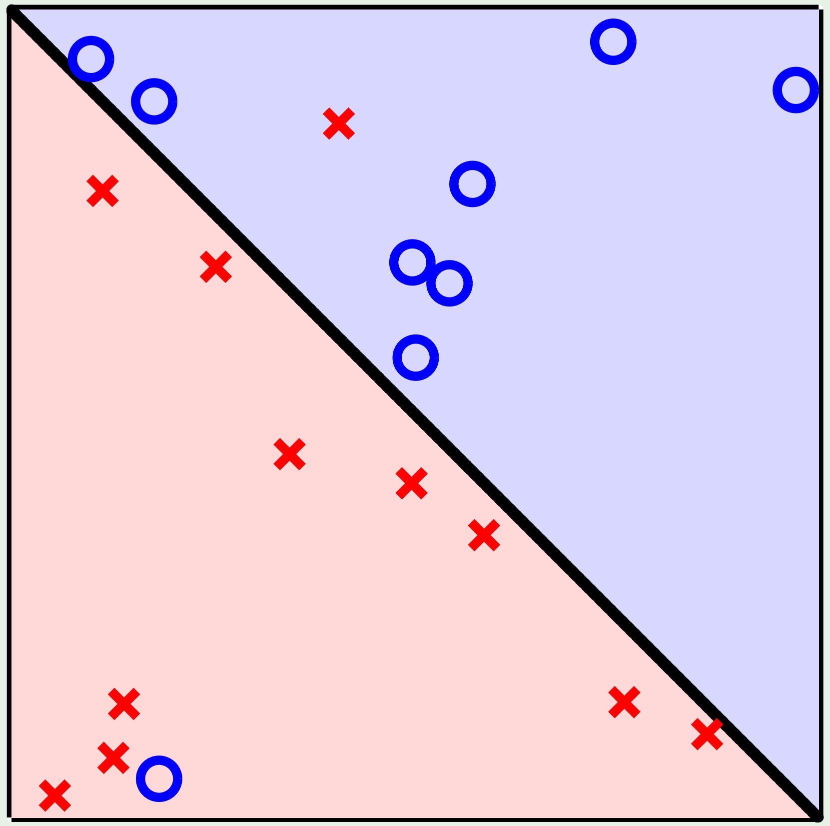
Outline

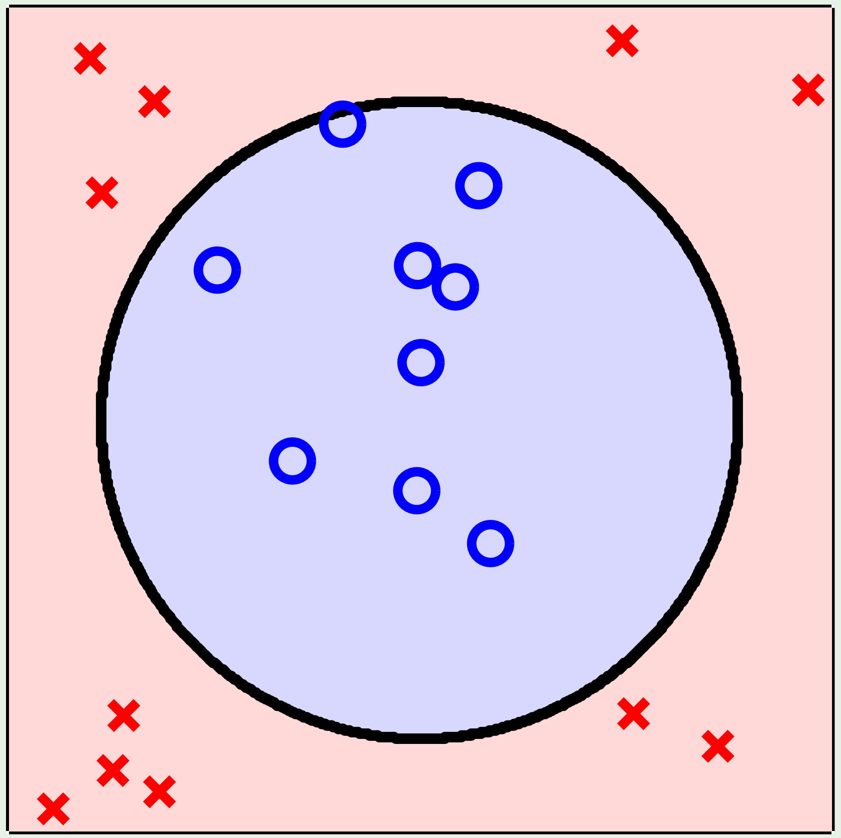
- The kernel trick
- Soft-margin SVM

Two types of non-separable

slightly:



seriously:

