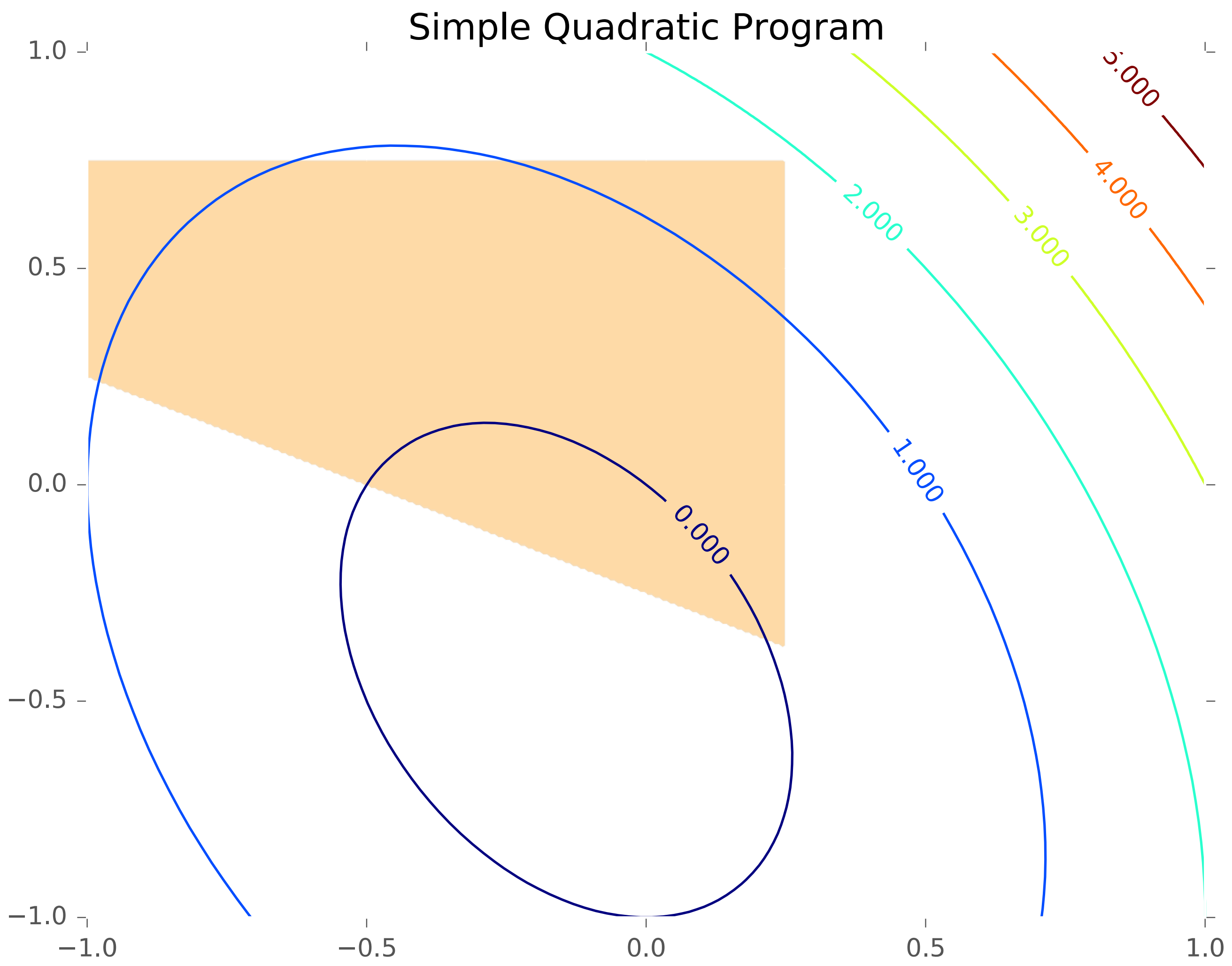
Solvers

- Simplex
- Interior point methods

Special Case

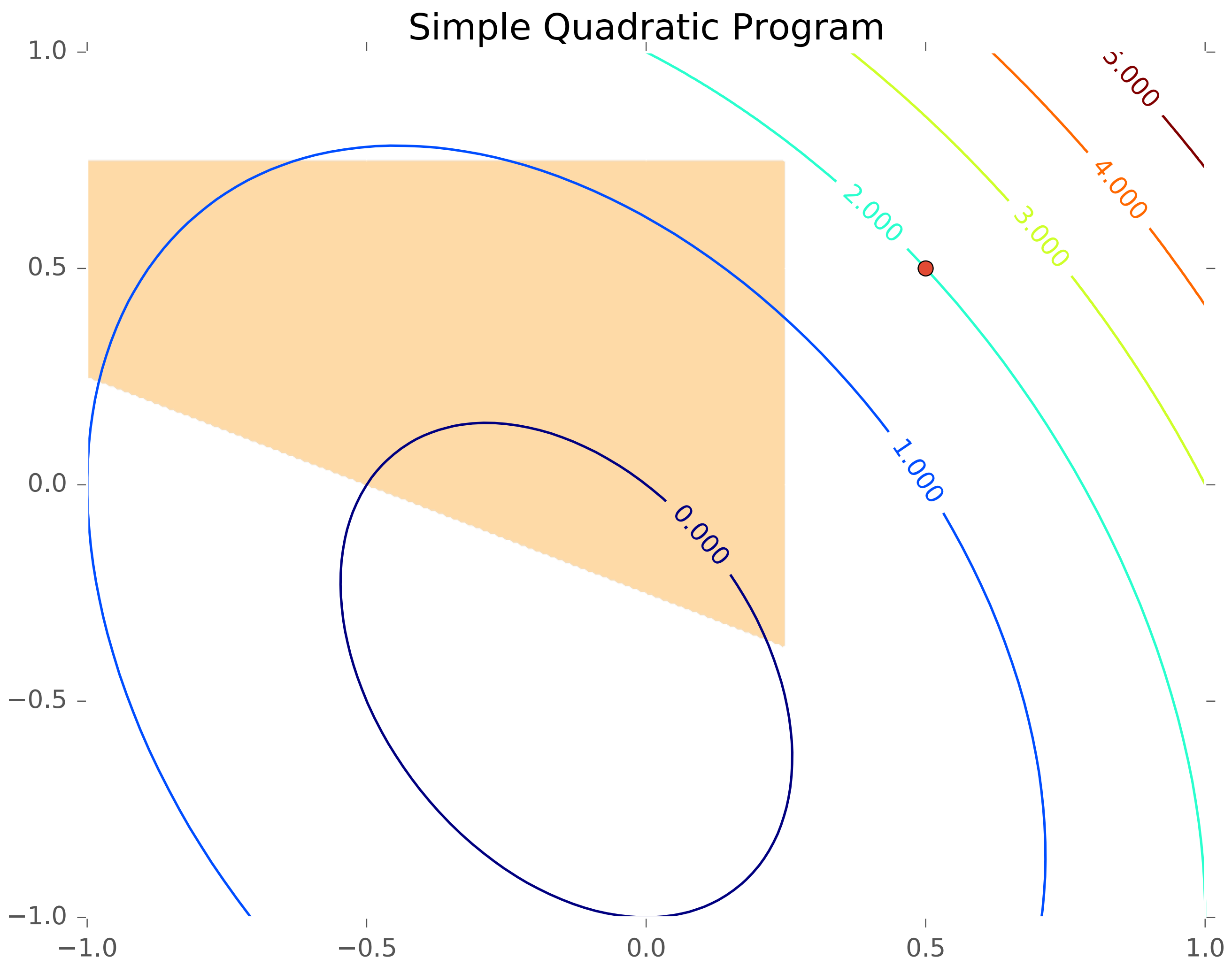
- If $\boldsymbol{Q}$ is positive-definite, the problem is convex
- $\boldsymbol{Q}$ is positive-definite if:
  $$\boldsymbol{v}^T \boldsymbol{Q} \boldsymbol{v} > 0 \quad \forall\, \boldsymbol{v} \in \mathbb{R}^M \backslash \boldsymbol{0}$$
- A symmetric $\boldsymbol{Q}$ is positive-definite if all of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

