Support Vector Machine (SVM)-I

Guowei Wei
Department of Mathematics
Michigan State University

References:
Duc D. Nguyen’s lecture notes
Wikipedia

Introduction

- One of top ten methods in machine learning
- Classification
- Regression, i.e., support vector regression (SVR)
- Supervised learning in general
- For unsupervised learning:

Support vector clustering (SVC) by Hava Siegelmann and Vladimir Vapnik

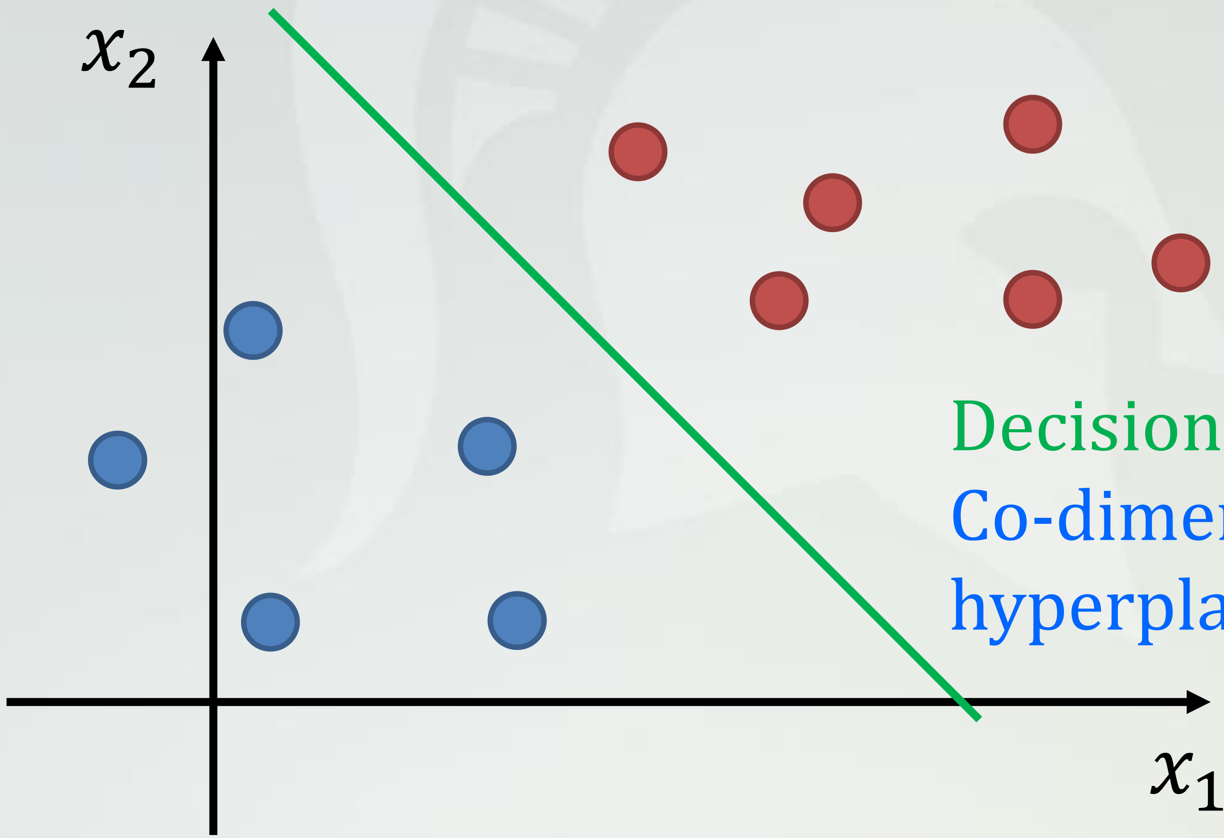
SVM for linear Classifiers

Decision Boundary

Training set: $\mathcal{D} = \left\{\left(\mathbf{x}^{(i)}, y^{(i)}\right) \middle| \mathbf{x}^{(i)} \in \mathbb{R}^n, y^{(i)} \in \{-1,1\}\right\}_{i=1}^{M}$

Red dots: $y = 1$; Blue dots: $y = -1$

(in SVM, we label negative class as -1 instead of 0)



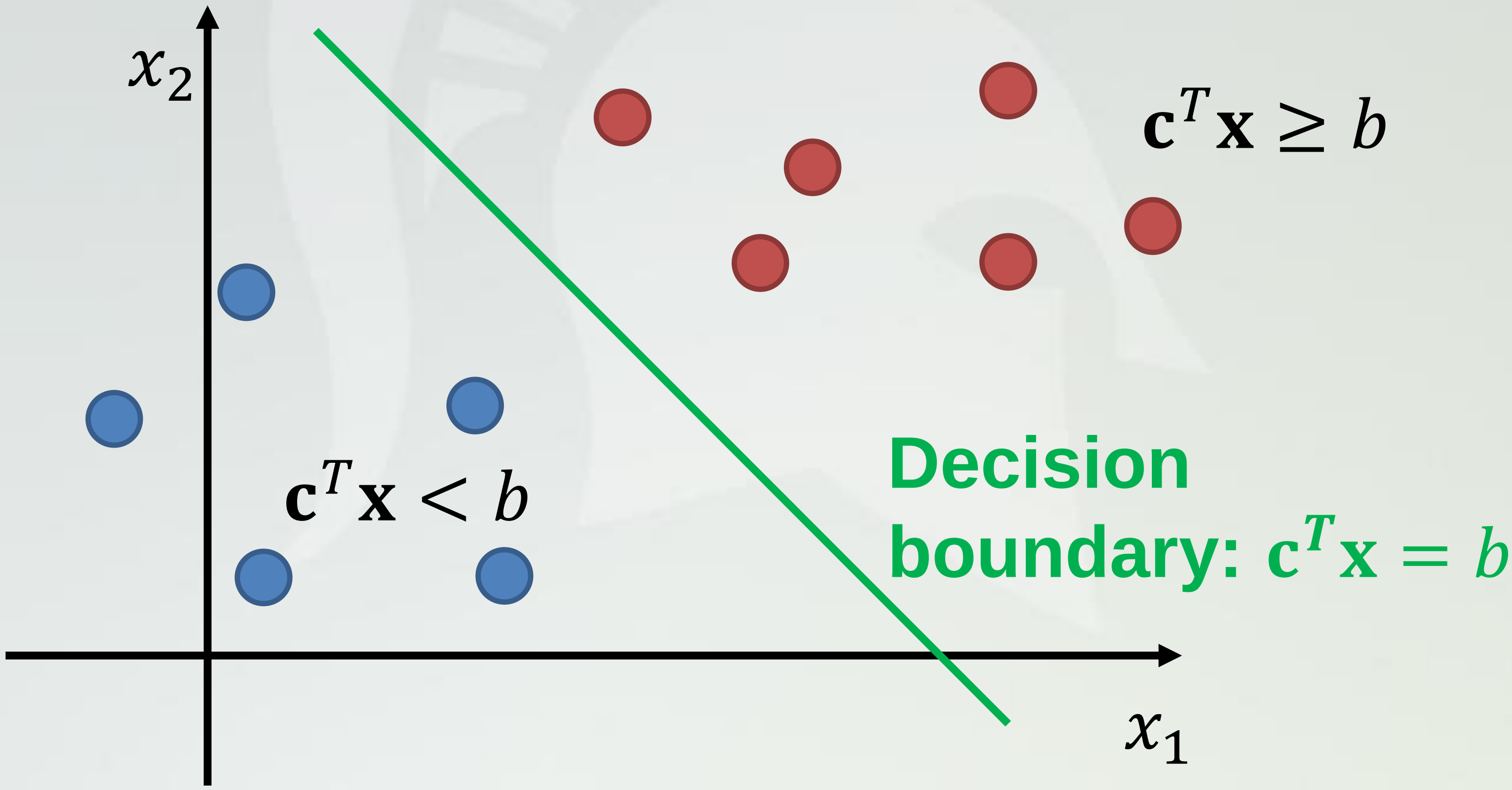
Decision boundary: Co-dimension 1 hyperplane