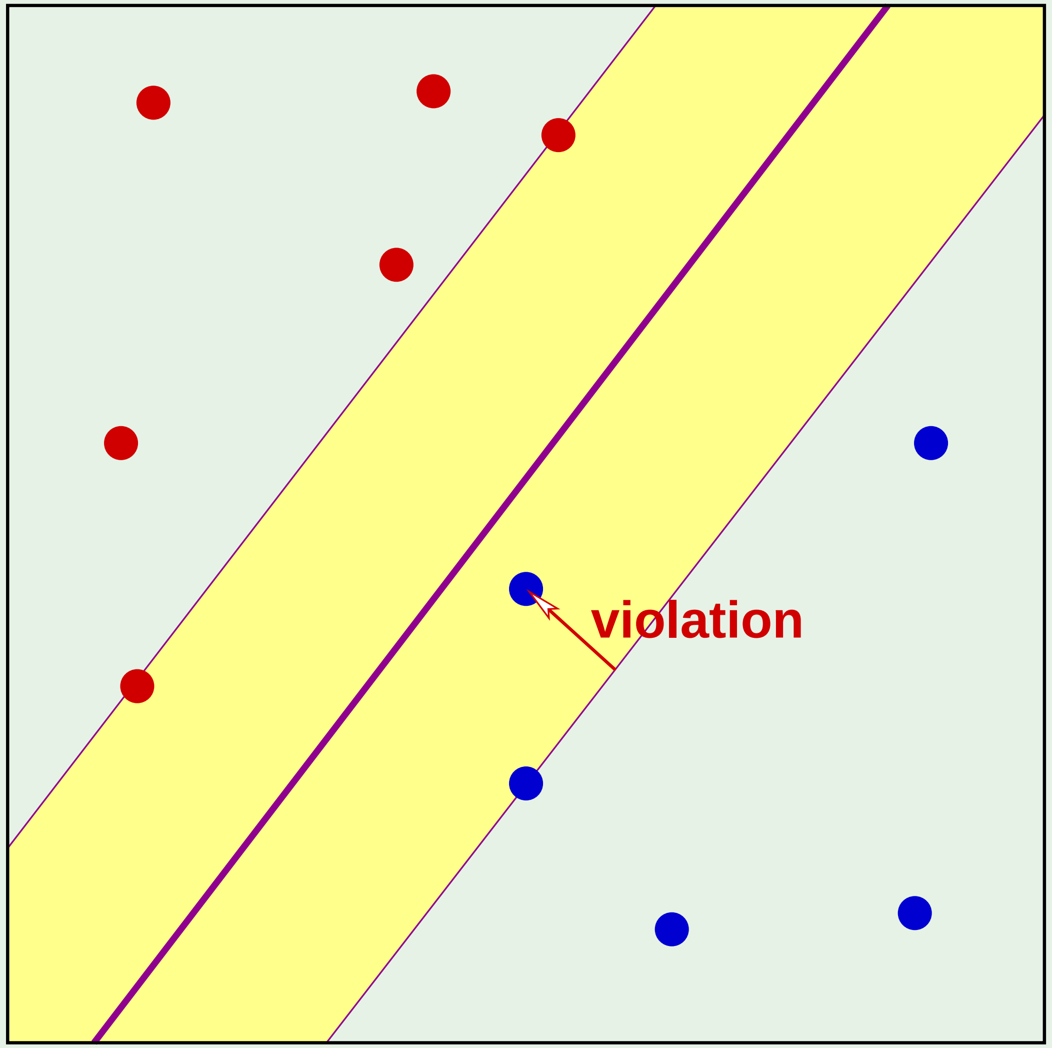


Error measure

Margin violation: $y_n\left(\mathbf{w}^\intercal\mathbf{x}_n + b\right) \geq 1$ fails

Quantify: $y_n\left(\mathbf{w}^\intercal\mathbf{x}_n + b\right) \geq 1 - \xi_n \qquad \xi_n \geq 0$

$$\text{Total violation} = \sum_{n=1}^{N} \xi_n$$



The new optimization

Minimize $\quad \frac{1}{2}\,\mathbf{w}^{\intercal}\mathbf{w} + C\sum_{n=1}^{N}\xi_n$

subject to $\quad y_n\left(\mathbf{w}^{\intercal}\mathbf{x}_n + b\right) \geq 1 - \xi_n \quad$ for $\quad n = 1, \ldots, N$

and $\quad \xi_n \geq 0 \quad$ for $\; n = 1, \ldots, N$

$$\mathbf{w} \in \mathbb{R}^d \;,\quad b \in \mathbb{R} \;,\quad \boldsymbol{\xi} \in \mathbb{R}^N$$

Lagrange formulation

$$\mathcal{L}(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2}\,\mathbf{w}^{\intercal}\mathbf{w} + C\sum_{n=1}^{N}\xi_n - \sum_{n=1}^{N}\alpha_n(y_n\,(\mathbf{w}^{\intercal}\mathbf{x}_n + b) - 1 + \xi_n) - \sum_{n=1}^{N}\beta_n\,\xi_n$$

