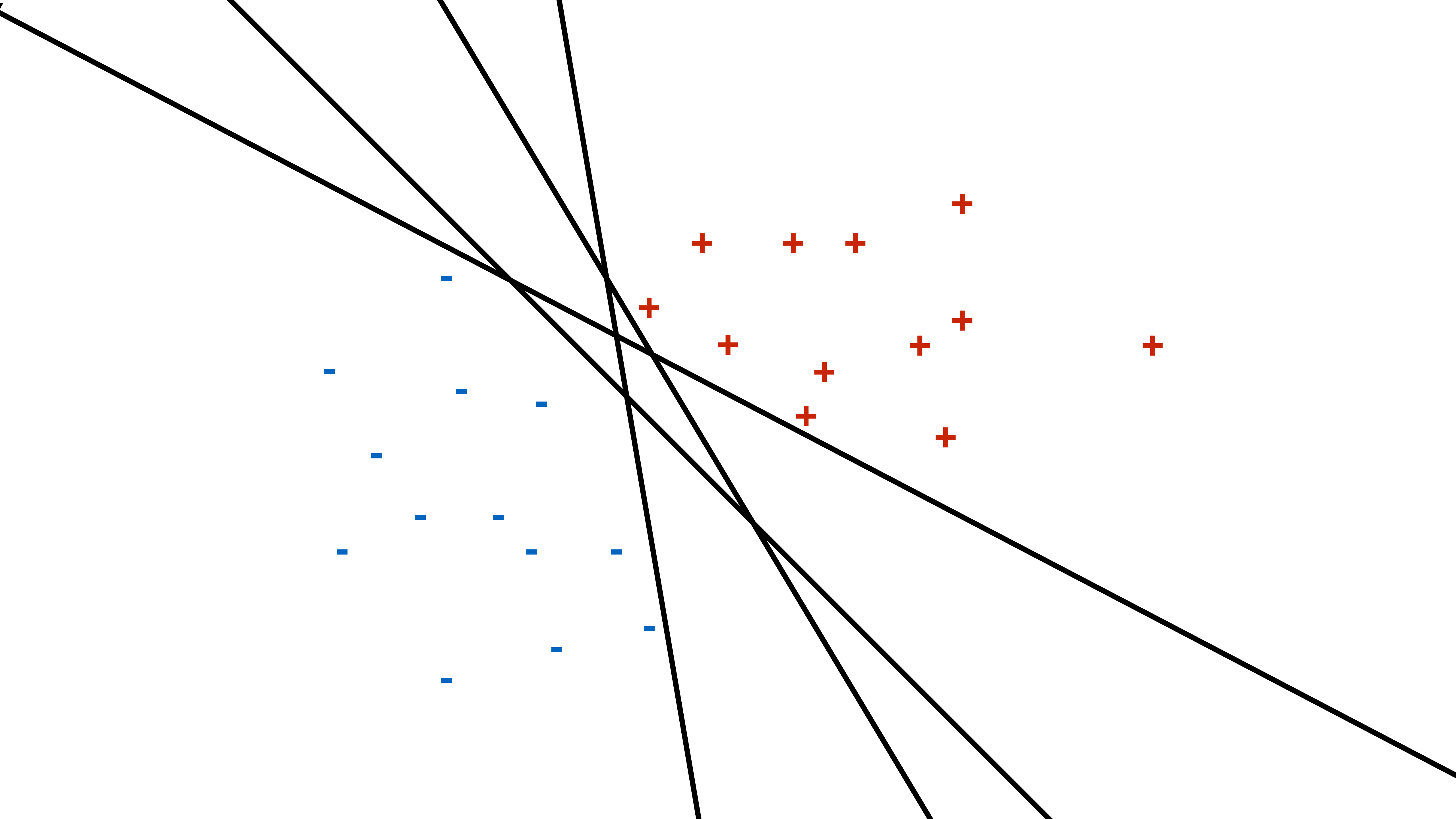


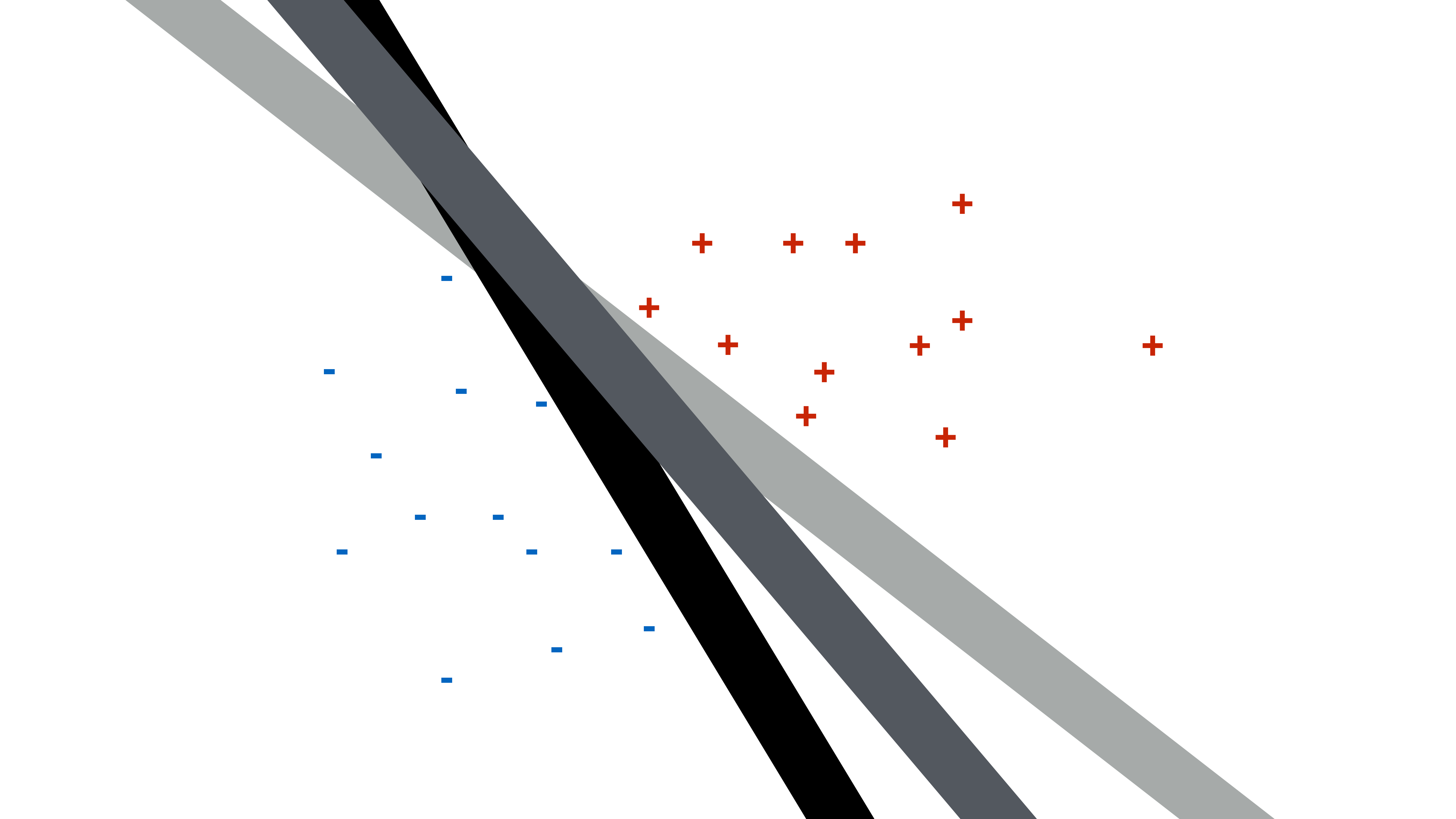
Support Vector Machines

