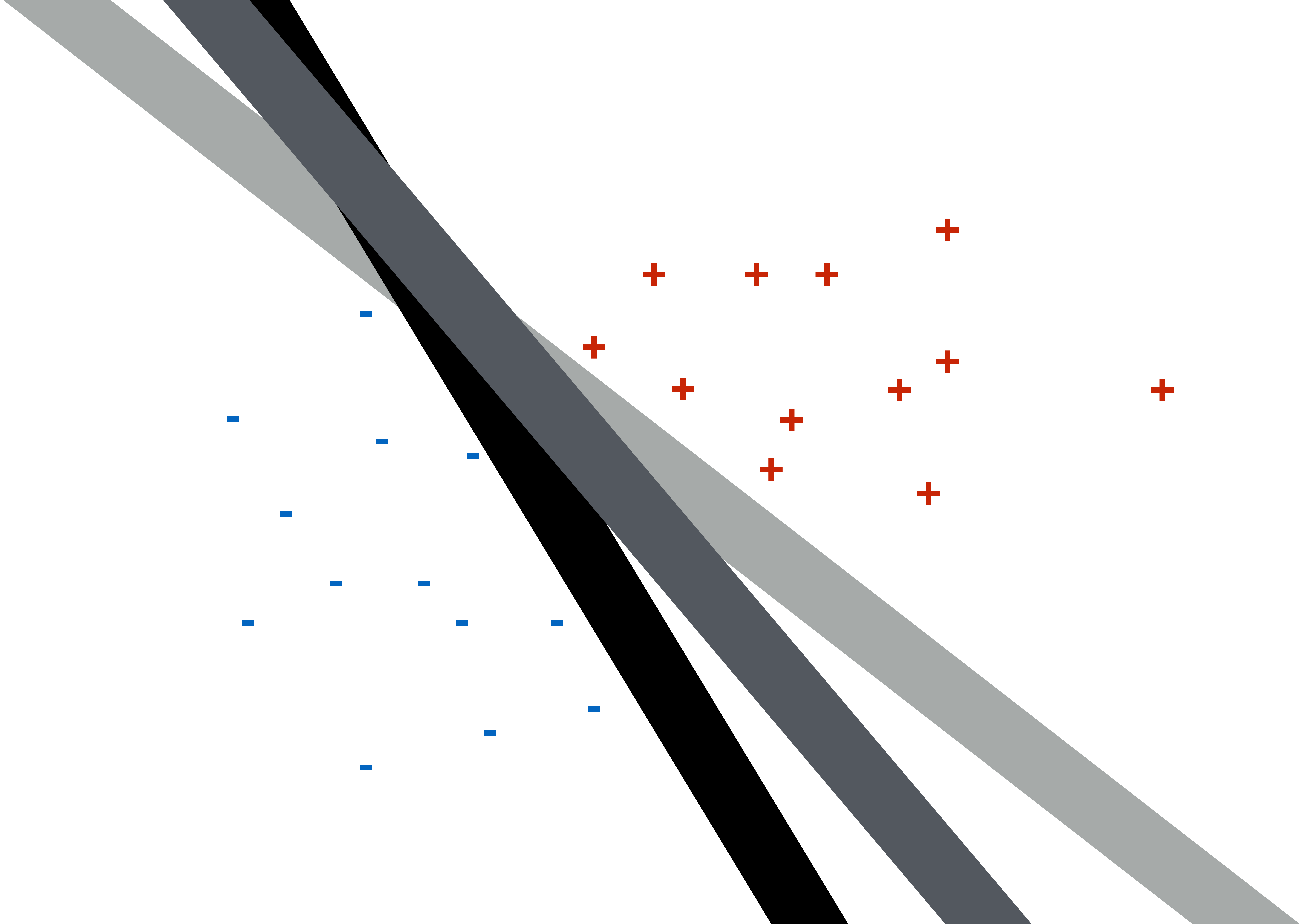