Quadratic Program



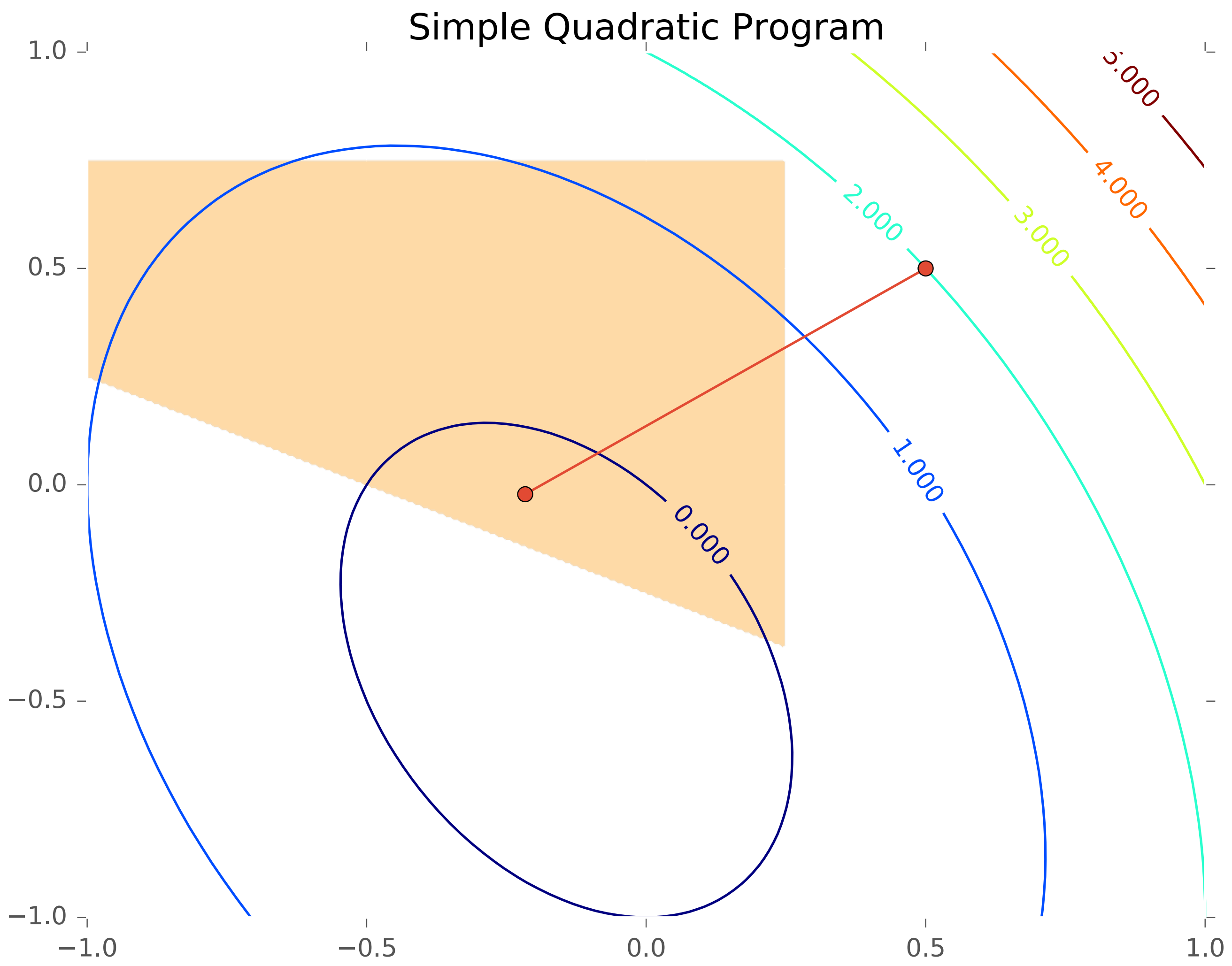
Slide credit: CMU MLD Matt Gormley

Quadratic Program



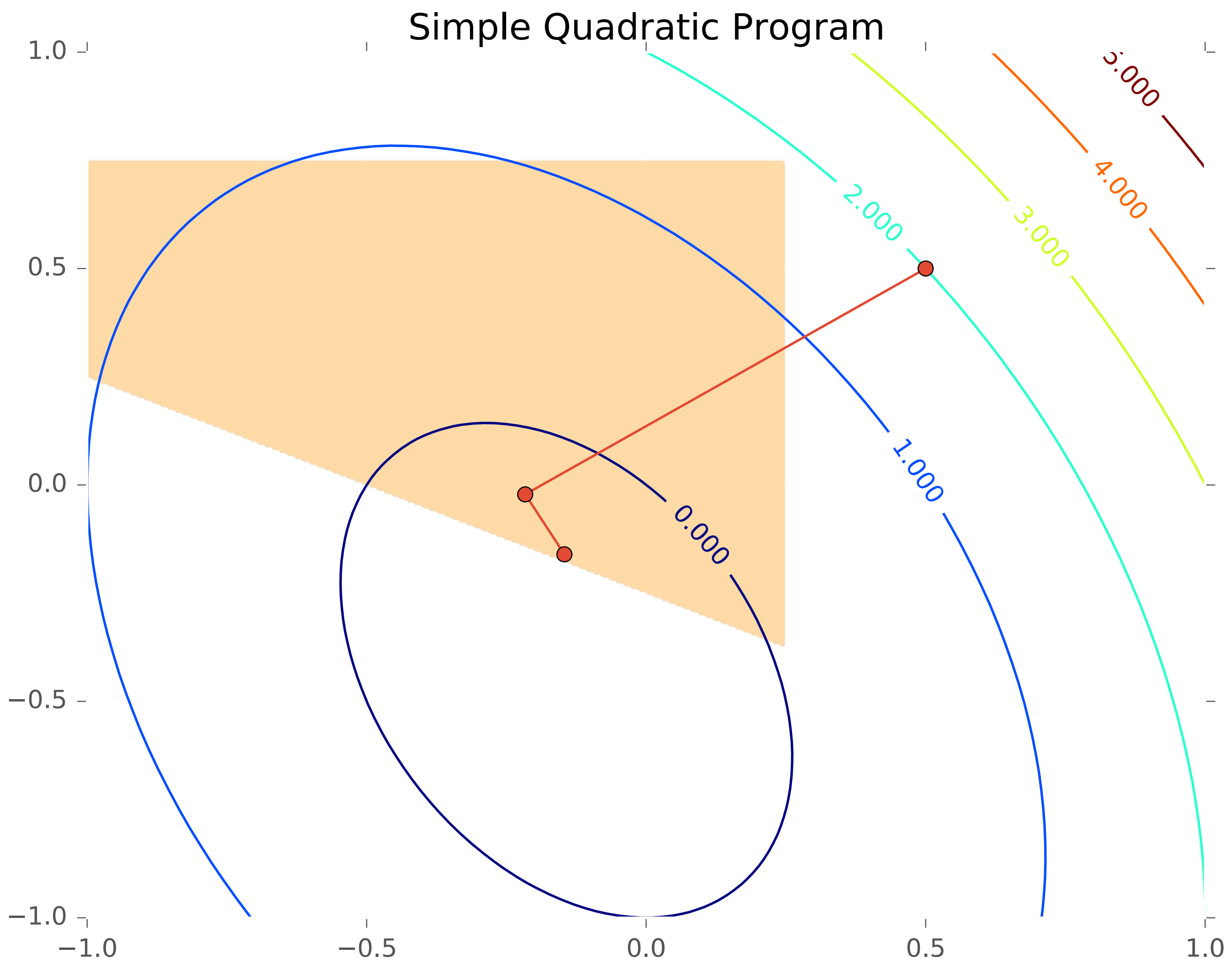
Slide credit: CMU MLD Matt Gormley

