Kernels, SVM

Kenny Marino, Colin White and
Nupur Chatterji

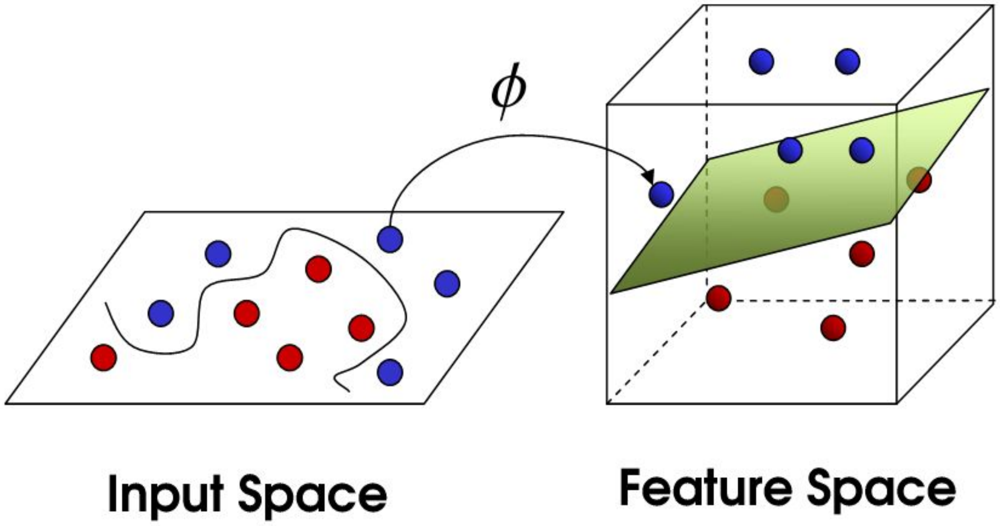
Kernel Overview

- What if data is not linearly separable?
  - Switch to a more complicated class of functions
  - Use a Kernel!

- Kernels are a "legal definition" of a dot product: there exists **Φ** such that K(x,y)=**Φ**(x) **Φ**(y)

Key Ideas

- By using the kernel we move into a higher dimension space
- Imperative to think of the Kernel Function **IMPLICITLY**



Why think about Φ implicitly?

- Feature space can grow rapidly
- Avoid computing actual values of coordinates in the feature space
  - “KERNEL TRICK”
  - Just take inner products
- Computationally cheaper than explicitly calculating values

Common Kernels and Commonly Kernelizable Algos

Linear: $K(x, z) = x \cdot z$

Polynomial: $K(x, z) = (x \cdot z)^d$ or $K(x, z) = (1 + x \cdot z)^d$

Gaussian: $K(x, z) = \exp\left[-\frac{||x-z||^2}{2\,\sigma^2}\right]$

Laplace Kernel: $K(x, z) = \exp\left[-\frac{||x-z||}{2\,\sigma^2}\right]$

- Perceptron
- SVM
- Linear Regression
- Ridge Regression
- K-Means

Margins

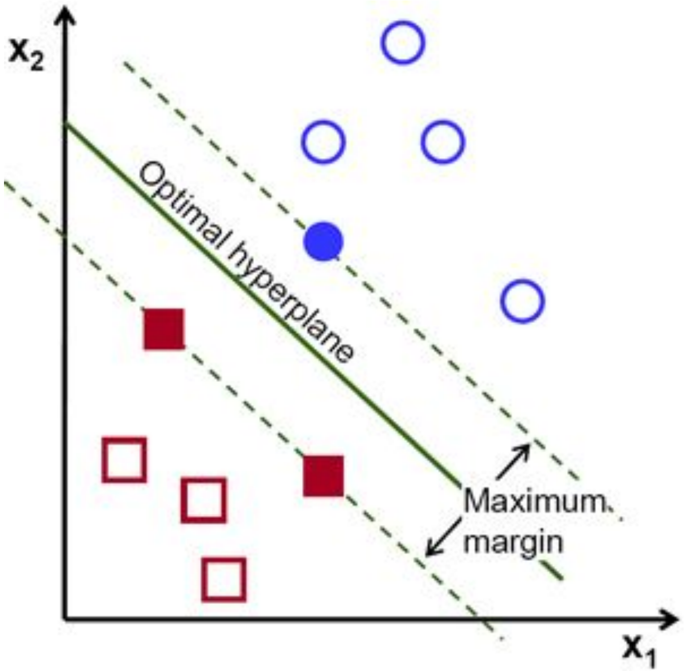
- Having a large margin will prevent overfitting
- Why not directly search for a large margin classifier?