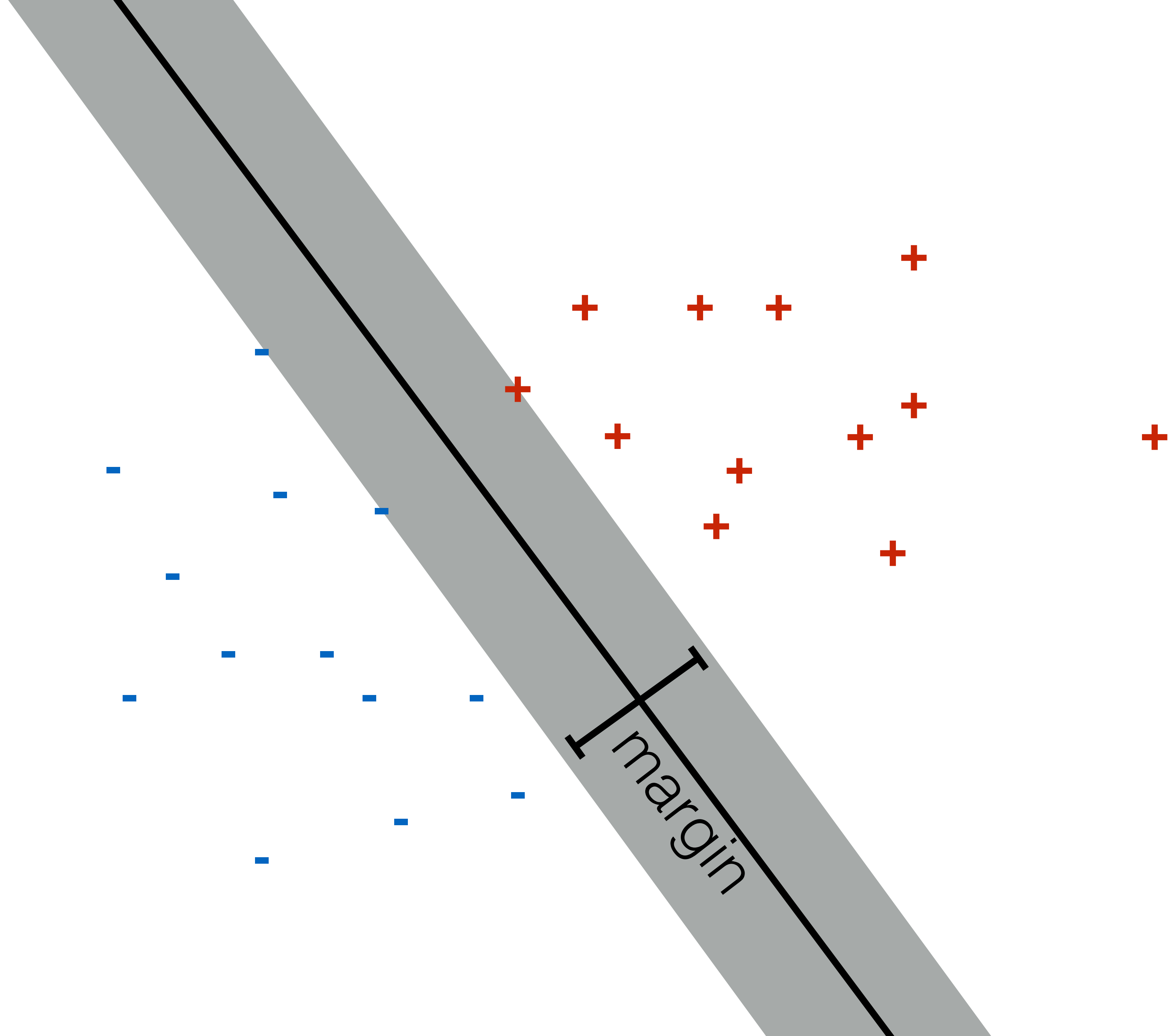
Machine Learning
CS4824/ECE4424
Bert Huang
Virginia Tech