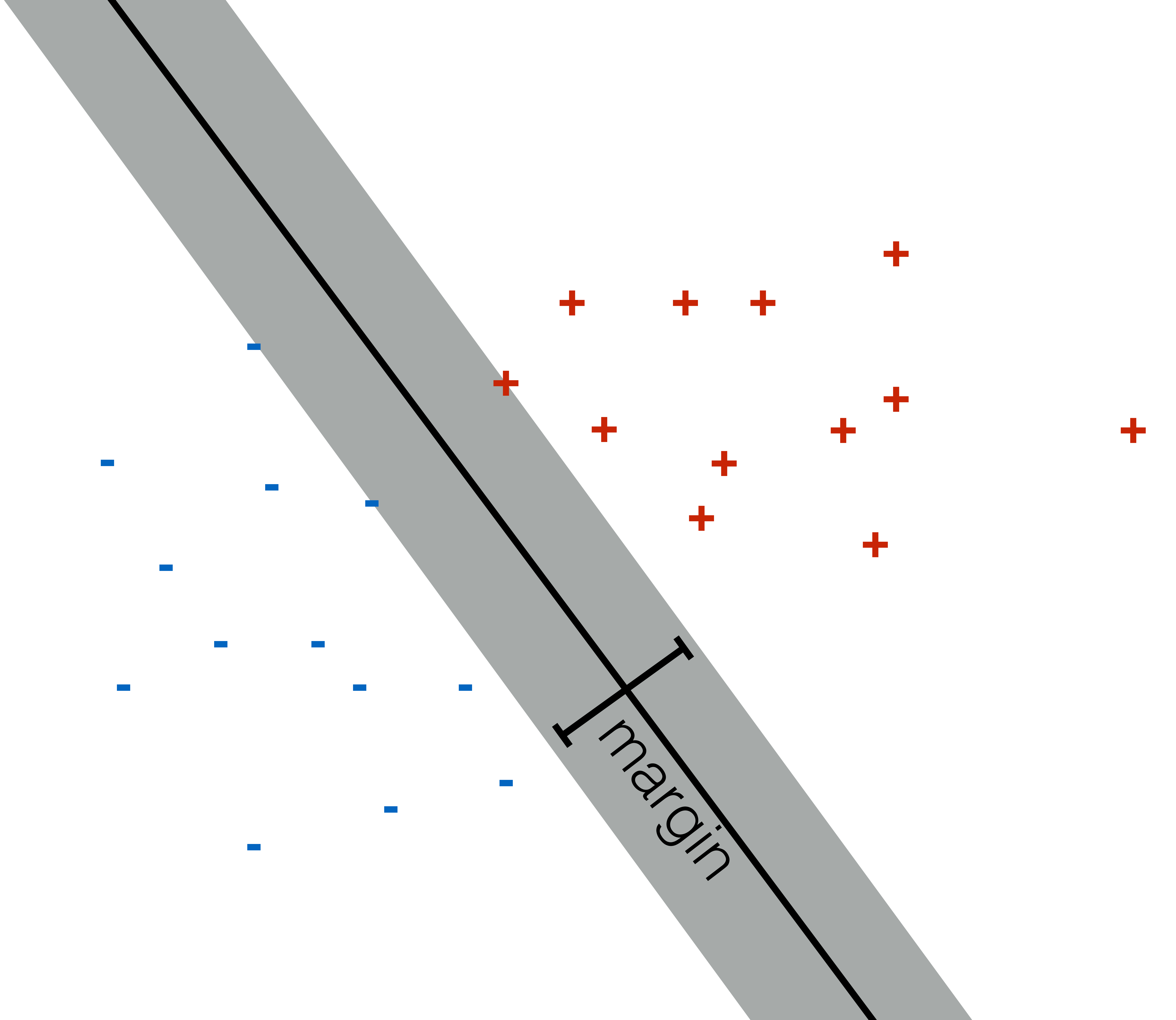
Support Vector Machines

