Support Vector Machines

Observation

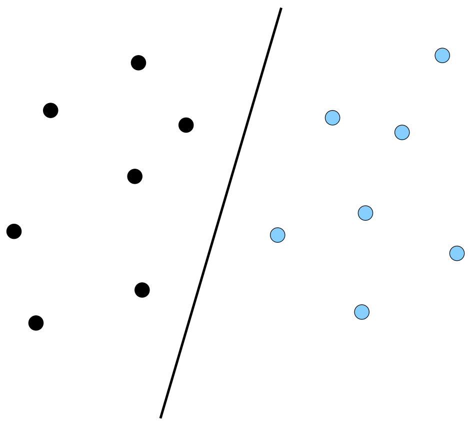
Almost everything becomes linearly separable when represented in high-dimensional spaces

"Ordinary" low-dimensional data can be "scattered" into a high-dimensional space.

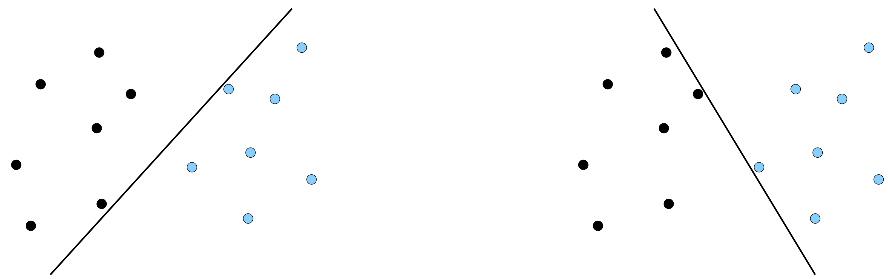
Two problems emerge

1. Many free parameters → bad generalization
2. Extensive computations

Linear Separation

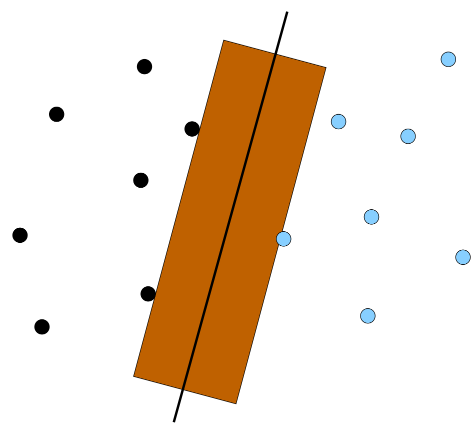


Many acceptable solutions → bad generalization



- Structural Risk

Hyperplane with margins



Less arbitrariness → better generalization

- Wide margins restrict the possible weights to choose from
- Less risk to choose bad weights by accident
- Reduced risk for bad generalization

Minimization of the structural risk ≡ maximization of the margin

> Out of all hyperplanes which solve the problem
> the one with widest margin will generalize best

Mathematical Formulation

- Separating Hyperplane

$$\vec{w}^T\vec{x} + b = 0$$

- Hyperplane with a margin

$$\vec{w}^T\vec{x} + b \geq 1 \qquad \text{when } t = 1$$
$$\vec{w}^T\vec{x} + b \leq -1 \qquad \text{when } t = -1$$

- Combined

$$t(\vec{w}^T\vec{x} + b) \geq 1$$

How wide is the margin?

1. Select two points, $\vec{p}$ and $\vec{q}$, on the two margins:

$$\vec{w}^T\vec{p} + b = 1 \qquad \vec{w}^T\vec{q} + b = -1$$

2. Distance between $\vec{p}$ and $\vec{q}$ *along* $\vec{w}$:

$$d = \vec{w}^T(\vec{p} - \vec{q})\frac{1}{||\vec{w}||}$$

3. Simplify:

$$d = \frac{\vec{w}^T\vec{p} - \vec{w}^T\vec{q}}{||\vec{w}||} = \frac{(1-b) - (-1-b)}{||\vec{w}||} = \frac{2}{||\vec{w}||}$$

> Maximal margin corresponds to minimal length of the weight vector

Best Separating Hyperplane

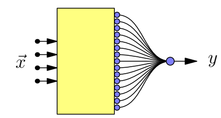
Minimize

$$\vec{w}^T\vec{w}$$

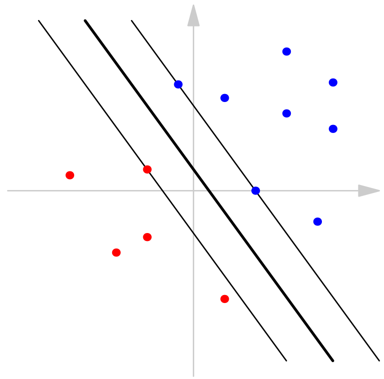
Constraints

$$t_i(\vec{w}^T\vec{x}_i + b) \geq 1 \qquad \forall i$$

Support Vector Machines



- Transform the input to a suitable high-dimensional space
- Choose the separation that has maximal margins



Support Vector Machines

- Advantages
  - Very good generalization
  - Works well even with few training samples
  - Fast classification
- Disadvantages
  - Non-local weight calculation
  - Hard to implement efficiently