Decision Boundary

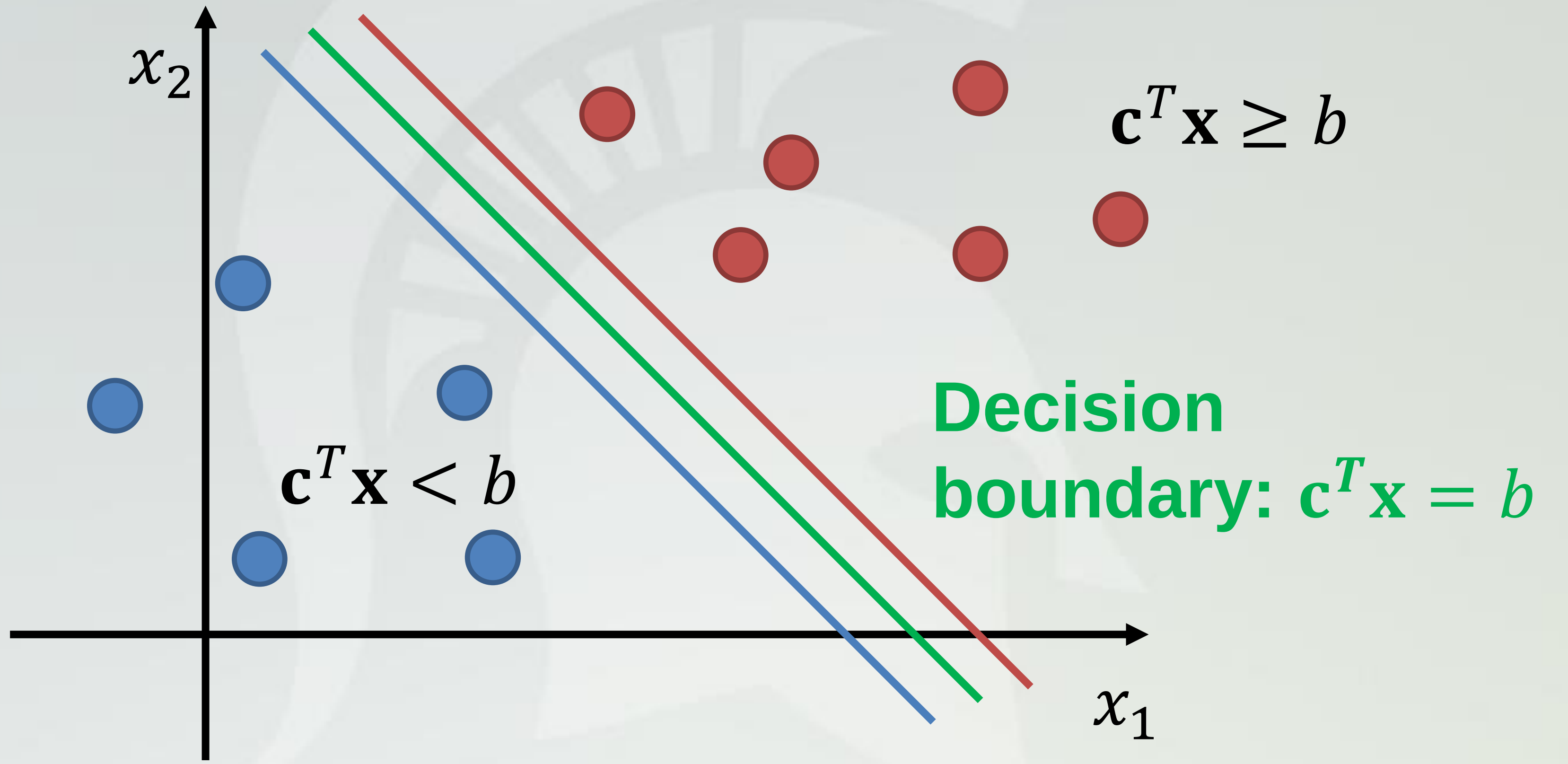
Recall:

$\mathbf{x} = (1, x_1, \ldots, x_n)^T$,

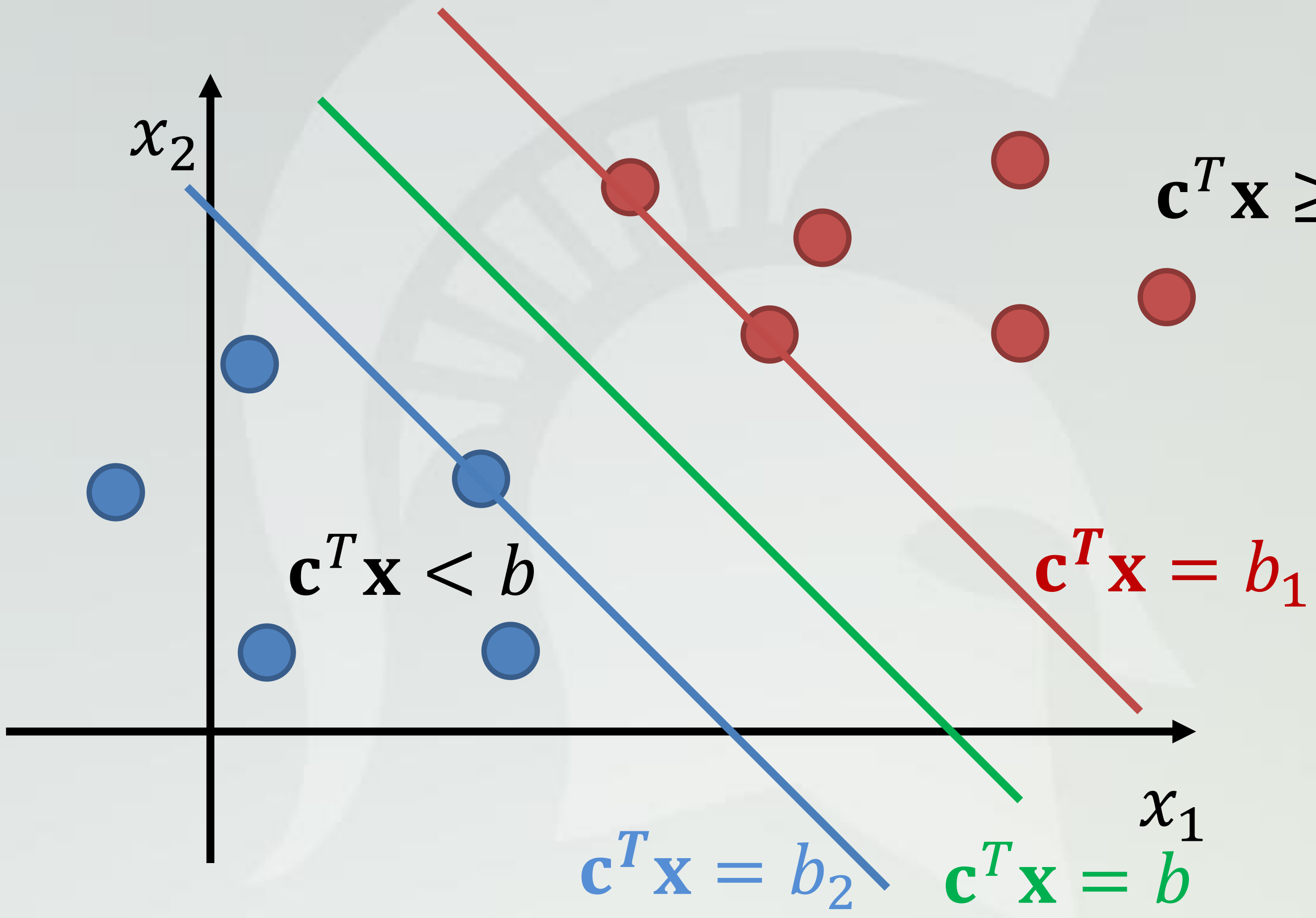
$\mathbf{c} = (c_0, c_1, \ldots, c_n)^T$