Machine Learning
CS5824/ECE5424
Bert Huang
Virginia Tech

Outline

- Large-margin and model complexity
- Formalizing large margin
- Quadratic program form
- Soft-margin
- Non-linearity



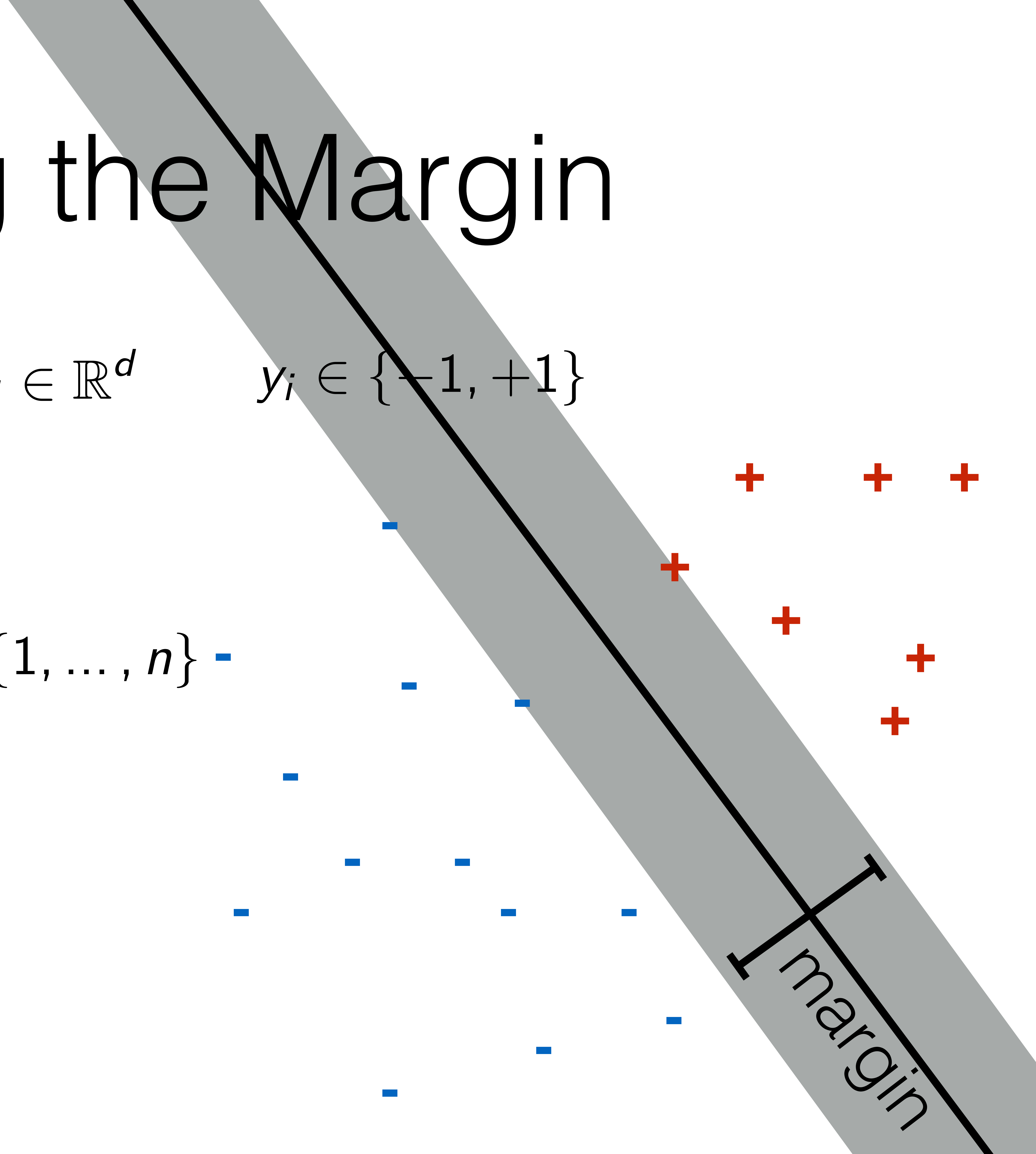

margin

Quantifying the Margin

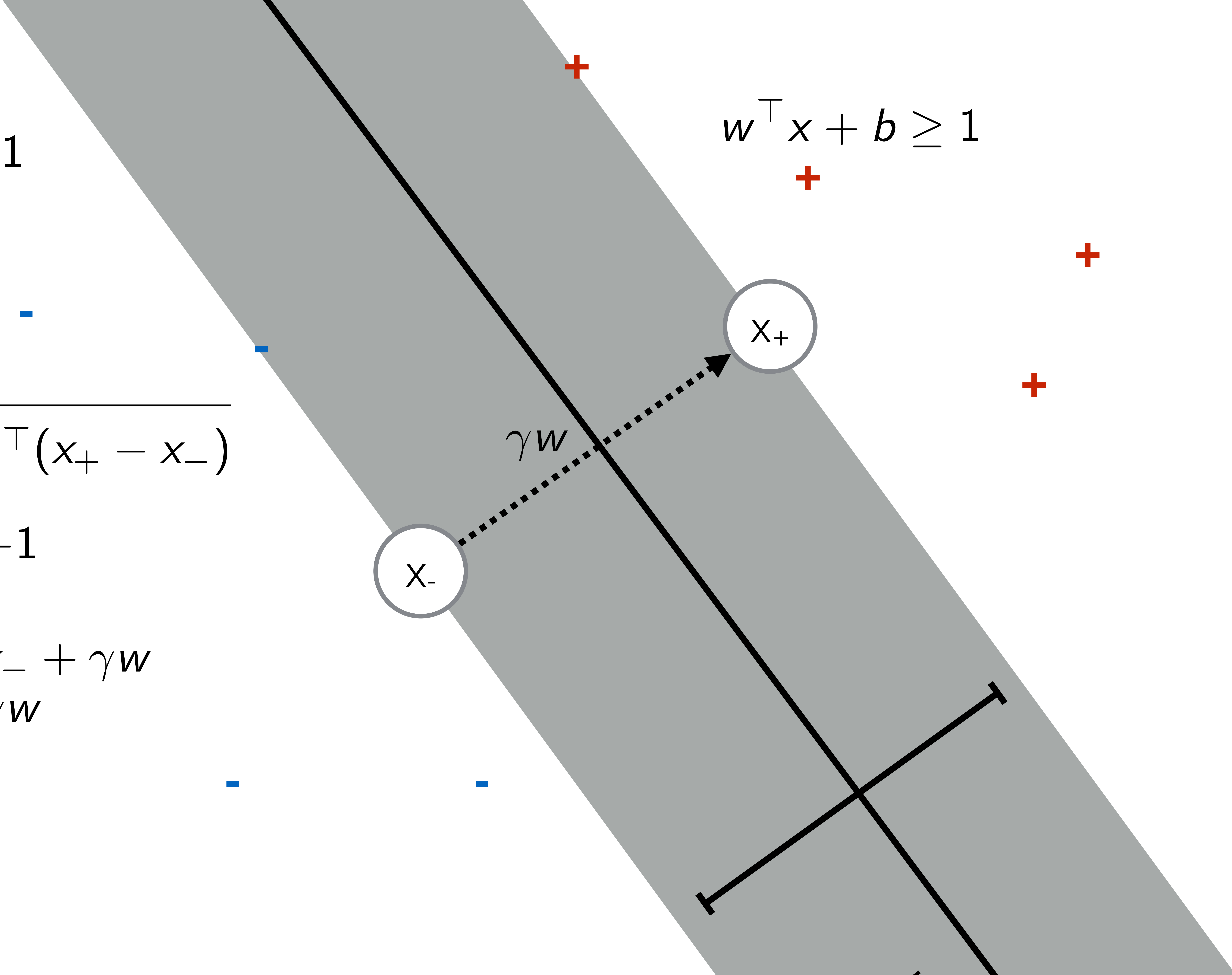
$D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ $\quad x_i \in \mathbb{R}^d$ $\quad y_i \in \{-1, +1\}$

~~$y_i(w^\top x_i + b) \geq 0$~~

$y_i(w^\top x_i + b) \geq 1 \quad \forall i \in \{1, \ldots, n\}$



$$w^\top x + b \leq -1$$



$$\text{margin} = \sqrt{(x_+ - x_-)^\top (x_+ - x_-)}$$

$$w^\top x_- + b = -1$$

$$w^\top x_+ + b = 1$$

$$x_+ = x_- + \gamma w$$

$$x_+ - x_- = \gamma w$$

$$\text{margin} = \sqrt{(x_+ - x_-)^\top (x_+ - x_-)}$$

$$w^\top x_- + b = -1$$

$$w^\top x_+ + b = 1$$

$$x_+ = x_- + \gamma w$$

$$x_+ - x_- = \gamma w$$

$$w^\top (x_- + \gamma w) + b = 1$$

$$w^\top x_- + b + \gamma w^\top w = 1$$

$$-1 + \gamma w^\top w = 1$$

$$x_+ - x_- = \frac{2w}{w^\top w}$$

$$\gamma = \frac{2}{w^\top w}$$

$$\text{margin} = \sqrt{\frac{4w^\top w}{w^\top w \times w^\top w}} = \frac{2}{\sqrt{w^\top w}}$$