Outline

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





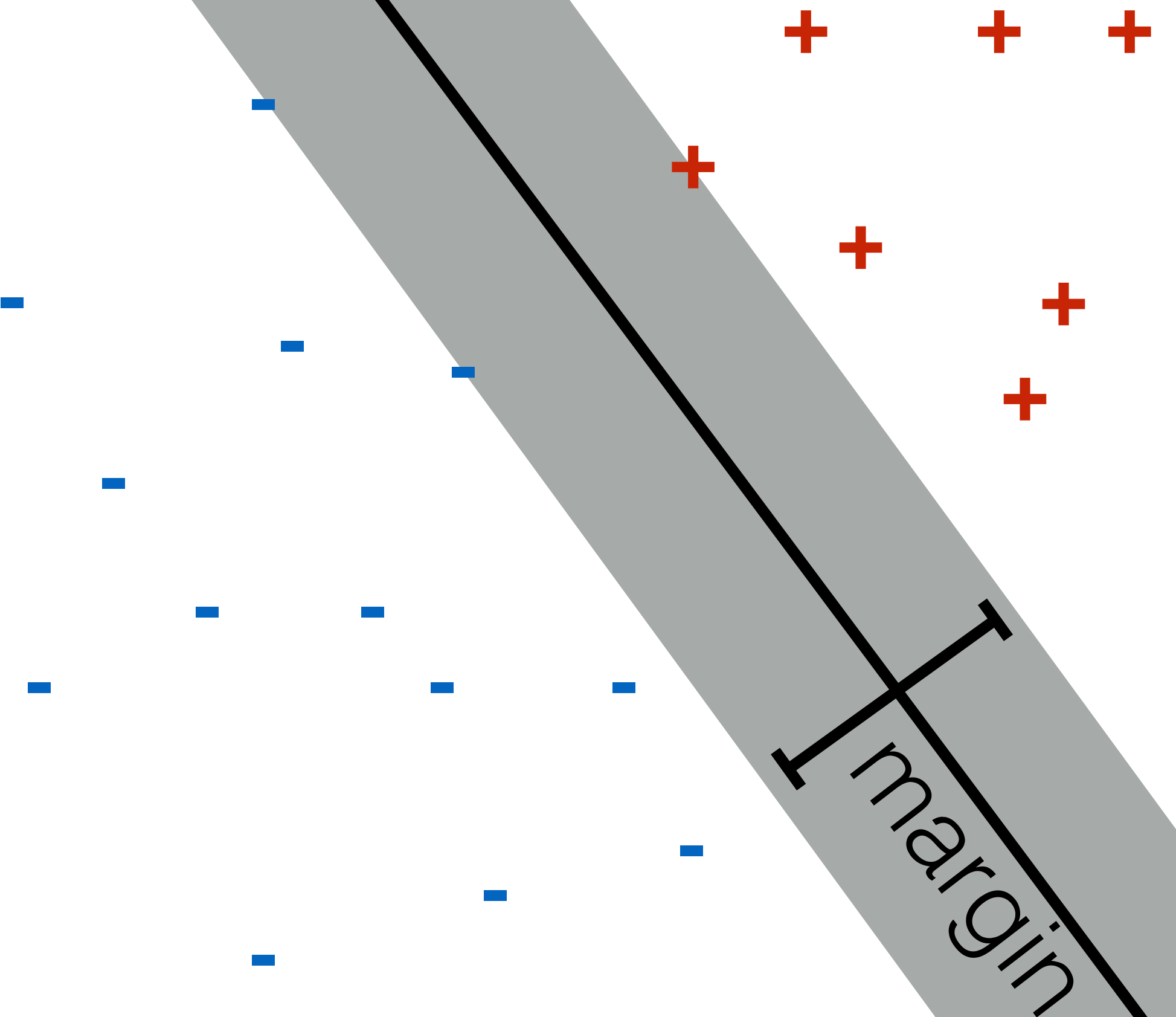


Quantifying the Margin

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

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