Quadratic Program



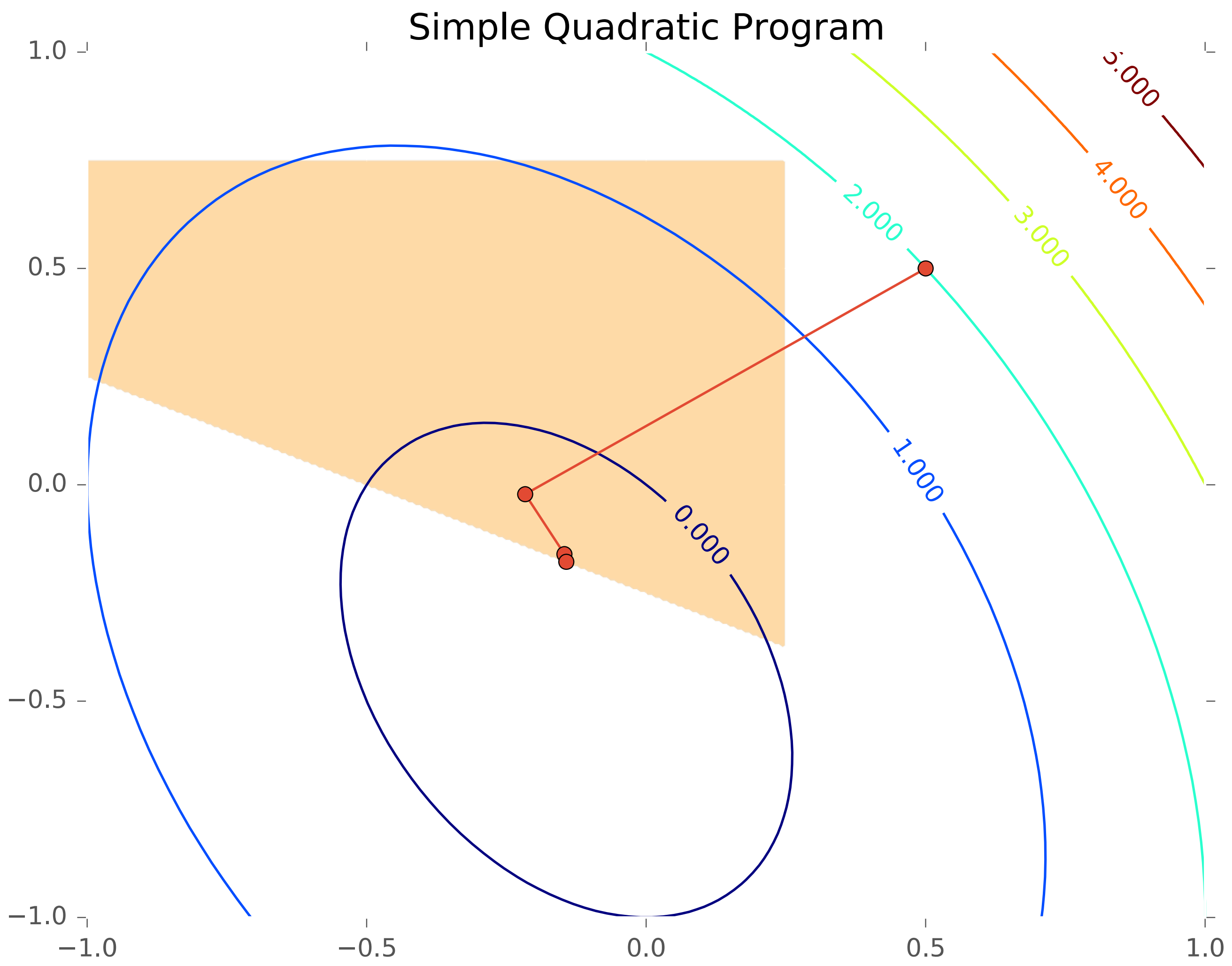
Slide credit: CMU MLD Matt Gormley

Quadratic Program



Slide credit: CMU MLD Matt Gormley

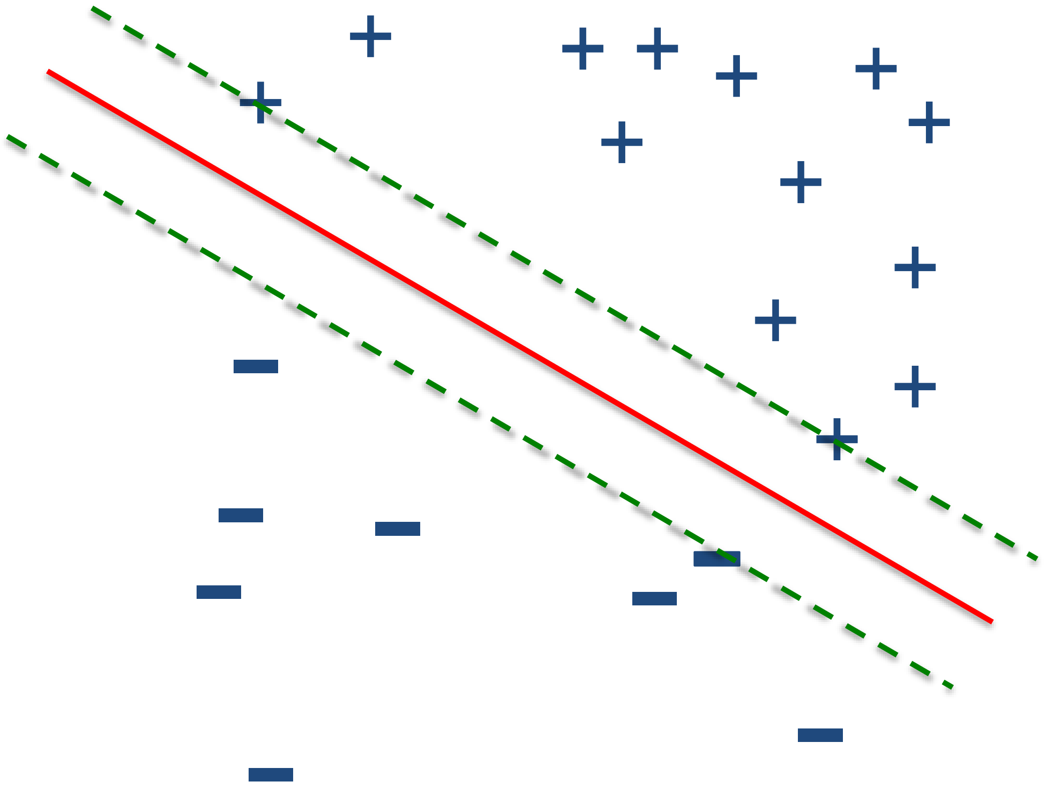
Quadratic Program



