Support vector machine

Ngô Minh Nhựt

Bộ môn Công nghệ Tri thức

2023

Large margin classifier

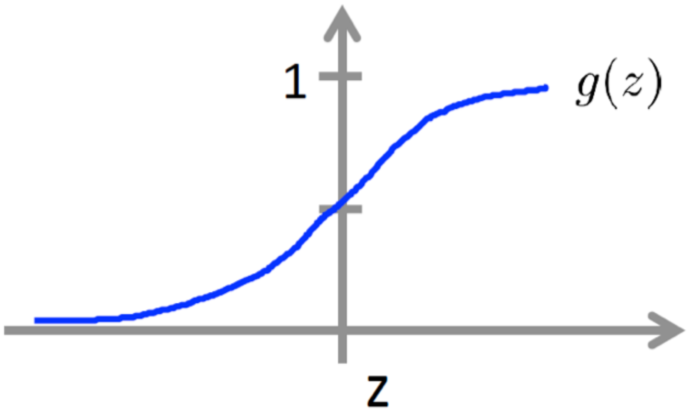
Phần 1

Hồi qui logistic

- Hypothesis:
  - $h_\theta(x) = g(\theta^T x)$
  - $h_\theta(x) = P(y=1|x;\theta)$

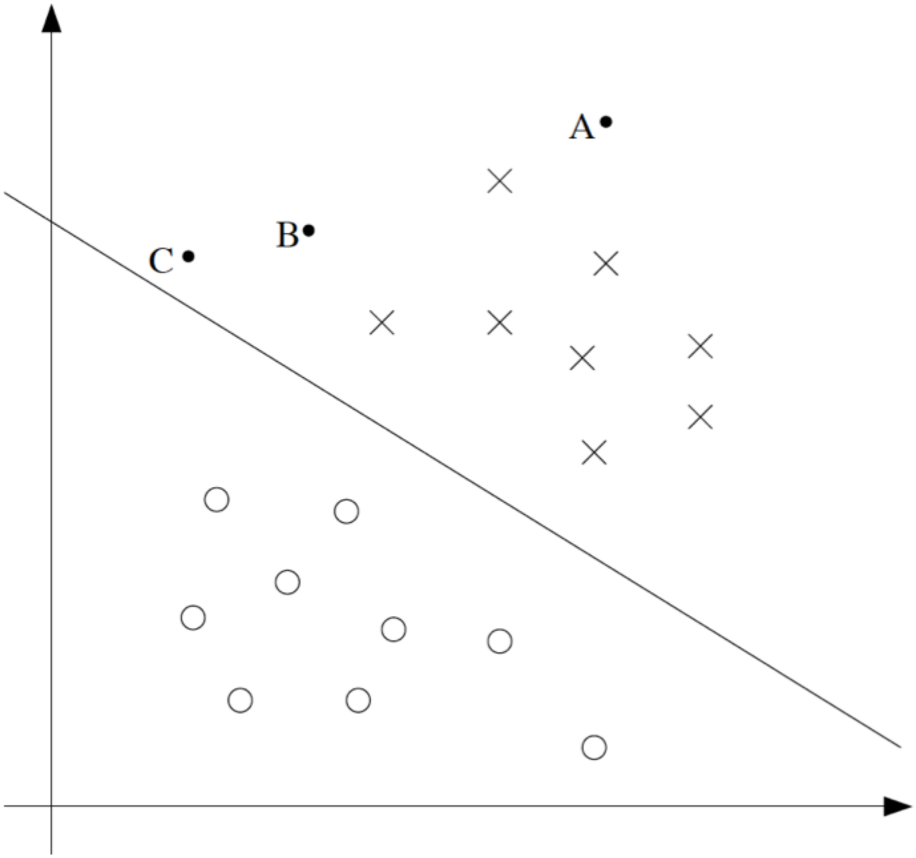
$$g(z) = \frac{1}{1+e^{-z}}$$

- $\theta^T x \gg 0$
  - Khả năng y = 1 càng cao
- $\theta^T x \ll 0$
  - Khả năng y = 0 càng cao