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



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

$$w^T x + b \geq 1$$

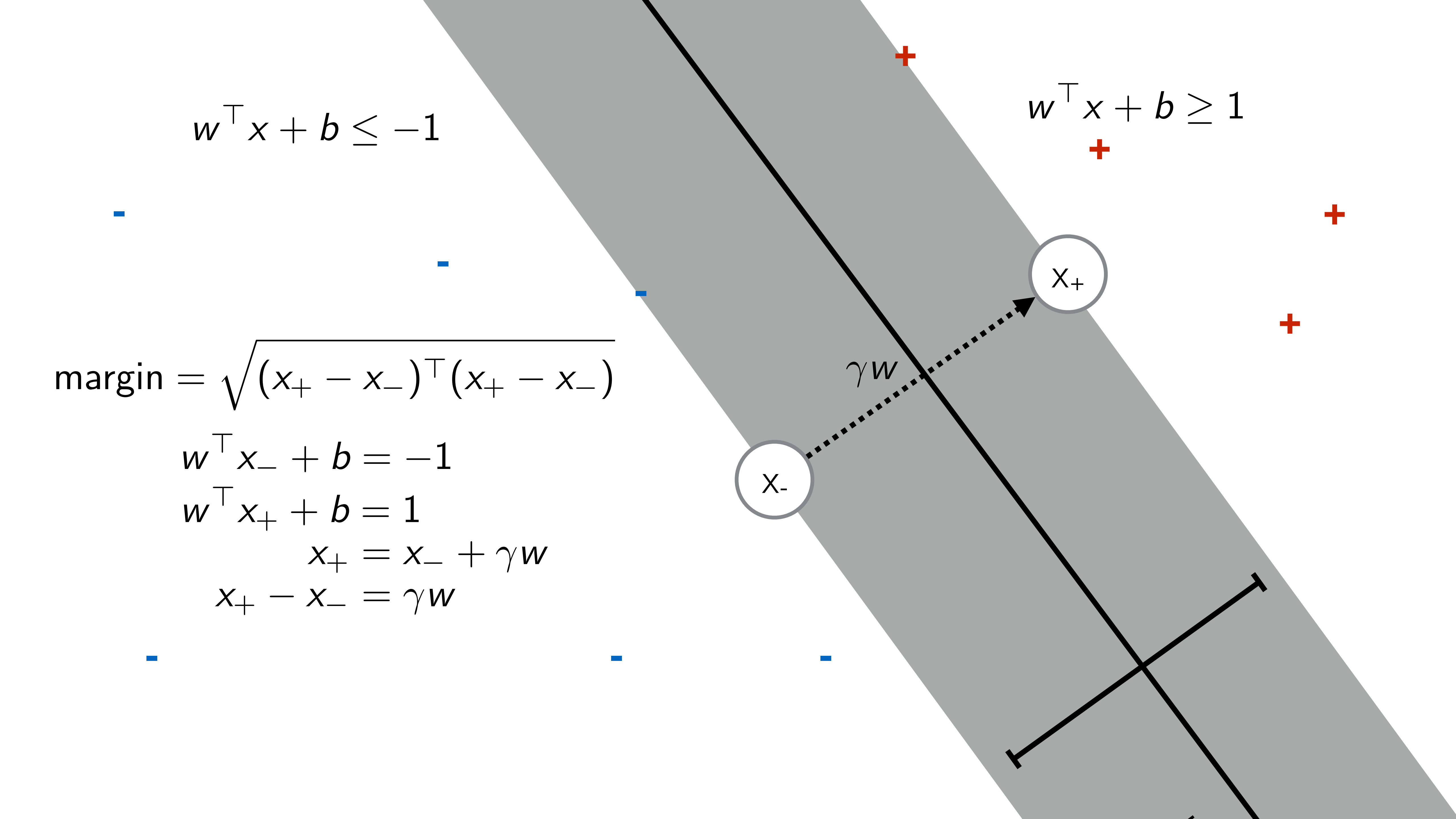
$$\text{margin} = \sqrt{(x_+ - x_-)^T (x_+ - x_-)}$$

