Kernel methods III
Kernel SVM

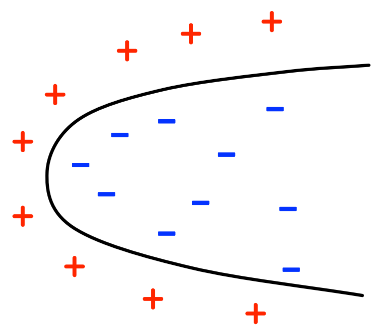
Sanjoy Dasgupta

University of California, San Diego

Topics we'll cover

1. Kernel SVM
2. Polynomial decision boundaries

Step 1: basis expansion



Actual boundary is something like $x_1 = x_2^2 + 5$.

- This is quadratic in $x = (x_1, x_2)$
- But it is linear in $\Phi(x) = (x_1, x_2, x_1^2, x_2^2, x_1 x_2)$

Basis expansion: embed data in a higher-dimensional feature space.
Then we can use a linear classifier!

Perceptron with basis expansion

Learning from data $(x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)}) \in \mathcal{X} \times \{-1, 1\}$

Primal form of the Perceptron:

- $w = 0$ and $b = 0$
- while there is some i with $y^{(i)}(w \cdot \Phi(x^{(i)}) + b) \leq 0$:
  - $w = w + y^{(i)}\,\Phi(x^{(i)})$
  - $b = b + y^{(i)}$

Problem: w and $\Phi(x)$ can be very high-dimensional.

Solution: work in the dual space, writing

$$w = \sum_j \alpha_j y^{(j)} \Phi(x^{(j)})$$

Step 2: the kernel trick

Dual form of the Perceptron:

- $\alpha = 0$ and $b = 0$
- while some i has $y^{(i)}\left(\sum_j \alpha_j y^{(j)} \Phi(x^{(j)}) \cdot \Phi(x^{(i)}) + b\right) \leq 0$:
  - $\alpha_i = \alpha_i + 1$
  - $b = b + y^{(i)}$

Classify a new point x: sign $\left(\sum_j \alpha_j y^{(j)} \Phi(x^{(j)}) \cdot \Phi(x) + b\right)$.

Does this work with SVMs?

$$\boxed{\begin{array}{ccc} \text{(PRIMAL)} & \displaystyle\min_{w\in\mathbb{R}^d, b\in\mathbb{R}, \xi\in\mathbb{R}^n} & \displaystyle\|w\|^2 + C\sum_{i=1}^n \xi_i \\ \text{s.t.:} & y^{(i)}(w \cdot x^{(i)} + b) \geq 1 - \xi_i & \text{for all } i = 1, 2, \ldots, n \\ & \xi \geq 0 & \end{array}}$$

$$\boxed{\begin{array}{c} \text{(DUAL)} \quad \displaystyle\max_{\alpha\in\mathbb{R}^n} \sum_{i=1}^n \alpha_i - \sum_{i,j=1}^n \alpha_i\alpha_j y^{(i)} y^{(j)} (x^{(i)} \cdot x^{(j)}) \\ \text{s.t.:} \quad \displaystyle\sum_{i=1}^n \alpha_i y^{(i)} = 0 \\ 0 \leq \alpha_i \leq C \end{array}}$$

Solution: $w = \sum_i \alpha_i y^{(i)} x^{(i)}$.

Kernel SVM

1. **Basis expansion.** Mapping $x \mapsto \Phi(x)$.
2. **Learning.** Solve the dual problem:

$$\boxed{\begin{array}{c} \max_{\alpha \in \mathbb{R}^n} \sum_{i=1}^{n} \alpha_i - \sum_{i,j=1}^{n} \alpha_i \alpha_j y^{(i)} y^{(j)} (\Phi(x^{(i)}) \cdot \Phi(x^{(j)})) \\ \text{s.t.: } \sum_{i=1}^{n} \alpha_i y^{(i)} = 0 \\ 0 \leq \alpha_i \leq C \end{array}}$$