Hồi qui logistic

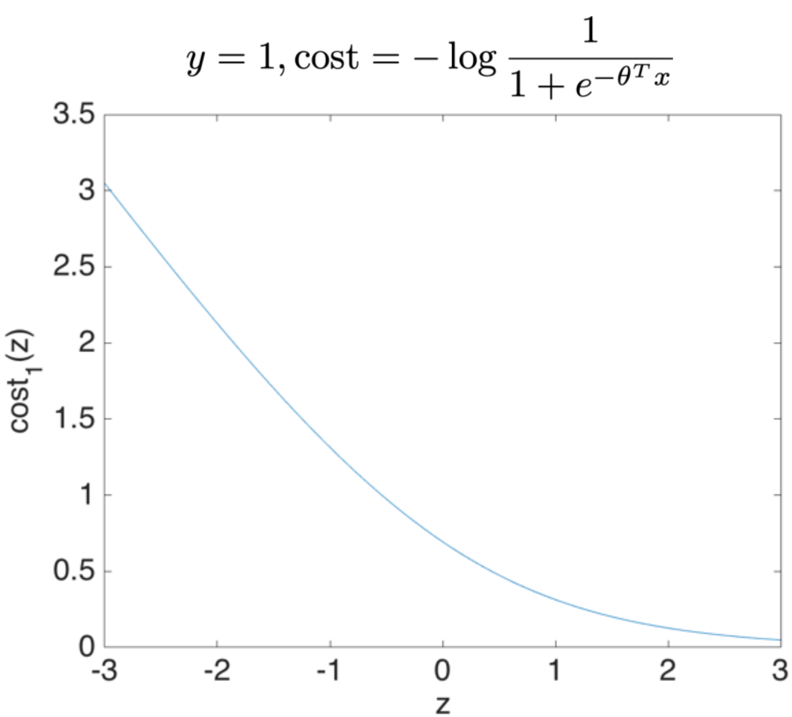
- Margin

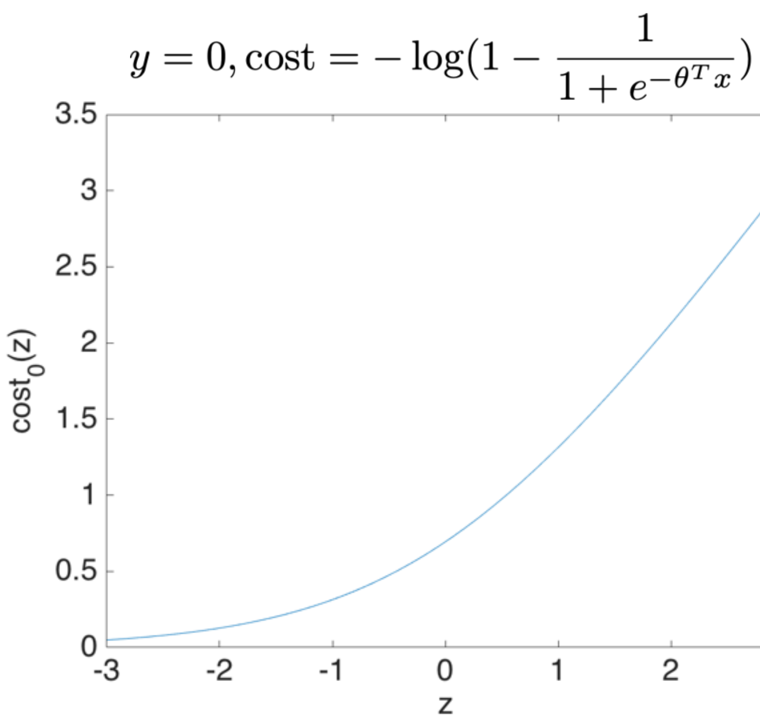


Hồi qui logistic

- Hàm chi phí cho một mẫu

$$-y\log\frac{1}{1+e^{-\theta^Tx}}-(1-y)\log(1-\frac{1}{1+e^{-\theta^Tx}})$$

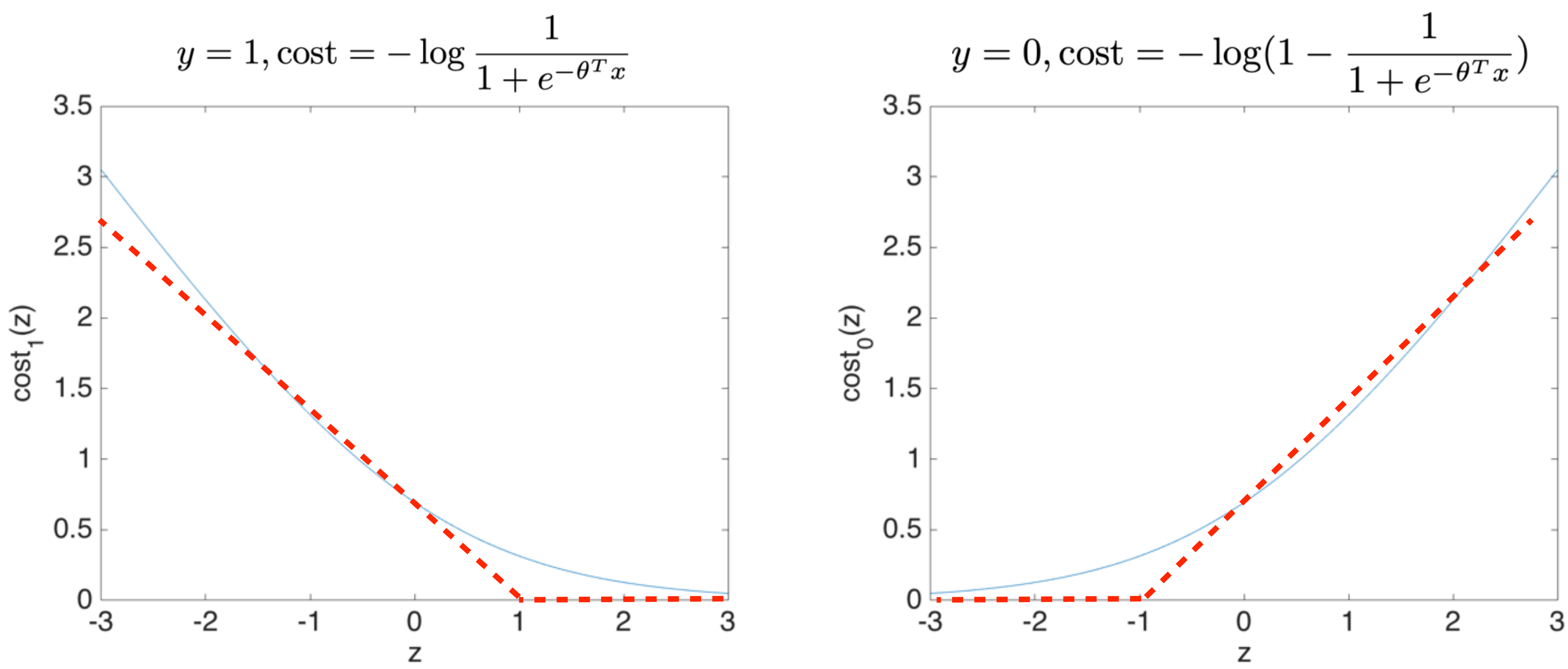




Hồi qui logistic

- Hàm chi phí cho một mẫu

$$-y \log \frac{1}{1+e^{-\theta^T x}} - (1-y)\log(1 - \frac{1}{1+e^{-\theta^T x}})$$



Support vector machine

- Hàm chi phí

$$J(\theta) = C\sum_{i=1}^{m}\left[y^{(i)}\text{cost}_1(\theta^T x^{(i)}) + (1-y^{(i)})\text{cost}_0(\theta^T x^{(i)})\right] + \frac{1}{2}\sum_{j=1}^{n}\theta_j^2$$

- Hồi qui logistic

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}(-\log h_\theta(x^{(i)})) + (1-y^{(i)})(-\log(1-h_\theta(x^{(i)})))\right] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$

Mô hình SVM

- Hypothesis
  - $y = 1 \quad \text{if} \quad \theta^T x \geq 1$
  - $y = 0 \quad \text{if} \quad \theta^T x \leq -1$
- Hàm chi phí

$$J(\theta) = C \sum_{i=1}^{m} \left[y^{(i)} \text{cost}_1(\theta^T x^{(i)}) + (1 - y^{(i)}) \text{cost}_0(\theta^T x^{(i)}) \right] + \frac{1}{2} \sum_{j=1}^{n} \theta_j^2$$