SVM

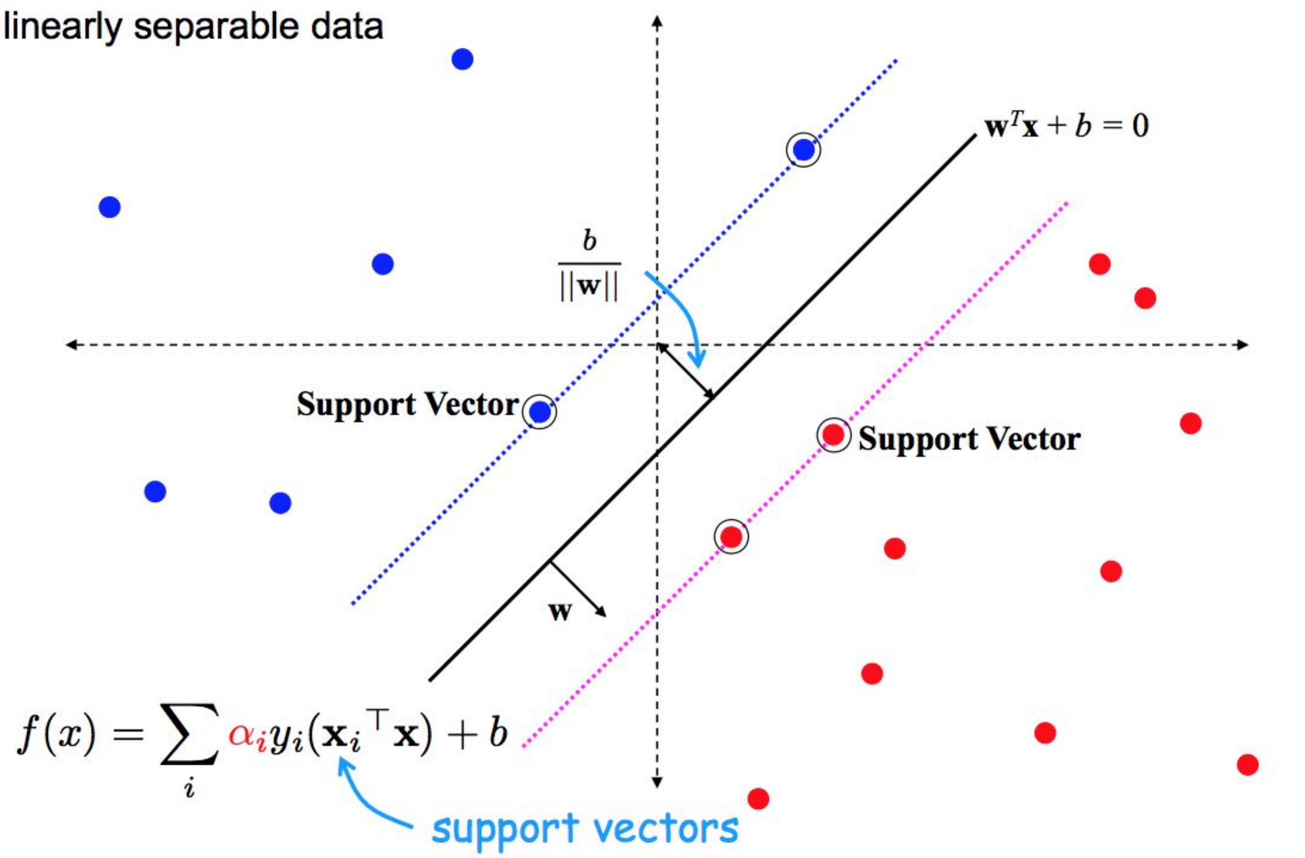
- Optimize for the maximum margin separator
- Most stable under noise