Slide credit: CMU MLD Matt Gormley

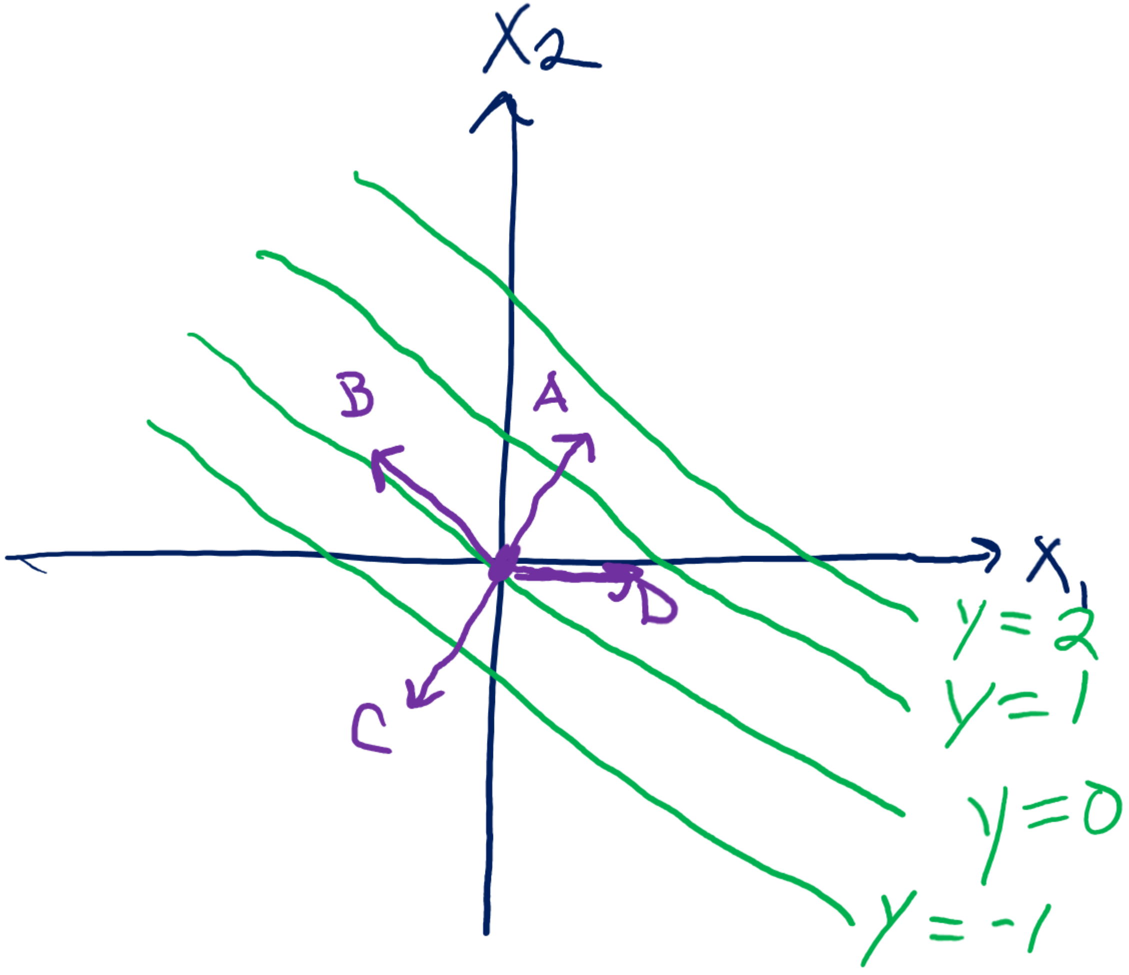
Support Vector Machines

Find linear separator with maximum margin



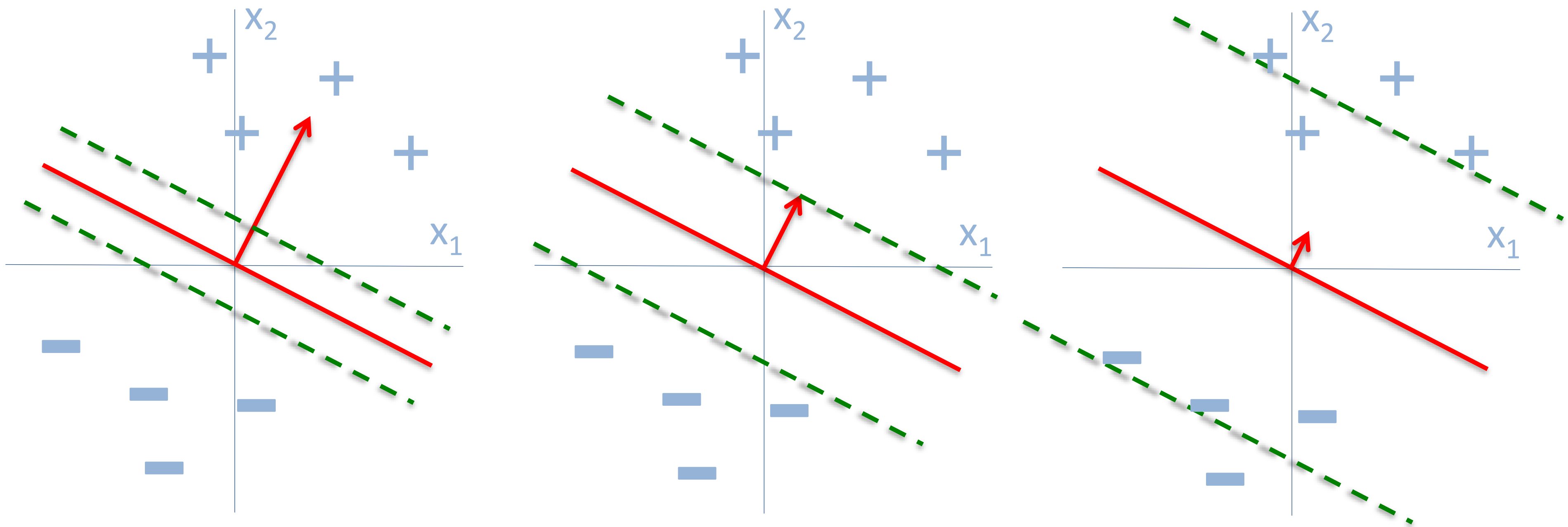
Piazza Poll 2

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



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:

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

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

Slide credit: CMU MLD Matt Gormley

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:

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

Slide credit: CMU MLD Matt Gormley