Decision Boundary

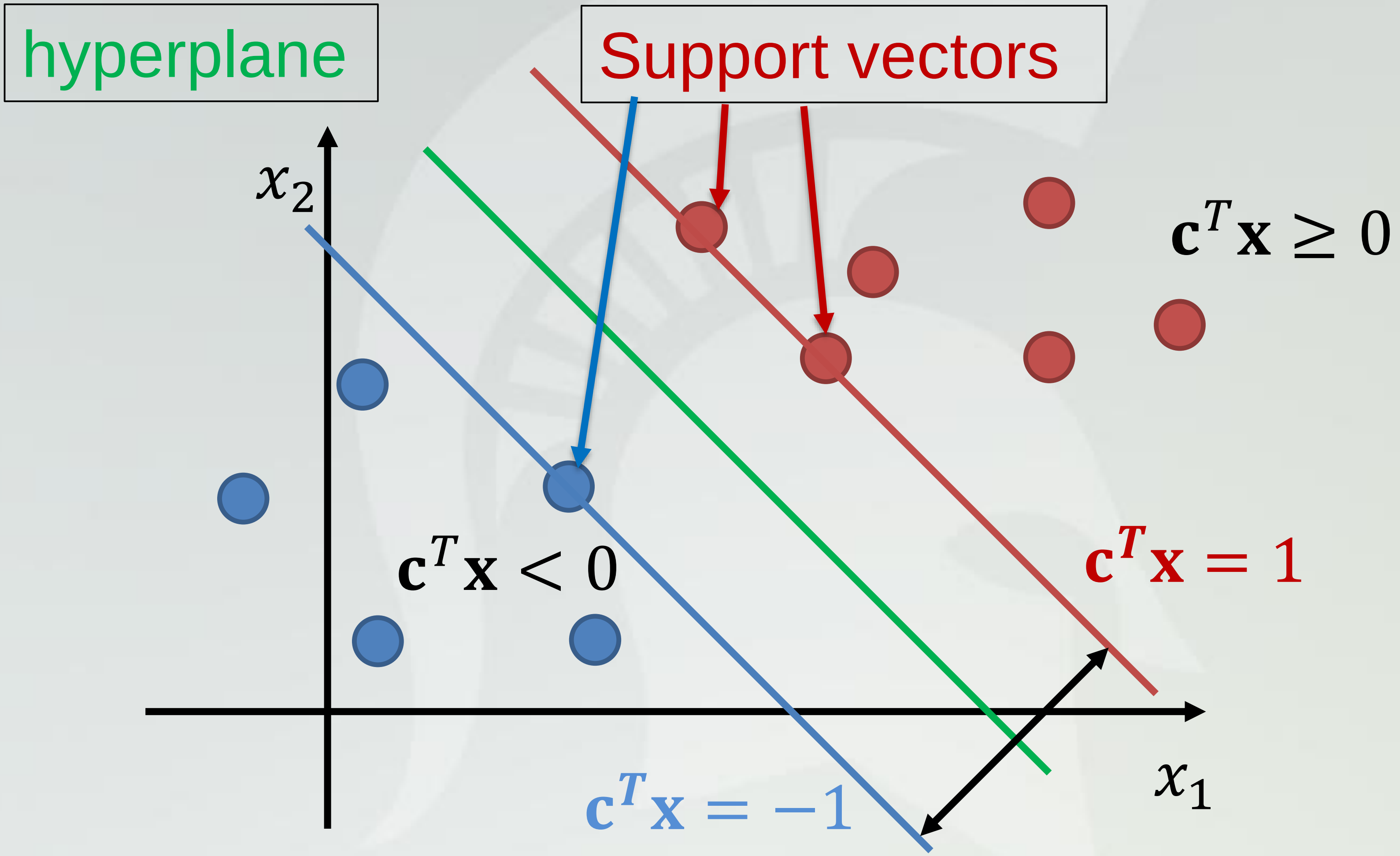


Decision Boundary



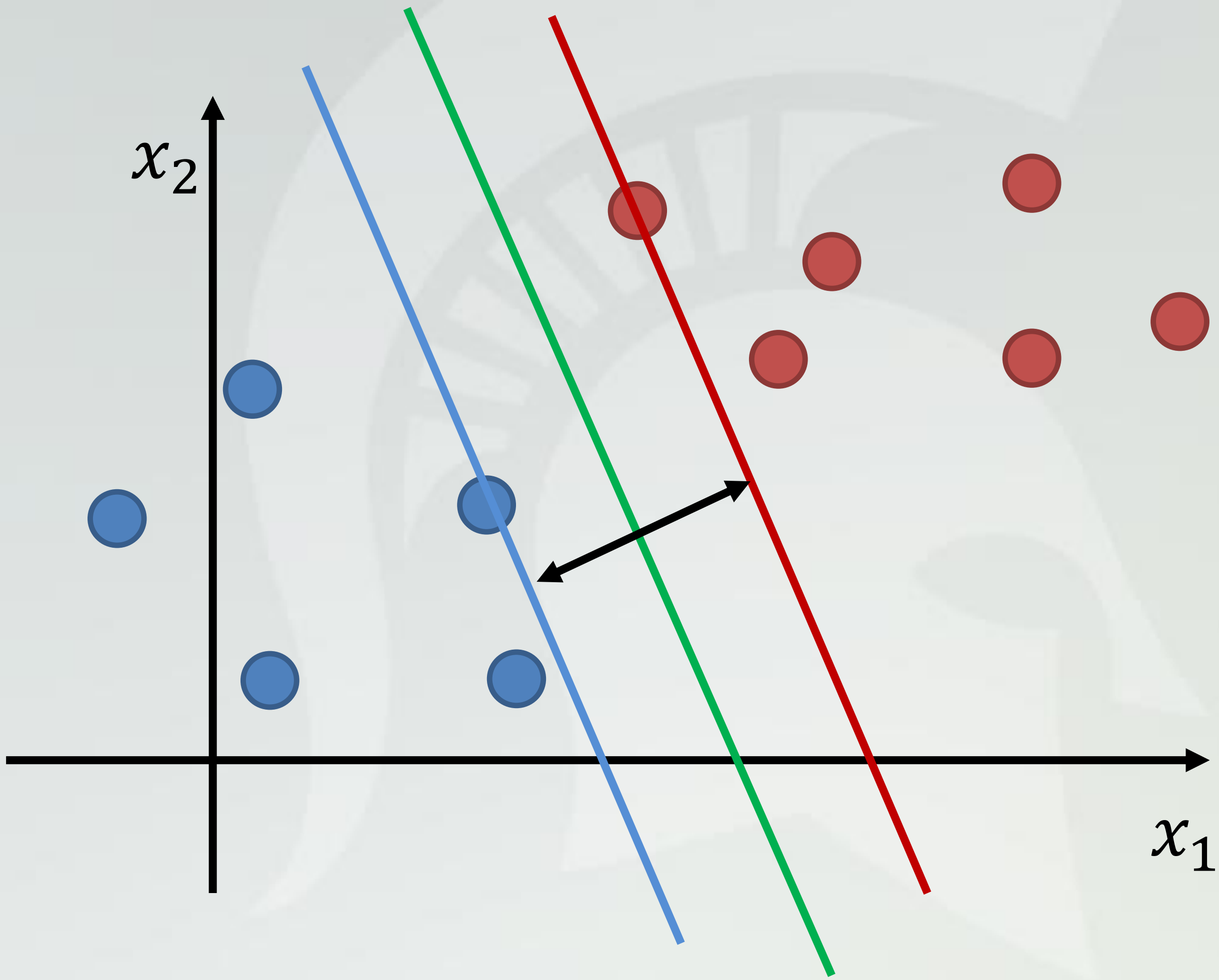
For simplicity: choose $b = 0$, $b_1 = 1$, $b_2 = -1$