Idea

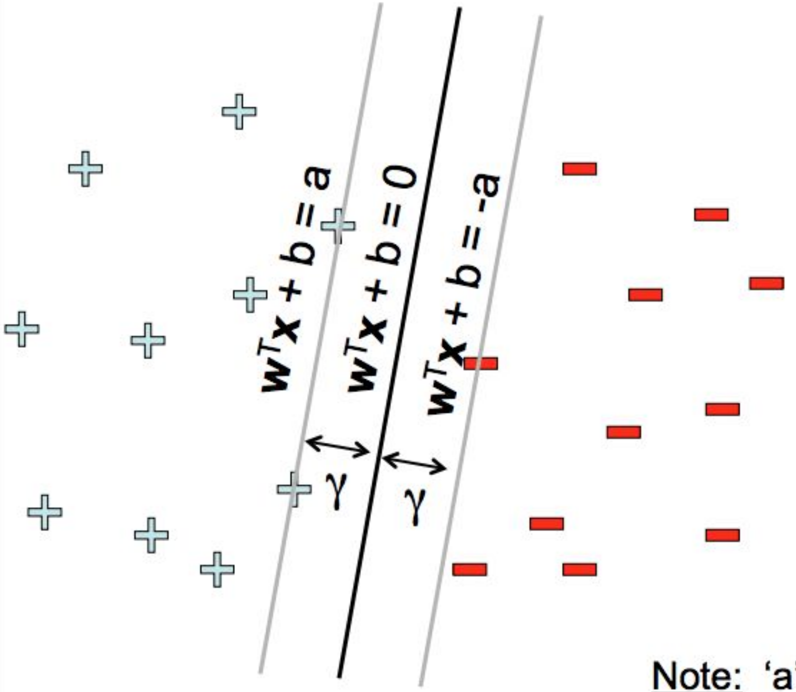
- Support Vector Machines attempt to directly learn the linear separator with the largest margin



Support Vector Machine



Maximizing the margin



Margin = Distance of closest examples from the decision line/ hyperplane

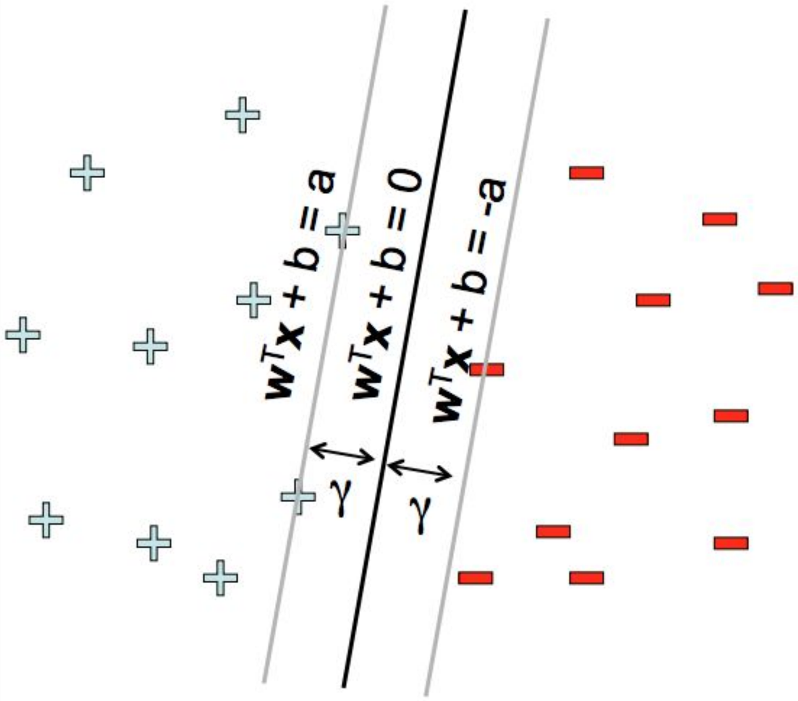
margin = γ = $a/\|w\|$

$$\max_{\mathbf{w},b} \ \gamma = a/\|w\|$$

$$\text{s.t. } (\mathbf{w}^T\mathbf{x}_j + b)\, y_j \geq a \quad \forall j$$

Note: 'a' is arbitrary (can normalize equations by a)

Support Vector Machine



$$\min_{\mathbf{w},b} \; \mathbf{w}^T\mathbf{w}$$

$$\text{s.t. } (\mathbf{w}^T\mathbf{x}_j + b)\, y_j \geq 1 \quad \forall j$$

Solve efficiently by quadratic programming (QP)

- Well-studied solution algorithms

Linear hyperplane defined by "<u>support vectors</u>"

Any Questions?

Slides taken from Tom Mitchell, Nina Balcan, Oxford University