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

$$w^T x_+ + b = 1$$

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

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



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

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

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

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

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

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

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

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

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

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

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

Large-Margin Linear Classification

$$\begin{aligned} \max_{w \in \mathbb{R}^d} \quad & \frac{2}{\sqrt{w^\top w}} \\ \text{s.t.} \quad & y_i(w^\top x_i + b) \geq 1 \quad \forall i \in \{1, \dots, n\} \end{aligned}$$

Large-Margin Linear Classification

$$\begin{aligned} \min_{w \in \mathbb{R}^d} \quad & \frac{1}{2} w^\top w \\ \text{s.t.} \quad & y_i(w^\top x_i + b) \geq 1 \quad \forall i \in \{1, \dots, n\} \end{aligned}$$

Quadratic Programming

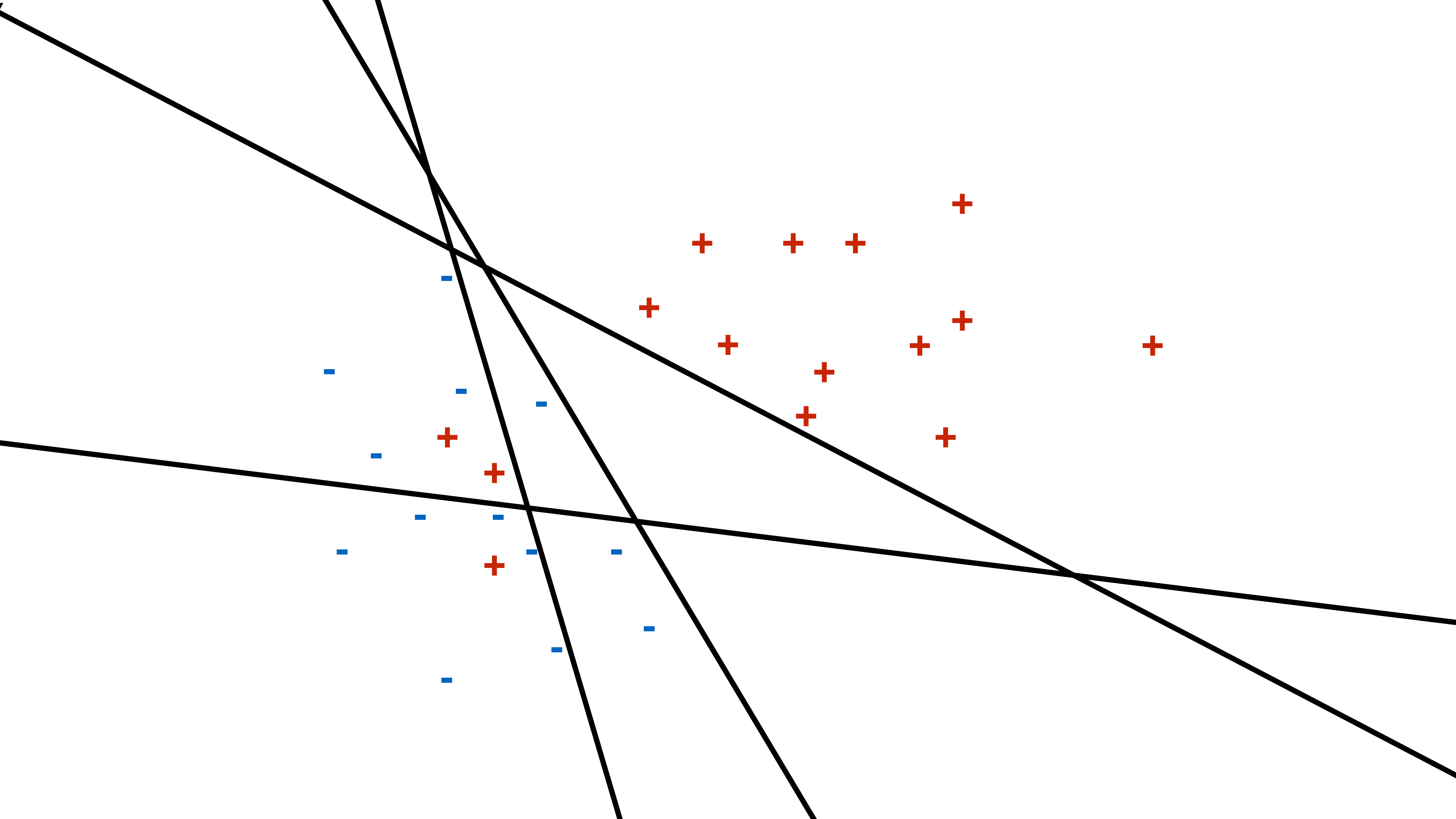
$$\min_x \quad \frac{1}{2} x^\top H x + f^\top x \quad \text{quadratic objective}$$