Large-Margin Linear Classification

$$\max_{w \in \mathbb{R}^d} \quad \frac{2}{\sqrt{w^\top w}}$$

$$\text{s.t.} \quad y_i(w^\top x_i + b) \geq 1 \qquad \forall i \in \{1, \ldots, n\}$$

Large-Margin Linear Classification

$$\min_{w \in \mathbb{R}^d} \quad \frac{1}{2} w^\top w$$

$$\text{s.t.} \quad y_i(w^\top x_i + b) \geq 1 \qquad \forall i \in \{1, \ldots, n\}$$

Quadratic Programming

$$\min_{x} \quad \frac{1}{2}x^\top Hx + f^\top x$$ quadratic objective

$$\text{s.t.} \quad A_{\text{ineq}}x \leq b_{\text{ineq}}$$ linear inequality constraints

$$A_{\text{eq}}x = b_{\text{eq}}$$ linear equality constraints

$$\min_{w \in \mathbb{R}^d} \quad \frac{1}{2} w^\top w$$

$$\text{s.t.} \quad y_i(w^\top x_i + b) \geq 1 \quad \forall i \in \{1, \ldots, n\}$$

Soft-Margin Form

$$\min_{w \in \mathbb{R}^d} \quad \frac{1}{2} w^\top w$$

$$\text{s.t.} \quad y_i(w^\top x_i + b) \geq 1 \qquad \forall i \in \{1, \ldots, n\}$$