Minimize w.r.t. $\mathbf{w}$, b, and $\boldsymbol{\xi}$ and maximize w.r.t. each $\alpha_n \geq 0$ and $\beta_n \geq 0$

$$\nabla_{\mathbf{w}}\mathcal{L} = \mathbf{w} - \sum_{n=1}^{N}\alpha_n y_n \mathbf{x}_n = \mathbf{0}$$

$$\frac{\partial \mathcal{L}}{\partial b} = -\sum_{n=1}^{N}\alpha_n y_n = 0$$

$$\frac{\partial \mathcal{L}}{\partial \xi_n} = C - \alpha_n - \beta_n = 0$$

and the solution is ...

Maximize $\mathcal{L}(\boldsymbol{\alpha}) = \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} y_n y_m \, \alpha_n \alpha_m \, \mathbf{x}_n^{\intercal} \mathbf{x}_m$ w.r.t. to $\boldsymbol{\alpha}$

subject to $0 \leq \alpha_n \leq C$ for $n = 1, \cdots, N$ and $\sum_{n=1}^{N} \alpha_n y_n = 0$

$$\Longrightarrow \quad \mathbf{w} = \sum_{n=1}^{N} \alpha_n y_n \mathbf{x}_n$$

minimizes $\frac{1}{2} \mathbf{w}^{\intercal} \mathbf{w} + C \sum_{n=1}^{N} \xi_n$