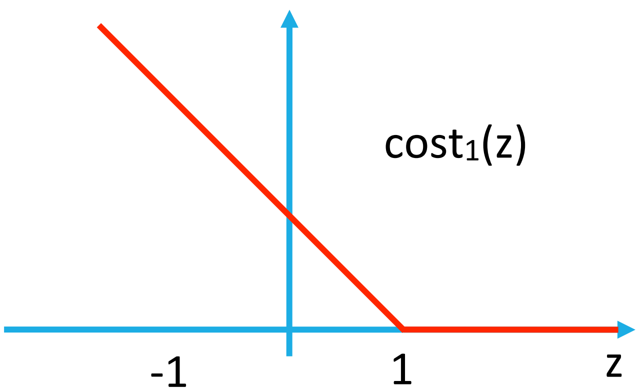
- Quá trình học:

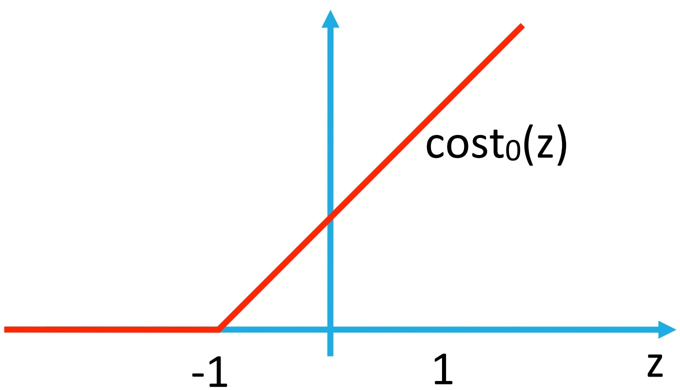
$$\min_{\theta} J(\theta)$$

Quá trình học

- Huấn luyện:

$$\min_{\theta} C\sum_{i=1}^{m}\left[y^{(i)}\text{cost}_1(\theta^T x^{(i)}) + (1-y^{(i)})\text{cost}_0(\theta^T x^{(i)})\right] + \frac{1}{2}\sum_{j=1}^{n}\theta_j^2$$





- Nếu y = 1, chúng ta cần $\theta^T x \geq 1$
- Nếu y = 0, chúng ta cần $\theta^T x \leq -1$

Quá trình học

- Mục tiêu

$$\min_{\theta} C \sum_{i=1}^{m} \left[y^{(i)} \text{cost}_1(\theta^T x^{(i)}) + (1 - y^{(i)}) \text{cost}_0(\theta^T x^{(i)}) \right] + \frac{1}{2} \sum_{j=1}^{n} \theta_j^2$$

  - Nếu y = 1, chúng ta cần $\theta^T x \geq 1$
  - Nếu y = 0, chúng ta cần $\theta^T x \leq -1$

➡ $$\min_{\theta} \frac{1}{2} \sum_{j=1}^{n} \theta_j^2$$

$$\begin{aligned} \text{s.t.} \quad & \theta^T x^{(i)} \geq 1 \quad \text{if} \quad y^{(i)} = 1 \\ & \theta^T x^{(i)} \leq -1 \quad \text{if} \quad y^{(i)} = 0 \end{aligned}$$

Vì sao large margin

- Tích vô hướng

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \qquad v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$u^T v = u_1 v_1 + u_2 v_2$$

- Độ lớn của vector u: $\|u\| = \sqrt{u_1^2 + u_2^2}$
- Hình chiếu của v lên u có độ dài: p
- Khi đó:

$$u^T v = \|u\| \|v\| \cos \varphi$$
$$= p\|u\|$$

  - φ là góc giữa u và v

Vì sao large margin

- $$\min_\theta \frac{1}{2}\sum_{j=1}^{n}\theta_j^2$$

$$\text{s.t.}\quad \theta^T x^{(i)} \geq 1 \quad \text{if} \quad y^{(i)} = 1$$

$$\theta^T x^{(i)} \leq -1 \quad \text{if} \quad y^{(i)} = 0$$

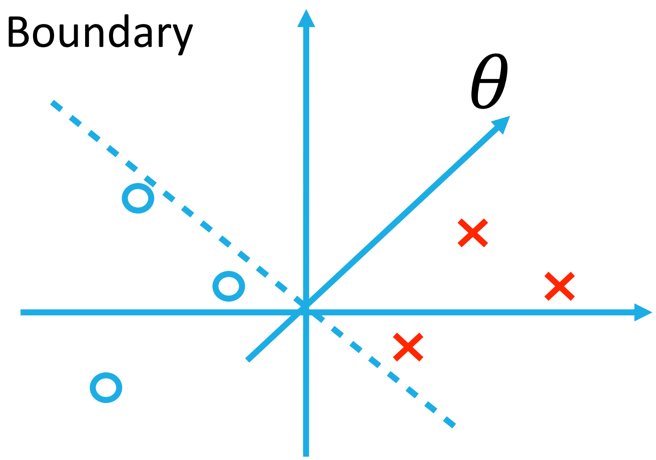
- Tích vô hướng:

$$\begin{aligned}\theta^T x^{(i)} &= p^{(i)}\|\theta\| \\ &= \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)}\end{aligned}$$