Soft-Margin Form

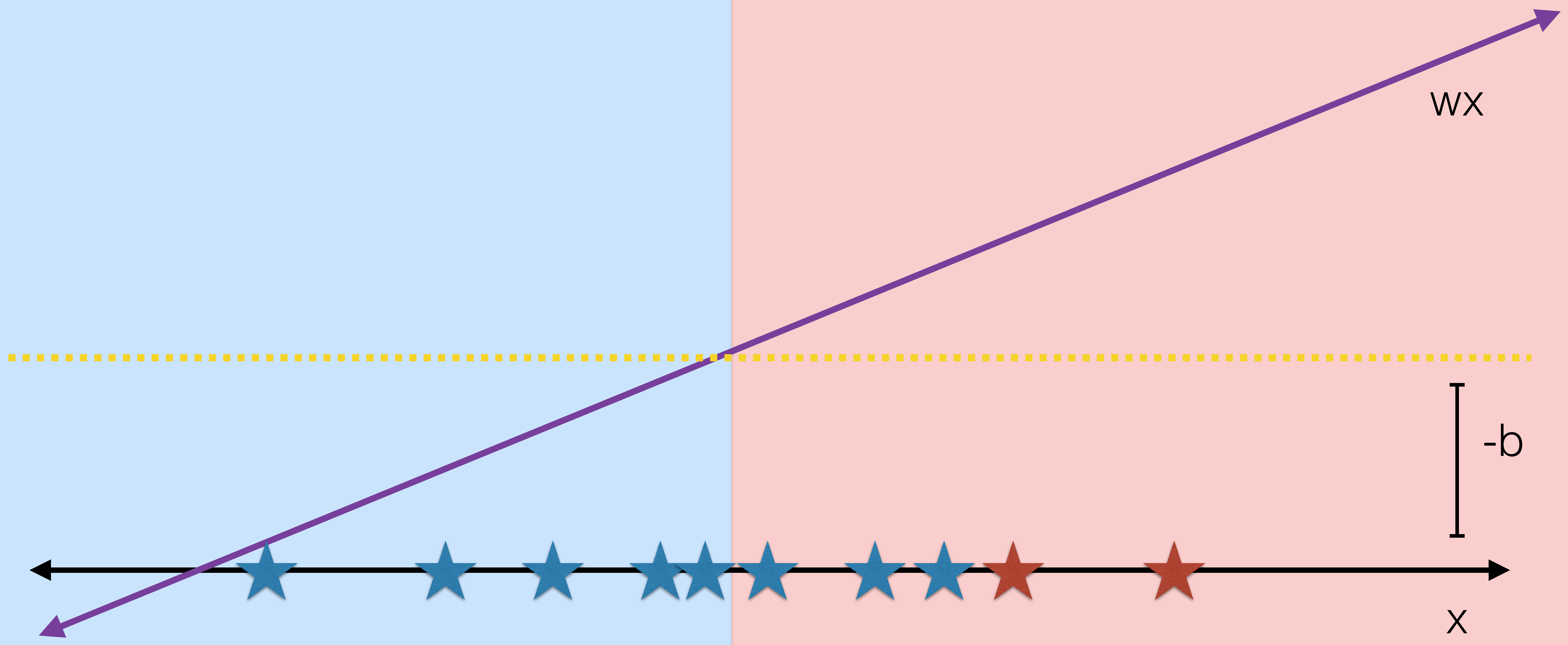
slack penalty

$$\min_{\substack{w \in \mathbb{R}^d \\ \xi \geq 0}} \quad \frac{1}{2} w^\top w + C \sum_{i=1}^{n} \xi_i$$

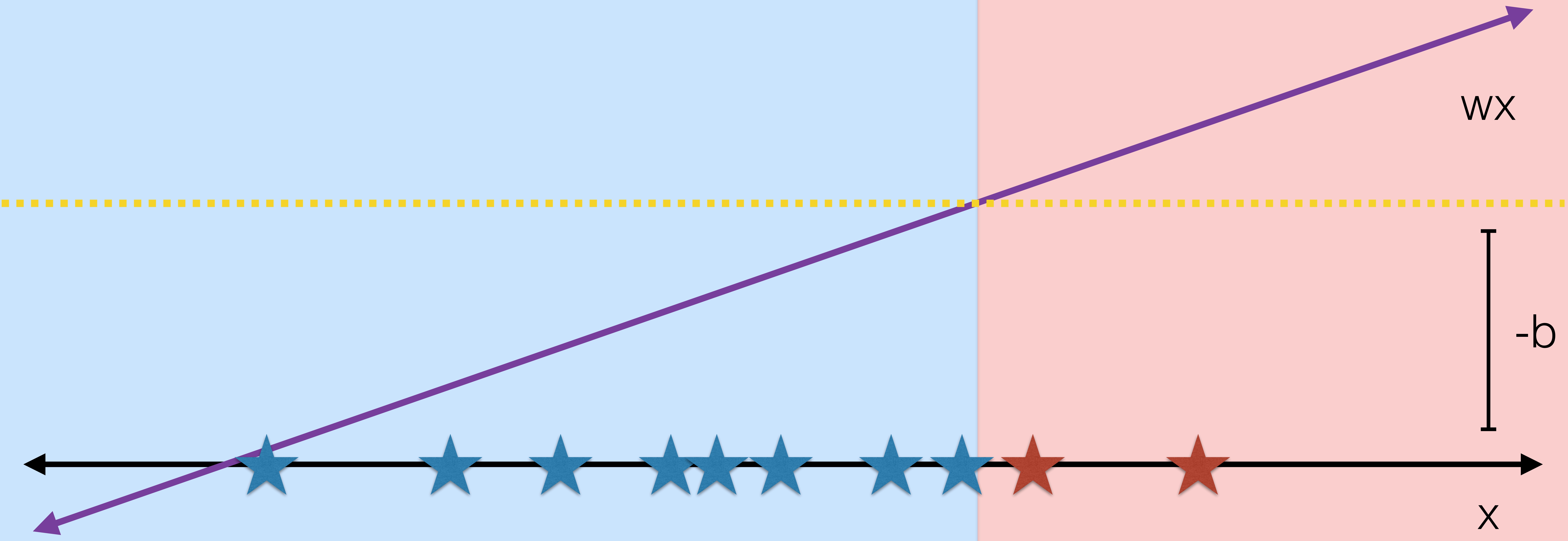
$$\text{s.t.} \quad y_i(w^\top x_i + b) \geq 1 - \xi_i \qquad \forall i \in \{1, \ldots, n\}$$