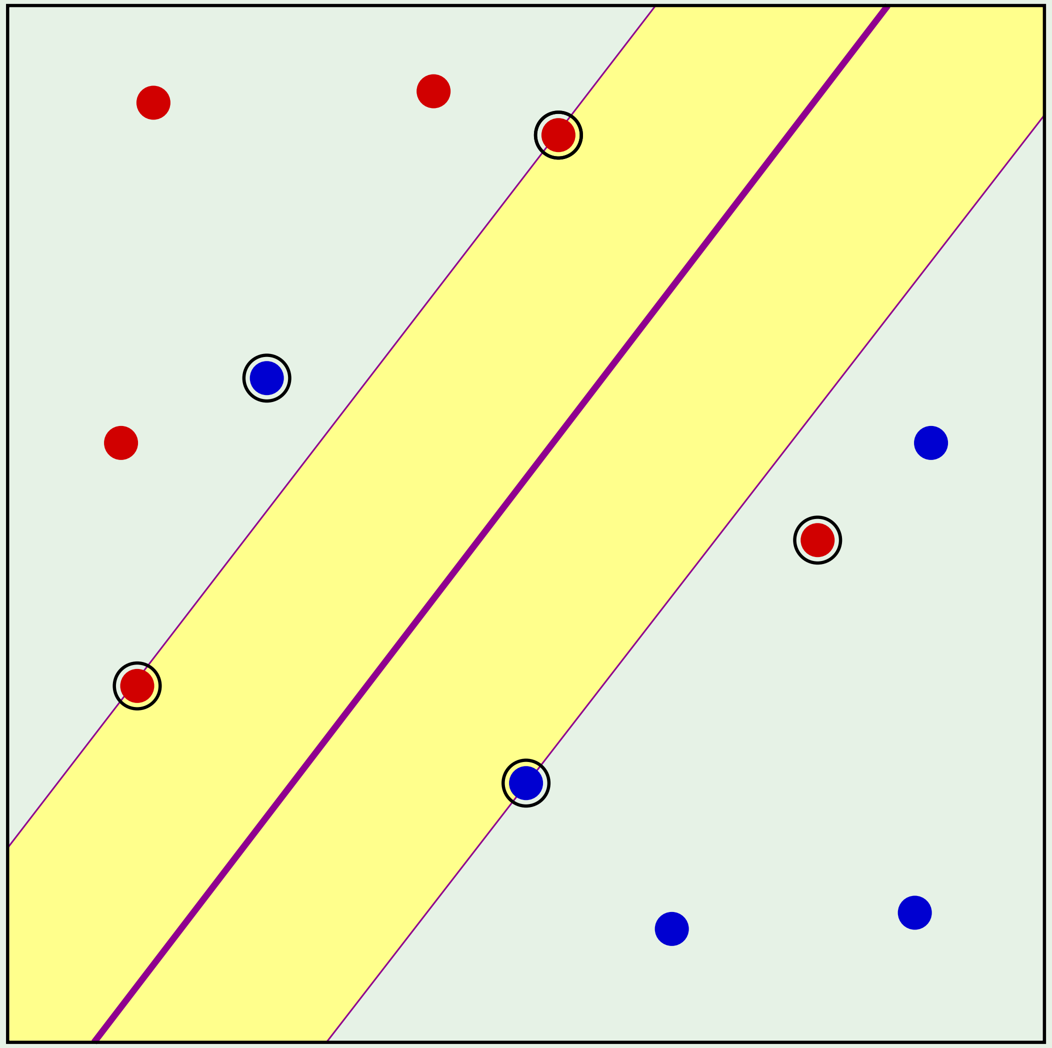
Types of support vectors

margin support vectors $(0 < \alpha_n < C)$

$$y_n\left(\mathbf{w}^{\intercal}\mathbf{x}_n + b\right) = 1 \qquad (\xi_n \;=\; 0)$$

non-margin support vectors $(\alpha_n = C)$

$$y_n\left(\mathbf{w}^{\intercal}\mathbf{x}_n + b\right) < 1 \qquad (\xi_n \;>\; 0)$$



Two technical observations

1. Hard margin: What if data is not linearly separable?

   "primal $\longrightarrow$ dual" breaks down

2. $\mathcal{Z}$: What if there is w_0?

   All goes to b and $w_0 \to 0$