Decision Boundary

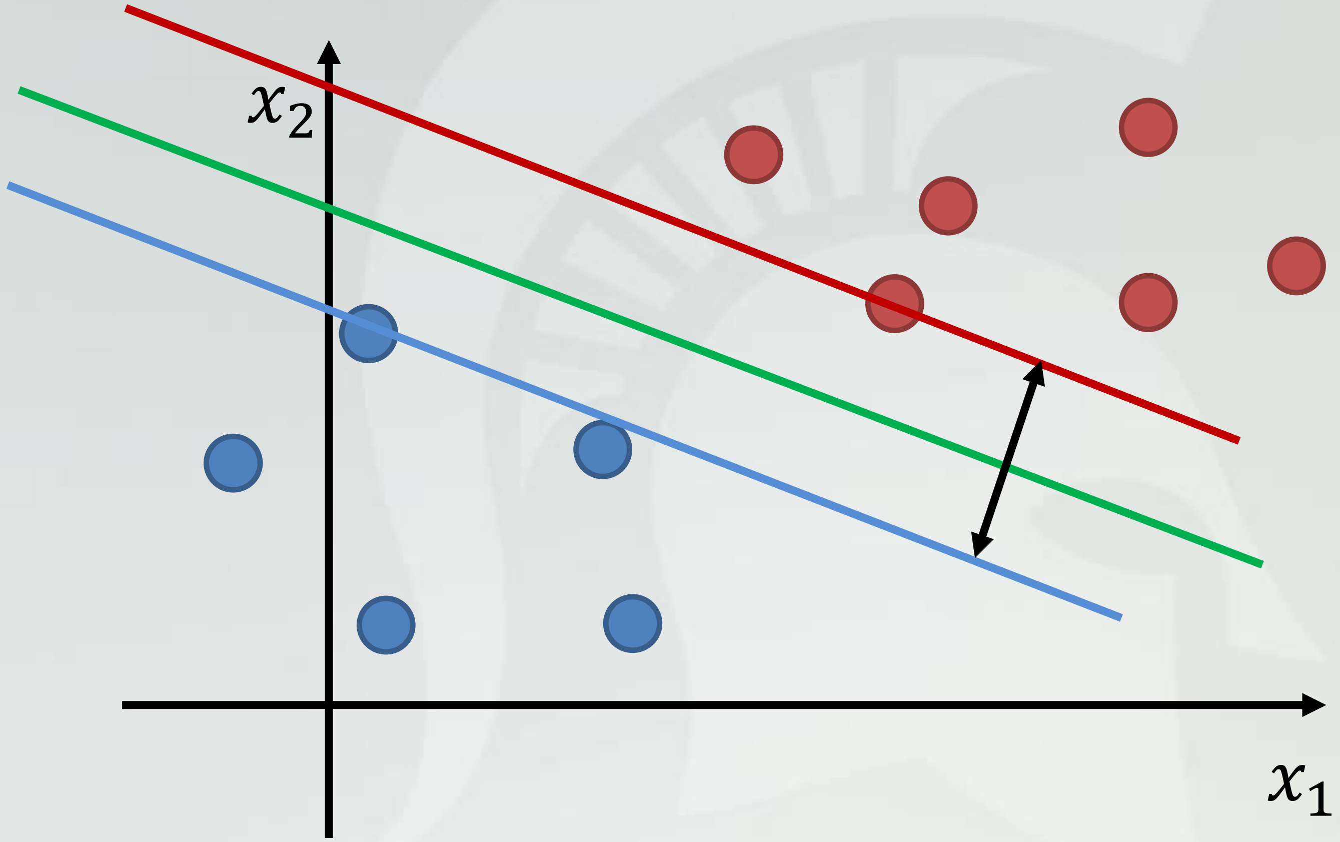


Margin: $\dfrac{2}{\sqrt{c_1^2+c_2^2+\cdots+c_n^2}}$

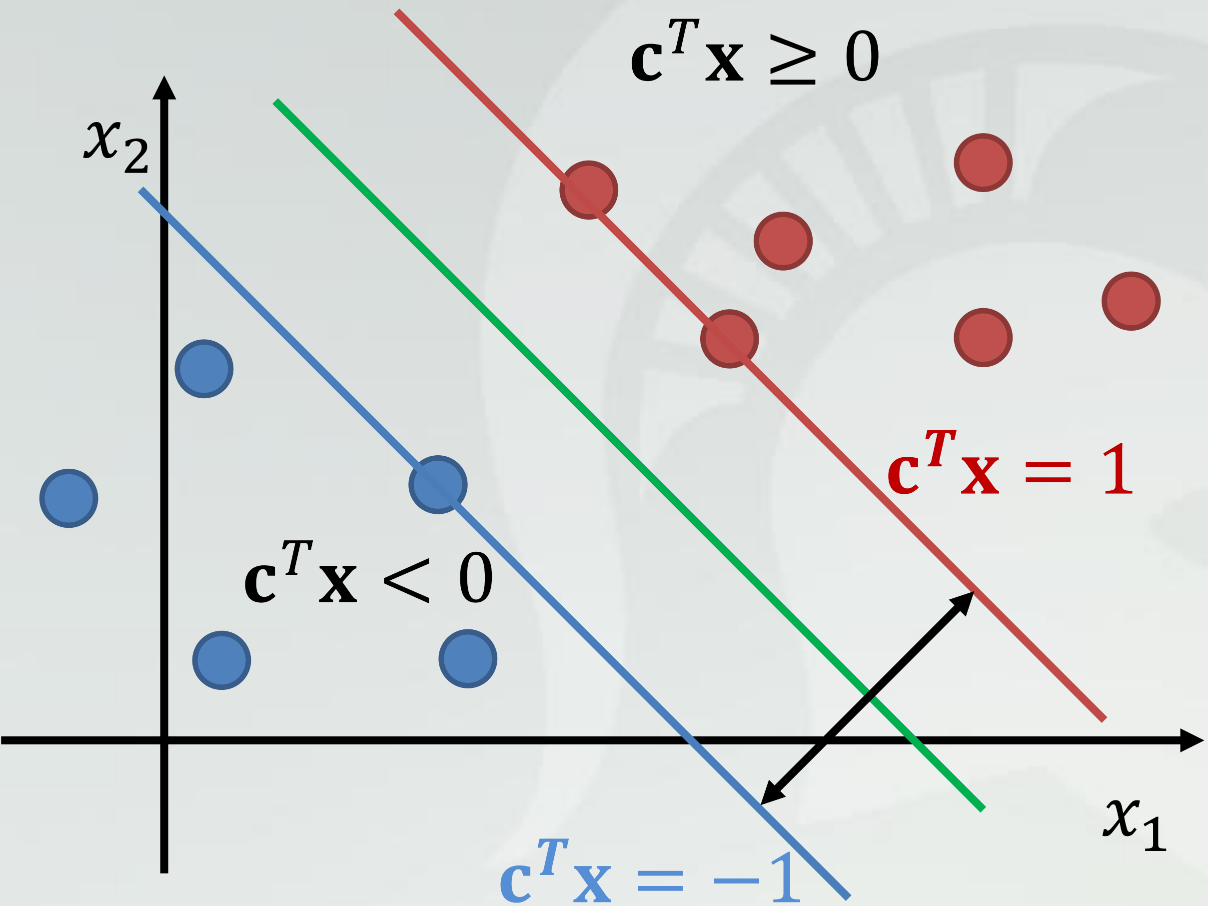
Decision Boundary



Decision Boundary



Optimization Objective



Margin: $\dfrac{2}{\sqrt{c_1^2+c_2^2+\cdots+c_n^2}}$

Maximize the margin:

$$\frac{2}{\sqrt{c_1^2+c_2^2+\cdots+c_n^2}}$$

Subject to

$\mathbf{c}^T\mathbf{x}^{(i)} \geq 1$ if $y^{(i)} = 1$

or

$\mathbf{c}^T\mathbf{x}^{(i)} \leq -1$ if $y^{(i)} = -1$

Optimization Objective

- Maximize

$$\frac{2}{\sqrt{c_1^2 + c_2^2 + \cdots + c_n^2}}$$

Subject to

$$\mathbf{c}^T \mathbf{x}^{(i)} \geq 1 \text{ if } \; y^{(i)} = 1$$

or

$$\mathbf{c}^T \mathbf{x}^{(i)} \leq -1 \text{ if } \; y^{(i)} = -1$$

- Equivalent to (dual problem):

Minimize:

$$\sqrt{c_1^2 + c_2^2 + \cdots + c_n^2}$$

Subject to $\mathbf{c}^T \mathbf{x}^{(i)} \geq 1$ if $y^{(i)} = 1$ or $\mathbf{c}^T \mathbf{x}^{(i)} \leq -1$ if $y^{(i)} = -1$