slack variables

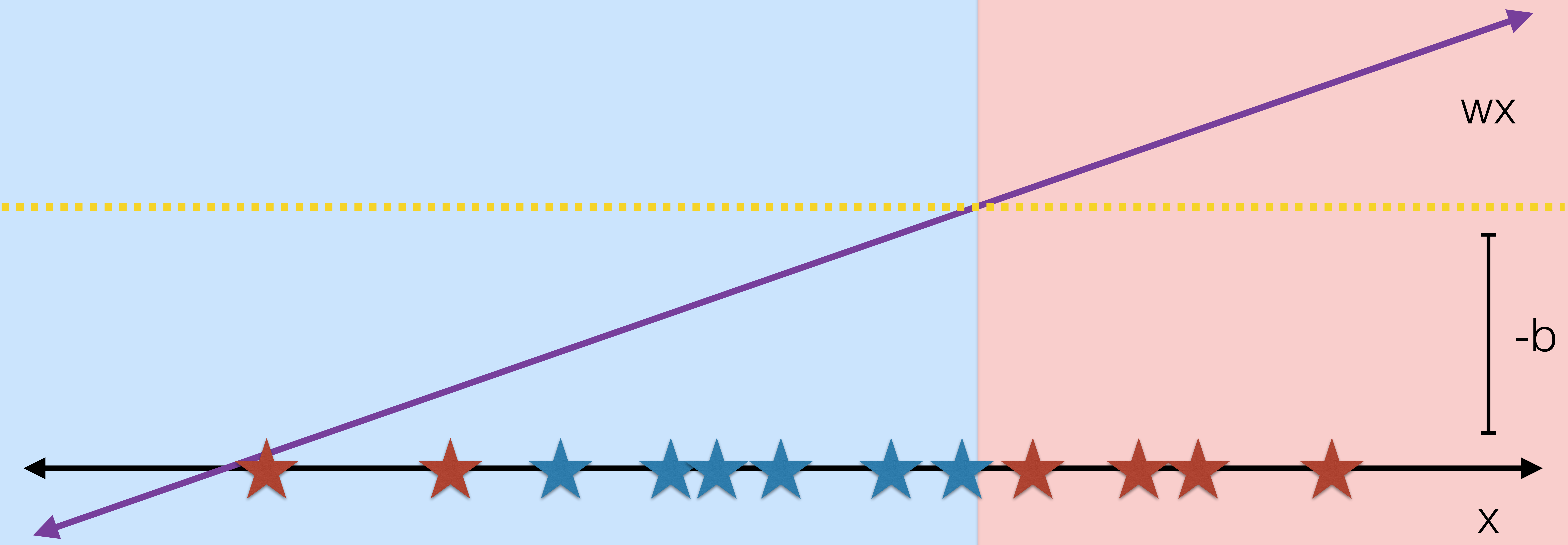
Nonlinear Decision Boundary



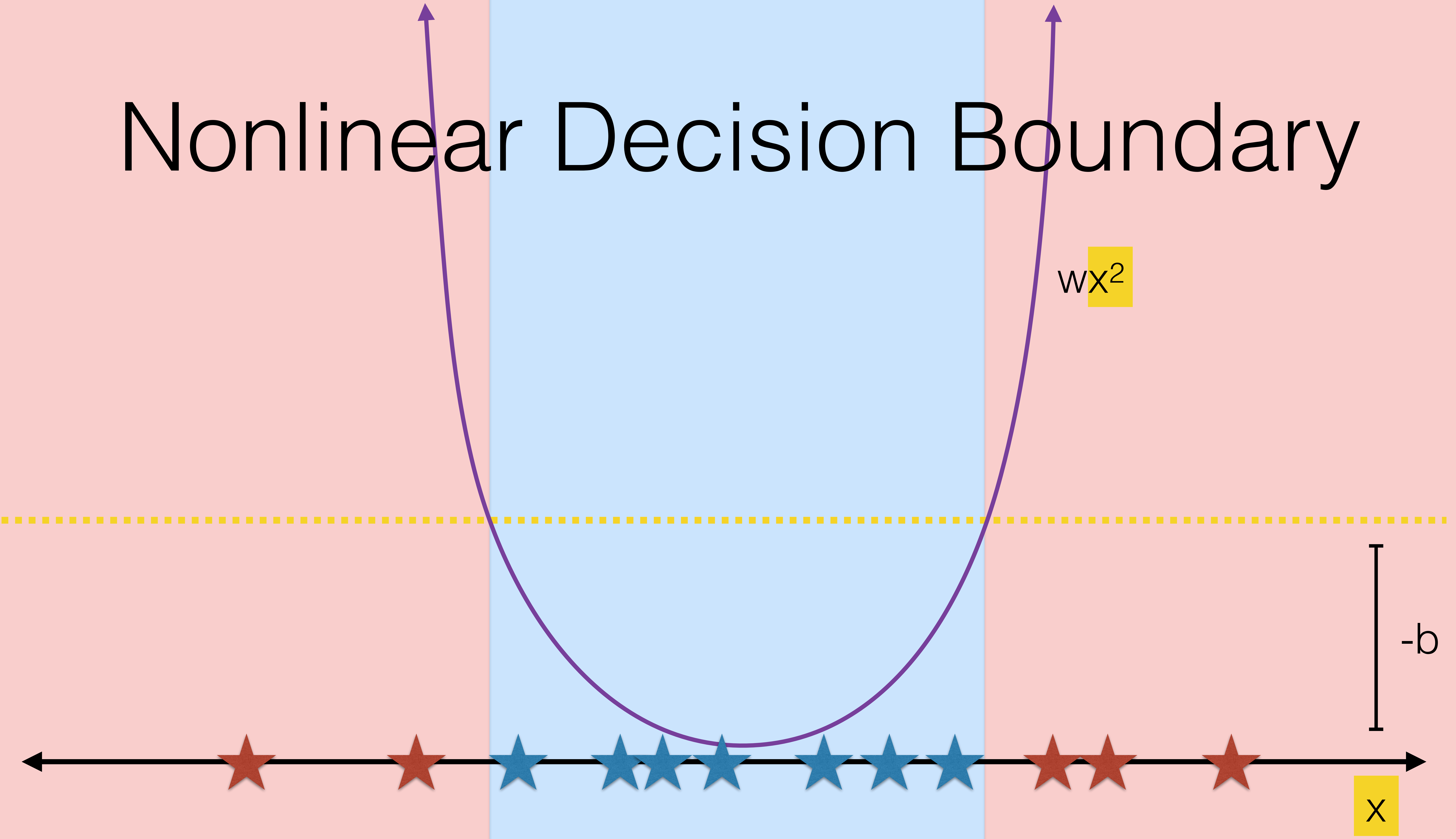
Nonlinear Decision Boundary



Nonlinear Decision Boundary



Nonlinear Decision Boundary



Polynomial Feature Map

$$\Phi(x) = [x^1, \dots, x^d, x^1x^1, \dots, x^1x^d, \dots, x^dx^1, \dots, x^dx^d, \dots]^\top$$

Third-order terms $\{x^1x^1x^1, x^1x^1x^2, \dots, x^1x^dx^d, \dots, x^dx^dx^d\}$

Fourth-order terms $\{x^ix^jx^kx^\ell | i, j, k, \ell \in \{1, \dots, d\}\}$

$$|\Phi(x)| = \sum_{a=1}^{M} d^a = O(d^M)$$

Soft-Margin Form w/ Feature Map

slack penalty

$$\min_{\substack{w \in \mathbb{R}^d \\ \xi \geq 0}} \quad \frac{1}{2} w^\top w + C \sum_{i=1}^{n} \xi_i$$

$$\text{s.t.} \quad y_i(w^\top \Phi(x_i) + b) \geq 1 - \xi_i \qquad \forall i \in \{1, \ldots, n\}$$

slack variables

(We rarely want to use this form.)

Summary

- Large-margin and model complexity
- Formalizing large margin
- Quadratic program form
- Soft-margin
- Feature maps for non-linearity

Optimization

- “Off-the-shelf” quadratic programming solvers
  - (usually interior-point methods with barrier functions)
- Gradient approaches using hinge-loss interpretation of slack penalty
- Dual form optimization
  - Leads to **kernel trick**