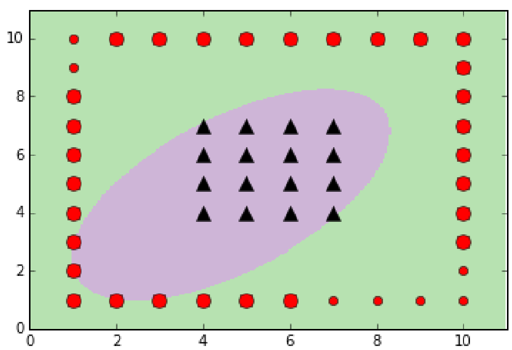
   This yields $w = \sum_i \alpha_i y^{(i)} \Phi(x^{(i)})$. Offset b also follows.

3. **Classification.** Given a new point x, classify as

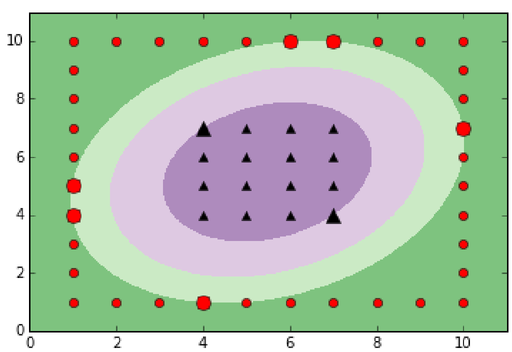
$$\text{sign}\left(\sum_i \alpha_i y^{(i)} (\Phi(x^{(i)}) \cdot \Phi(x)) + b \right).$$

Kernel Perceptron vs. Kernel SVM: examples

Perceptron:

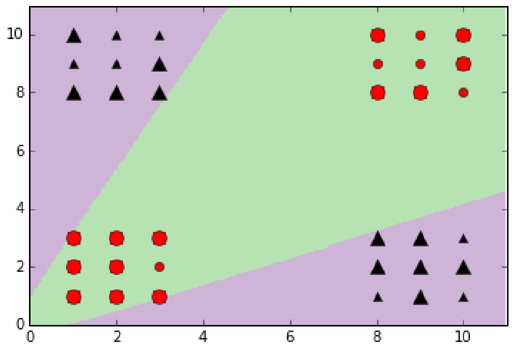


SVM:

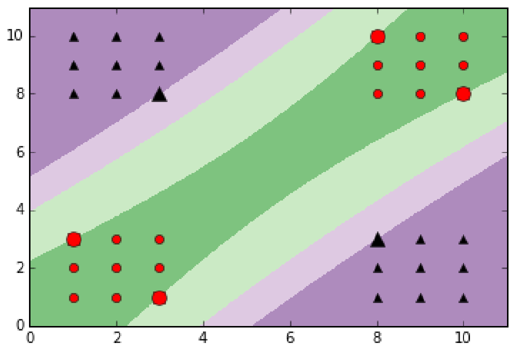


Kernel Perceptron vs. Kernel SVM: examples

Perceptron:

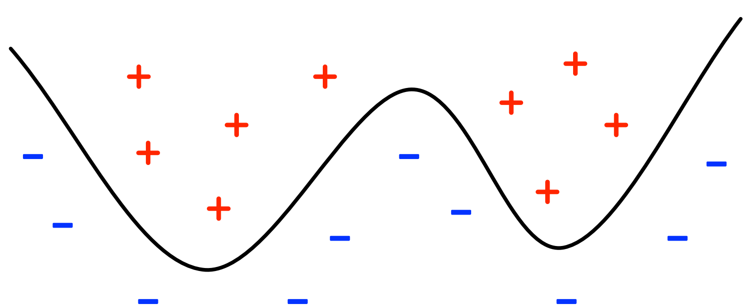


SVM:



Polynomial decision boundaries

When the decision surface is a polynomial of order p:



- Let $\Phi(x)$ consist of all terms of order $\leq p$, such as $x_1 x_2^2 x_3^{p-3}$.
  (How many such terms are there, roughly?)
- Same trick works: $\Phi(x) \cdot \Phi(z) = (1 + x \cdot z)^p$.