Optimization Objective

- Minimize:

$$\sqrt{c_1^2 + c_2^2 + \cdots + c_n^2}$$

Subject to $\mathbf{c}^T\mathbf{x}^{(i)} \geq 1$ if $y^{(i)} = 1$ or $\mathbf{c}^T\mathbf{x}^{(i)} \leq -1$ if $y^{(i)} = -1$

- Equivalent to

Minimize:

Loss function

$$\sqrt{c_1^2 + c_2^2 + \cdots + c_n^2}$$

Subject to $y^{(i)}\mathbf{c}^T\mathbf{x}^{(i)} \geq 1$

Predictor?

$$p_{\mathbf{c}}(\mathbf{x}) = \mathbf{c}^T\mathbf{x} = c_0 + c_1x_1 + \cdots + c_nx_n$$ Predictor

Minimize: Loss function

$$L(\mathbf{c}) = L(c_0, c_1, \ldots, c_n) = \sqrt{c_1^2 + c_2^2 + \cdots + c_n^2}$$

Subject to $y^{(i)}p_{\mathbf{c}}(\mathbf{x}^{(i)}) = y^{(i)}\mathbf{c}^T\mathbf{x}^{(i)} \geq 1$

- Classifier: Take threshold=0

  if $p_{\mathbf{c}}(\mathbf{x}) \geq 0$ then $y = 1$

  if $p_{\mathbf{c}}(\mathbf{x}) < 0$ then $y = -1$

Loss Function

$$p_{\mathbf{c}}(\mathbf{x}) = \mathbf{c}^T\mathbf{x} = c_0 + c_1x_1 + \cdots + c_nx_n$$ Predictor

Minimize:

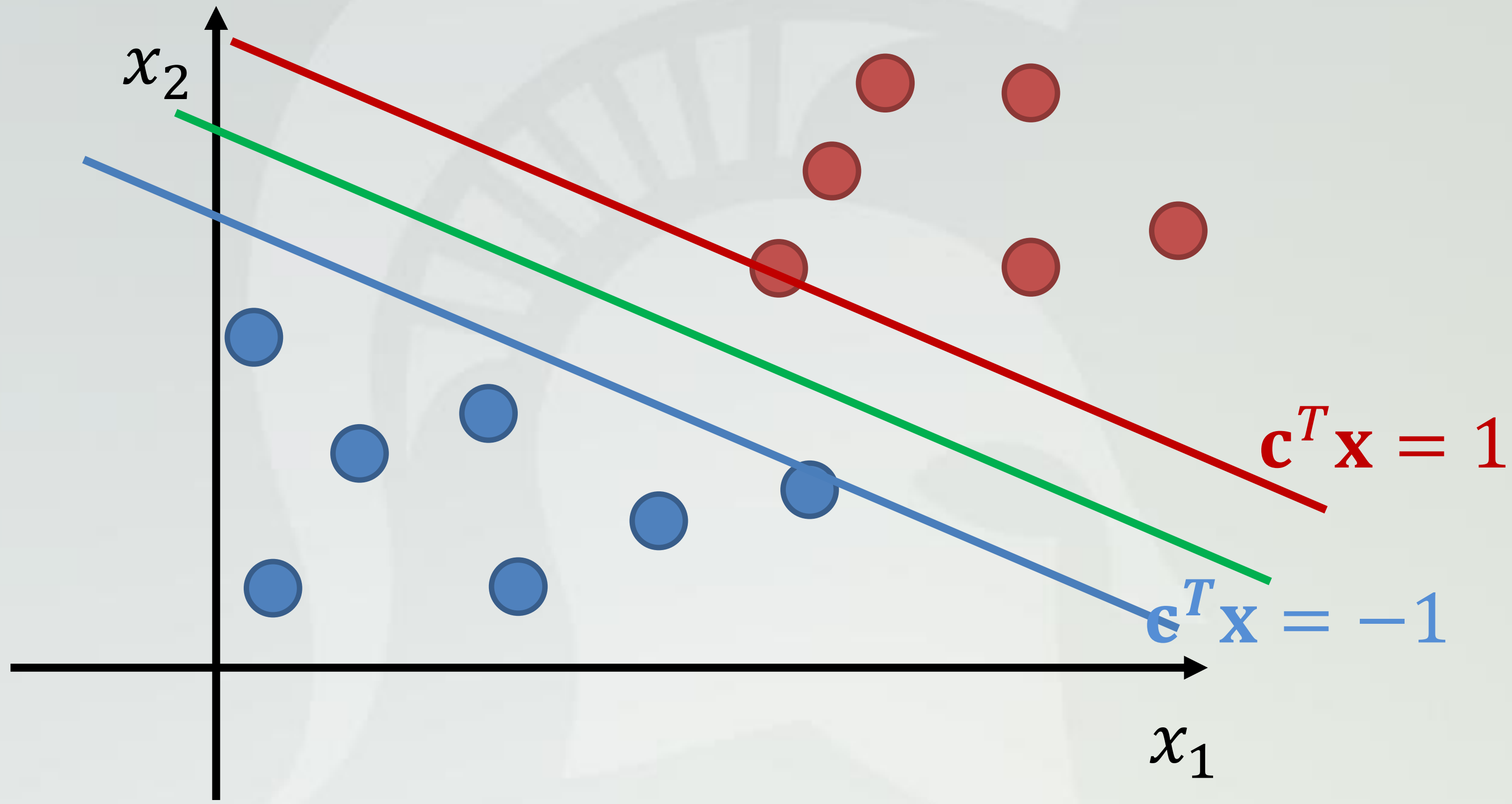
Loss function

$$L(\mathbf{c}) = L(c_0, c_1, \ldots, c_n) = \sqrt{c_1^2 + c_2^2 + \cdots + c_n^2}$$

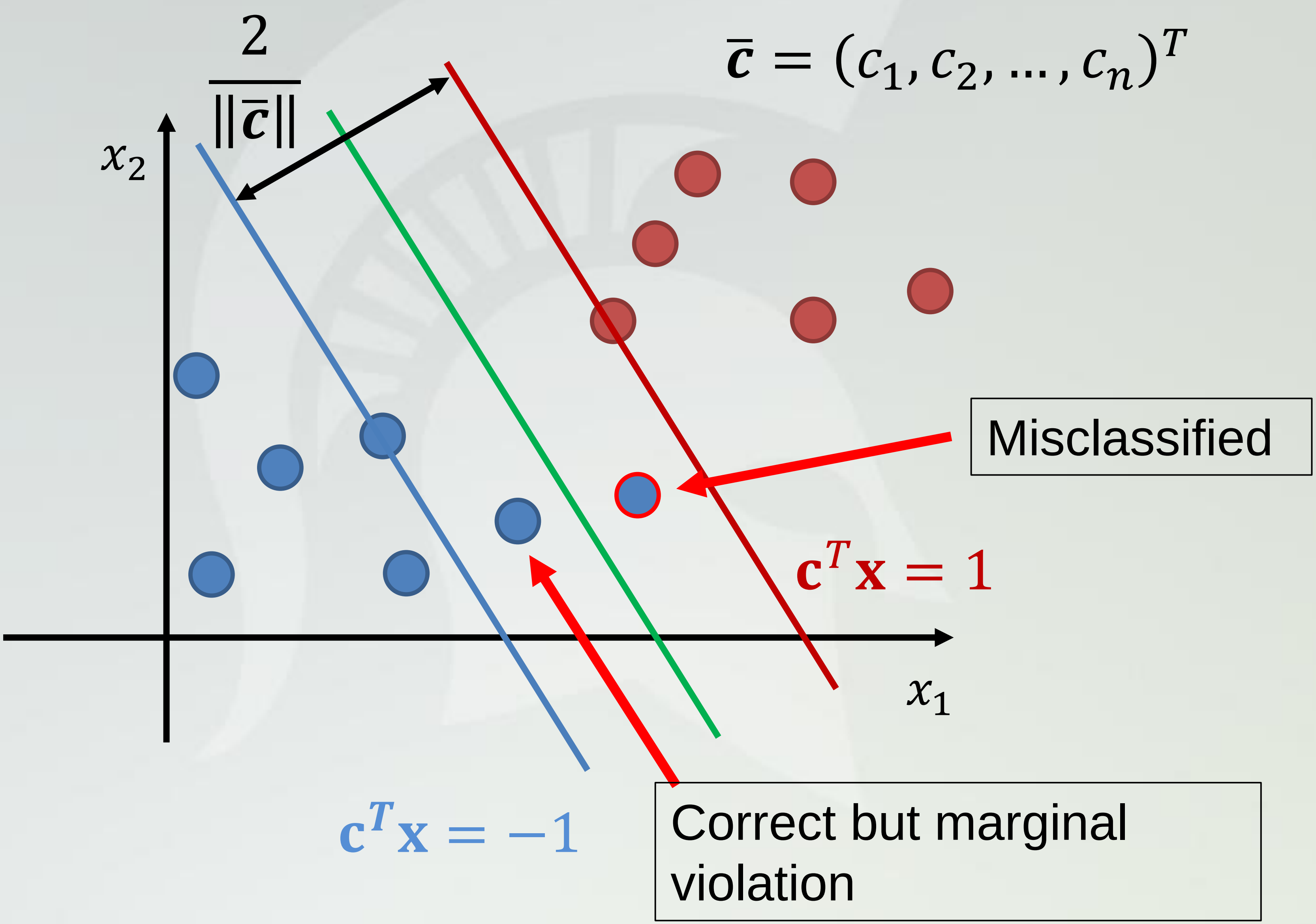
Subject to $y^{(i)}\mathbf{c}^T\mathbf{x}^{(i)} \geq 1$