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

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

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

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



Soft-Margin Form

$$\begin{aligned} \min_{w \in \mathbb{R}^d} \quad & \frac{1}{2} w^\top w \\ \text{s.t.} \quad & y_i(w^\top x_i + b) \geq 1 \quad \forall i \in \{1, \dots, n\} \end{aligned}$$

Soft-Margin Form

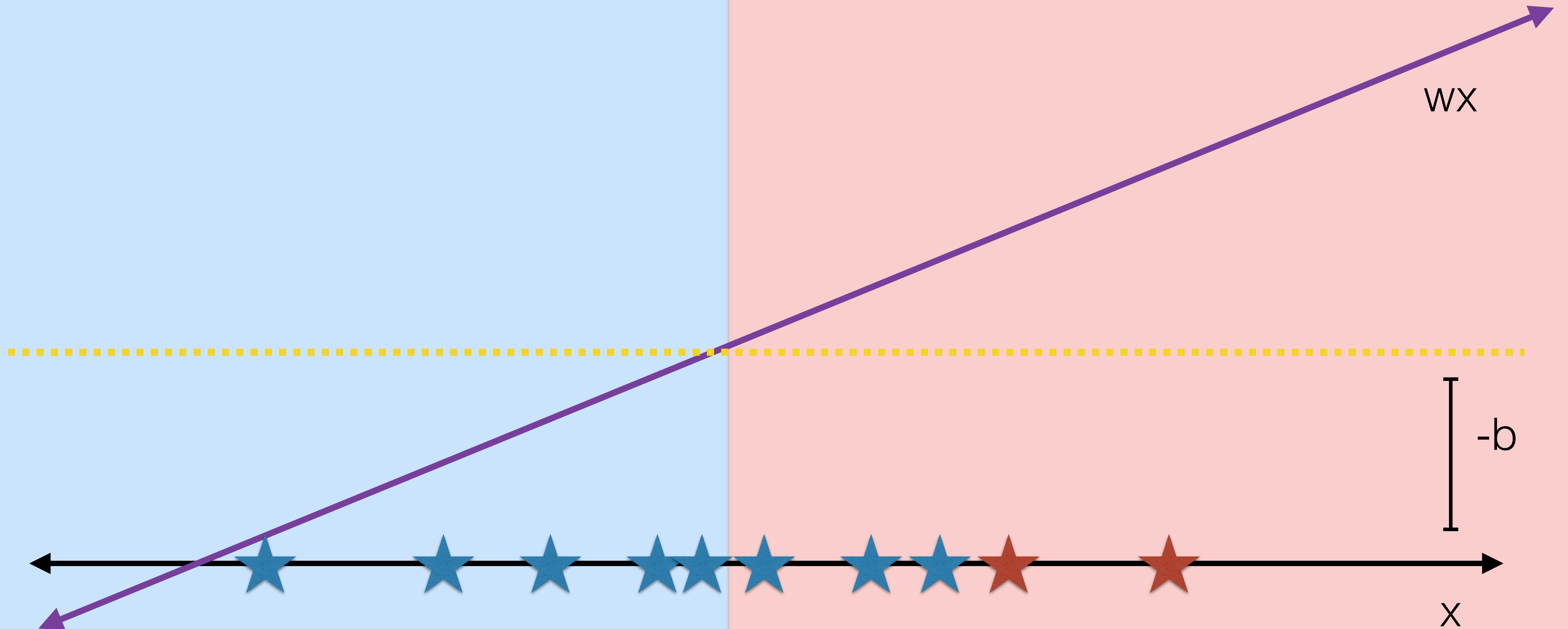
slack penalty

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

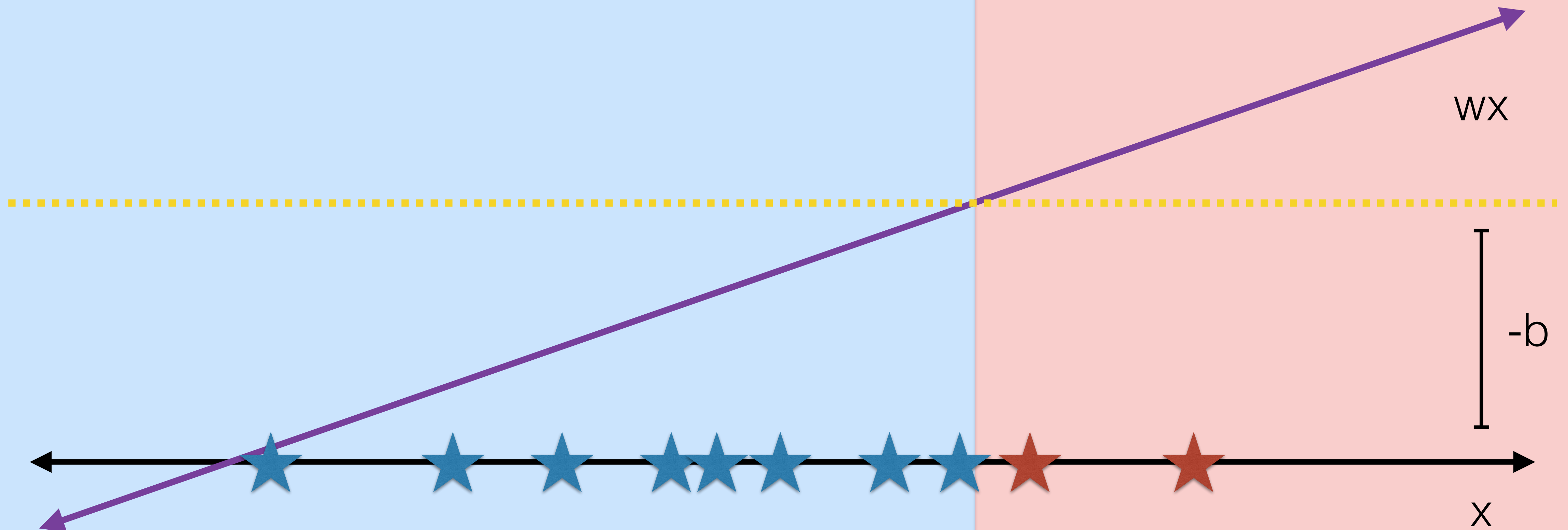
$$\text{s.t.} \quad y_i(w^\top x_i + b) \geq 1 - \xi_i \quad \forall i \in \{1, \dots, n\}$$

slack variables

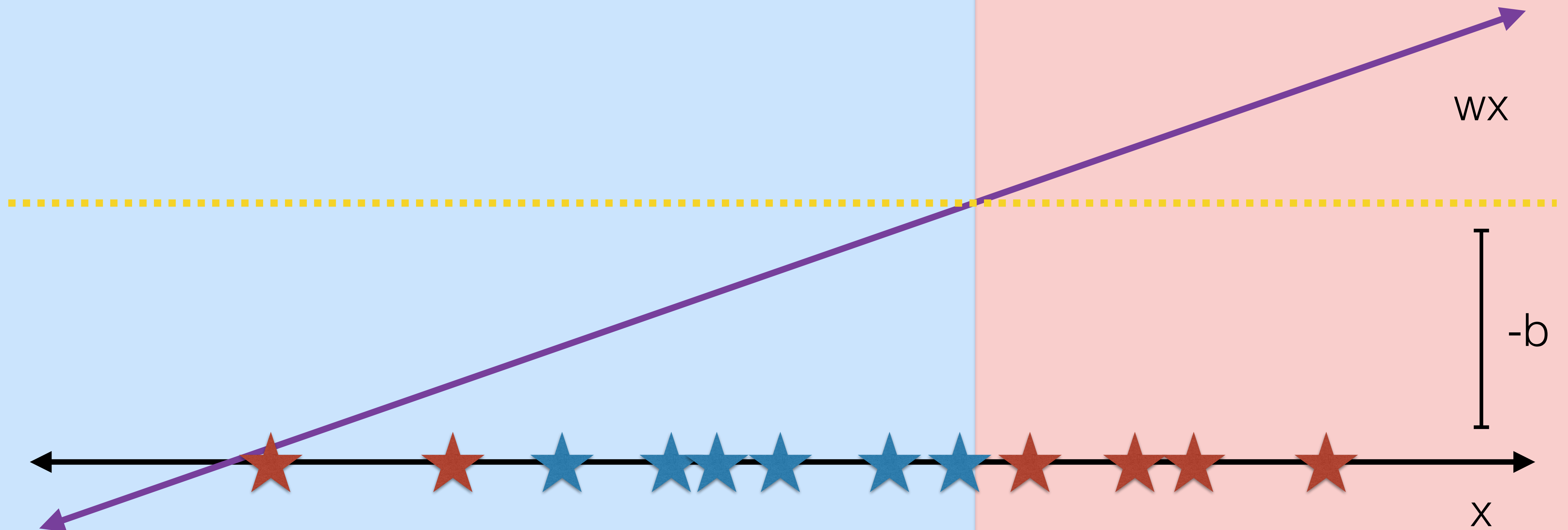
Nonlinear Decision Boundary



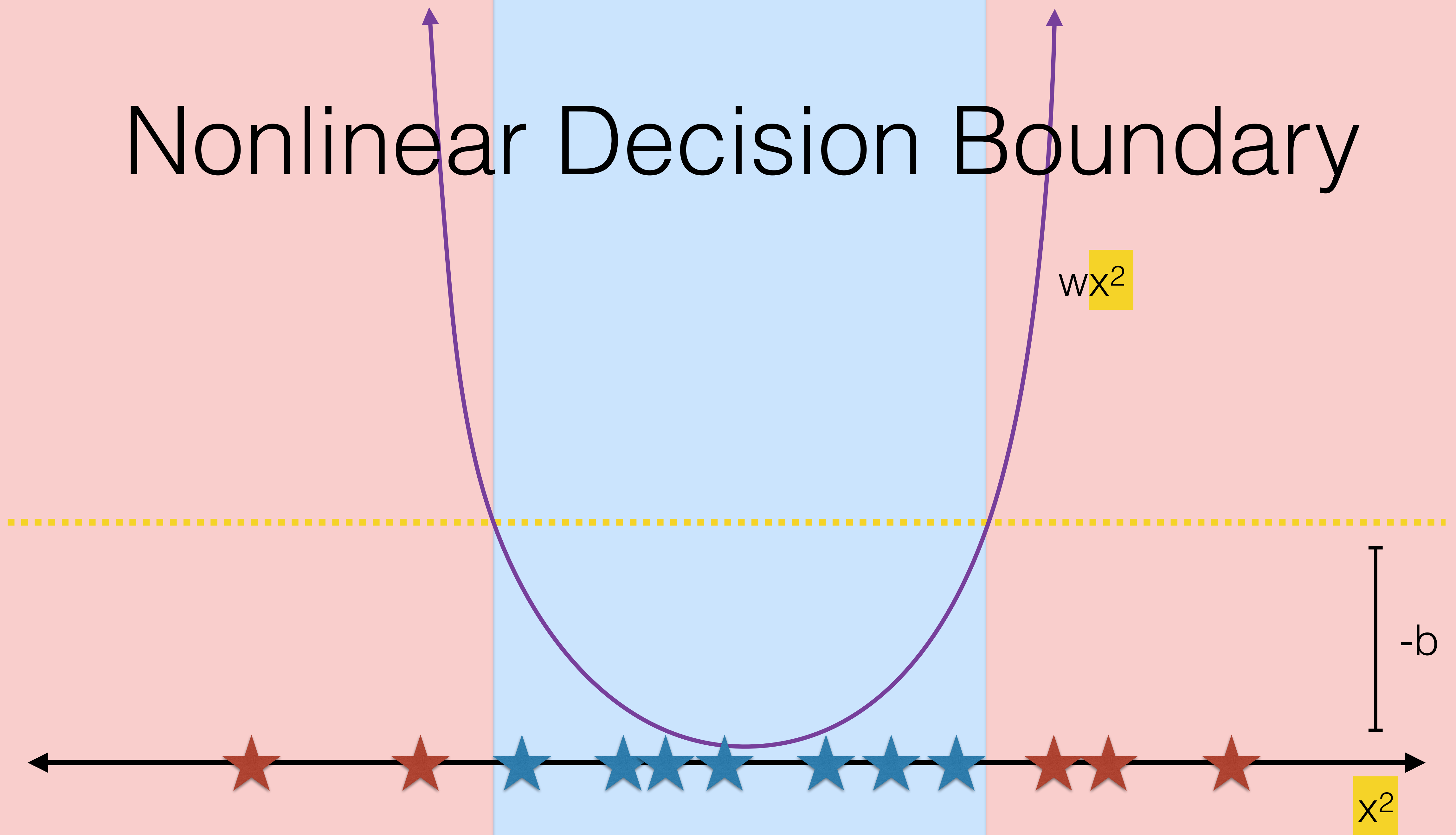
Nonlinear Decision Boundary



Nonlinear Decision Boundary



Nonlinear Decision Boundary



Polynomial Feature Map

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

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

Fourth-order terms $\{x^i x^j x^k x^\ell \mid i, j, k, \ell \in \{1, \dots, d\}\}$

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

Soft-Margin Form w/ Feature Map

$$\begin{aligned} & \text{slack penalty} \\ \min_{\substack{w \in \mathbb{R}^d \\ \xi \geq 0}} & \quad \frac{1}{2} w^\top w + C \sum_{i=1}^n \xi_i \\ \text{s.t.} & \quad y_i(w^\top \Phi(x_i) + b) \geq 1 - \xi_i \quad \forall i \in \{1, \dots, n\} \end{aligned}$$

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**