Simplified condition

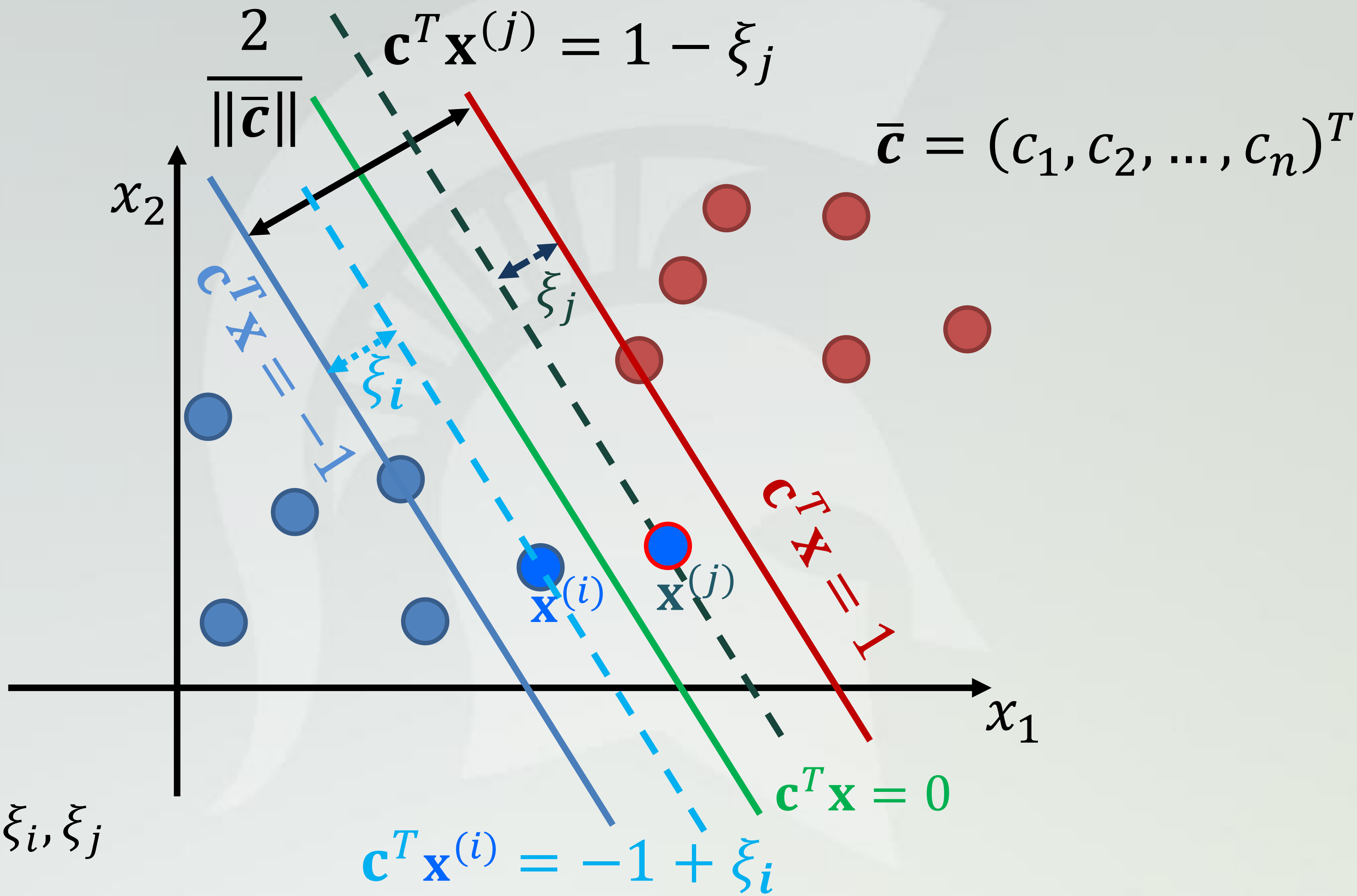
Hard Margin



Soft Margin


$\frac{2}{\|\bar{c}\|}$
$\bar{c} = (c_1, c_2, \ldots, c_n)^T$
x_2
x_1
Misclassified
$\mathbf{c}^T\mathbf{x} = 1$
$\mathbf{c}^T\mathbf{x} = -1$
Correct but marginal violation

Soft Margin



$0 \leq \xi_i, \xi_j$

$\xi_i < 1$ (Correct but marginal violation)

$\xi_j > 1$ (incorrect) If $\xi_k = 0$: perfect

Minimize ξ_k!

Loss Function for Soft Margin

Modified loss function

Predictor

$$p_{\mathbf{c}}(\mathbf{x}) = \mathbf{c}^T\mathbf{x} = c_0 + c_1x_1 + \cdots + c_nx_n$$

Minimize:

Loss function

$$L(\mathbf{c}) = L(c_0, c_1, \ldots, c_n) = \sqrt{c_1^2 + c_2^2 + \cdots + c_n^2}$$

Subject to $y^{(i)}\mathbf{c}^T\mathbf{x}^{(i)} \geq 1 - \xi_i$, with $\xi_i \geq 0$

Modified condition

Loss Function for Soft Margin

Predictor

$$p_{\mathbf{c}}(\mathbf{x}) = \mathbf{c}^T\mathbf{x} = c_0 + c_1 x_1 + \cdots + c_n x_n$$

Minimize:

Loss function

$$L(\mathbf{c}, \boldsymbol{\xi}) = L(c_0, c_1, \dots, c_n, \xi_1, \xi_2, \dots, \xi_M) =$$

$$\sqrt{c_1^2 + c_2^2 + \cdots + c_n^2} + \sum_{i=1}^{M} \xi_i$$

Regularization

Subject to $y^{(i)}\mathbf{c}^T\mathbf{x}^{(i)} \geq 1 - \xi_i$, with $\xi_i \geq 0$

Loss Function for Soft Margin

Predictor

$$p_{\mathbf{c}}(\mathbf{x}) = \mathbf{c}^T\mathbf{x} = c_0 + c_1x_1 + \cdots + c_nx_n$$

Minimize:

Loss function

$$L(\mathbf{c}, \boldsymbol{\xi}) = L(c_0, c_1, \dots, c_n, \xi_1, \xi_2, \dots, \xi_M) =$$

$$\sqrt{c_1^2 + c_2^2 + \cdots + c_n^2} + \lambda \sum_{i=1}^{M} \xi_i$$

Subject to $y^{(i)}\mathbf{c}^T\mathbf{x}^{(i)} \geq 1 - \xi_i,\;$ with $\xi_i \geq 0$

λ: regularization parameter

If $\lambda \to \infty$?

then $\sum_{i=1}^{M} \xi_i \to 0 \Rightarrow \xi_i = 0 \Rightarrow$ hard margin

Simplify Loss Function

$$p_{\mathbf{c}}(\mathbf{x}) = \mathbf{c}^T\mathbf{x} = c_0 + c_1x_1 + \cdots + c_nx_n$$

Minimize:

$$L(\mathbf{c}, \boldsymbol{\xi}) = L(c_0, c_1, \ldots, c_n, \xi_1, \xi_2, \ldots, \xi_M) =$$

$$\sqrt{c_1^2 + c_2^2 + \cdots + c_n^2} + \lambda \sum_{i=1}^{M} \xi_i$$

Subject to $y^{(i)}\mathbf{c}^T\mathbf{x}^{(i)} \geq 1 - \xi_i$, with $\xi_i \geq 0$

Hinge loss

$$\xi_i = \max(0, 1 - y^{(i)}\mathbf{c}^T\mathbf{x}^{(i)})$$

Simplify Loss Function

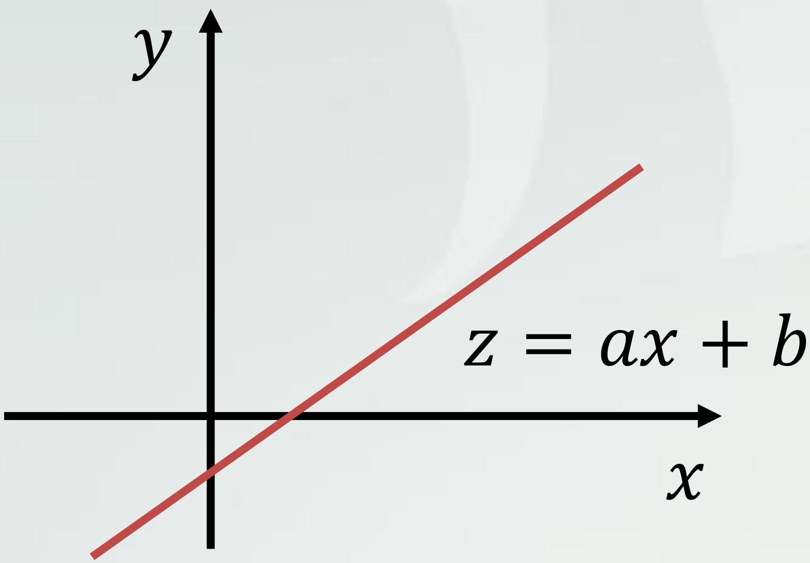
$$p_{\mathbf{c}}(\mathbf{x}) = \mathbf{c}^T\mathbf{x} = c_0 + c_1x_1 + \cdots + c_nx_n$$

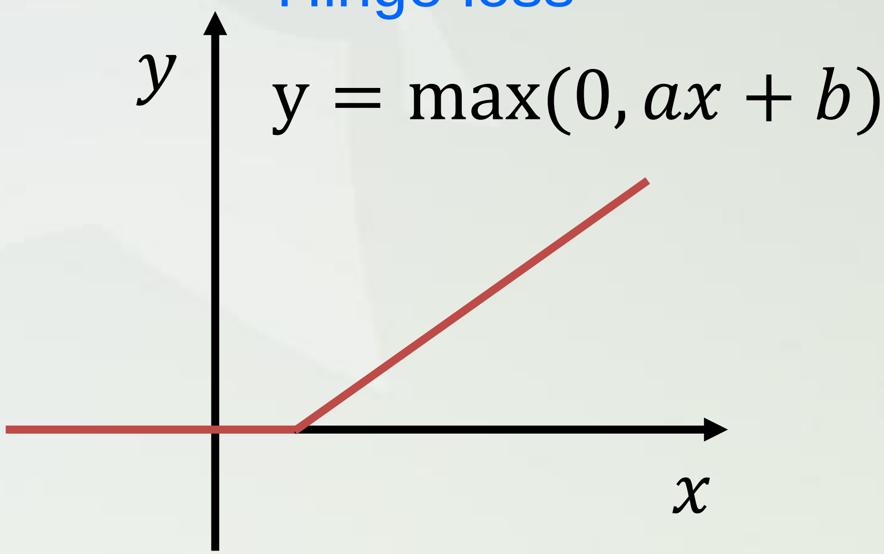
Minimize:

$$L(\mathbf{c}, \boldsymbol{\xi}) = L(c_0, c_1, \dots, c_n, \xi_1, \xi_2, \dots, \xi_M) =$$

$$\sqrt{c_1^2 + c_2^2 + \cdots + c_n^2} + \lambda \sum_{i=1}^{M} \max(0, 1 - y^{(i)}\mathbf{c}^T\mathbf{x}^{(i)})$$

Hinge loss





How to Minimize Loss Function

Minimize:

$$L(\mathbf{c}) = L(c_0, c_1, \dots, c_n) =$$

$$\sqrt{c_1^2 + c_2^2 + \cdots + c_n^2} + \lambda \sum_{i=1}^{M} \max(0, 1 - y^{(i)} \mathbf{c}^T \mathbf{x}^{(i)})$$

- Our loss function is convex

Convex

Convex

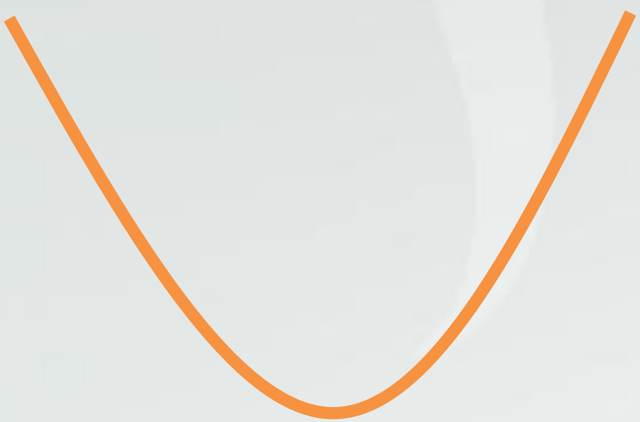
Not Convex

How to Minimize Loss Function

Minimize:

$$L(\mathbf{c}) = L(c_0, c_1, \dots, c_n) =$$

$$\sqrt{c_1^2 + c_2^2 + \cdots + c_n^2} + \lambda \sum_{i=1}^{M} \max(0, 1 - y^{(i)} \mathbf{c}^T \mathbf{x}^{(i)})$$



Convex

+

Convex

= Convex

How to Minimize Loss Function

Minimize:

$$L(\mathbf{c}) = L(c_0, c_1, \ldots, c_n) =$$

$$\sqrt{c_1^2 + c_2^2 + \cdots + c_n^2} + \lambda \sum_{i=1}^{M} \max(0, 1 - y^{(i)} \mathbf{c}^T \mathbf{x}^{(i)})$$

- The loss function is convex
- In convex function, local minimum is the global minimum
- Loss function can be optimized by
  - Quadratic optimization method
  - Gradient descent (continuity condition)?

Sub-gradient descent

For non-differentiable objective functions

$$\mathbf{c} := \mathbf{c} - \alpha \nabla_{\mathbf{c}} L(\mathbf{c})$$

$$:= \mathbf{c} - \alpha \nabla_{\mathbf{c}} \left(\sqrt{c_1^2 + c_2^2 + \cdots + c_n^2} + \lambda \sum_{i=1}^{M} \max(0, 1 - y^{(i)} \mathbf{c}^T \mathbf{x}^{(i)}) \right)$$

$$:= \mathbf{c} - \alpha \nabla_{\mathbf{c}} \left(\sqrt{c_1^2 + c_2^2 + \cdots + c_n^2} \right) - \lambda \sum_{i=1}^{M} \nabla_{\mathbf{c}} \left(\max(0, 1 - y^{(i)} \mathbf{c}^T \mathbf{x}^{(i)}) \right)$$