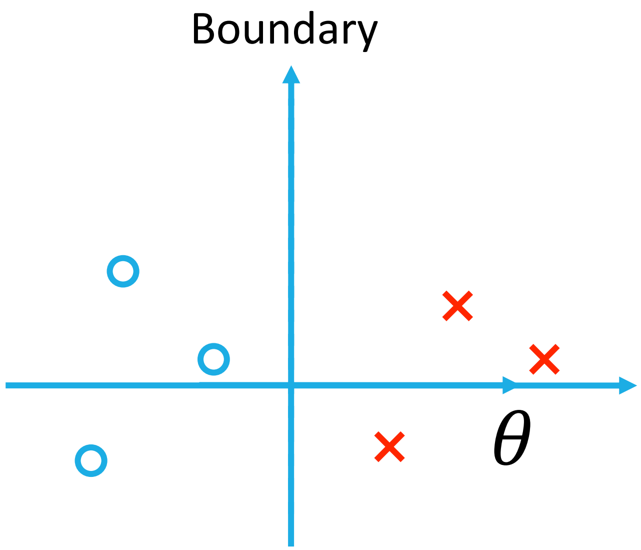
Vì sao large margin

- $$\min_\theta \frac{1}{2}\sum_{j=1}^{n}\theta_j^2$$
  $$\text{s.t.} \quad p^{(i)}\|\theta\| \geq 1 \quad \text{if} \quad y^{(i)} = 1$$
  $$p^{(i)}\|\theta\| \leq -1 \quad \text{if} \quad y^{(i)} = 0$$
- Với $p^{(i)}$ là độ dài hình chiếu của $x^{(i)}$ lên θ
- $p^{(i)}$ là khoảng cách giữa mẫu tới đường phân lớp
- Do $p^{(i)}\|\theta\| \geq 1$
  - $p^{(i)}$ càng nhỏ, độ dài θ phải càng lớn
  - $p^{(i)}$ càng lớn, độ dài θ có thể nhỏ

Vì sao large margin



Narrow margin



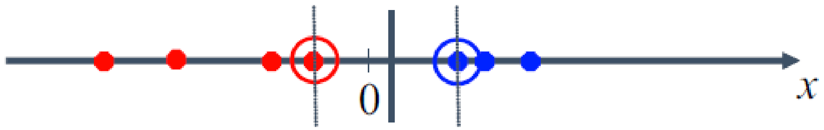
Large margin

Ánh xạ dữ liệu với Kernel

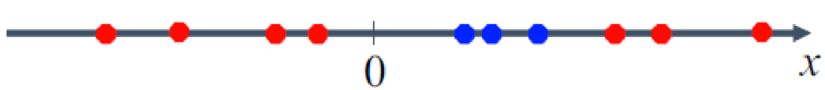
Phần 2

Phân lớp phi tuyến

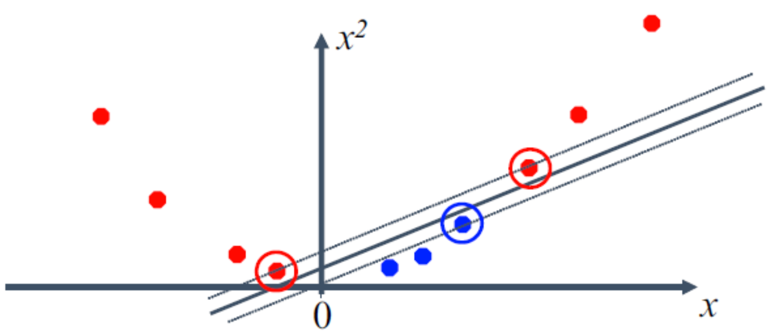
- Bộ phân lớp tuyến tính hoạt động tốt trên dữ liệu phân tách tuyến tính



- Một số dữ liệu không phân tách tuyến tính

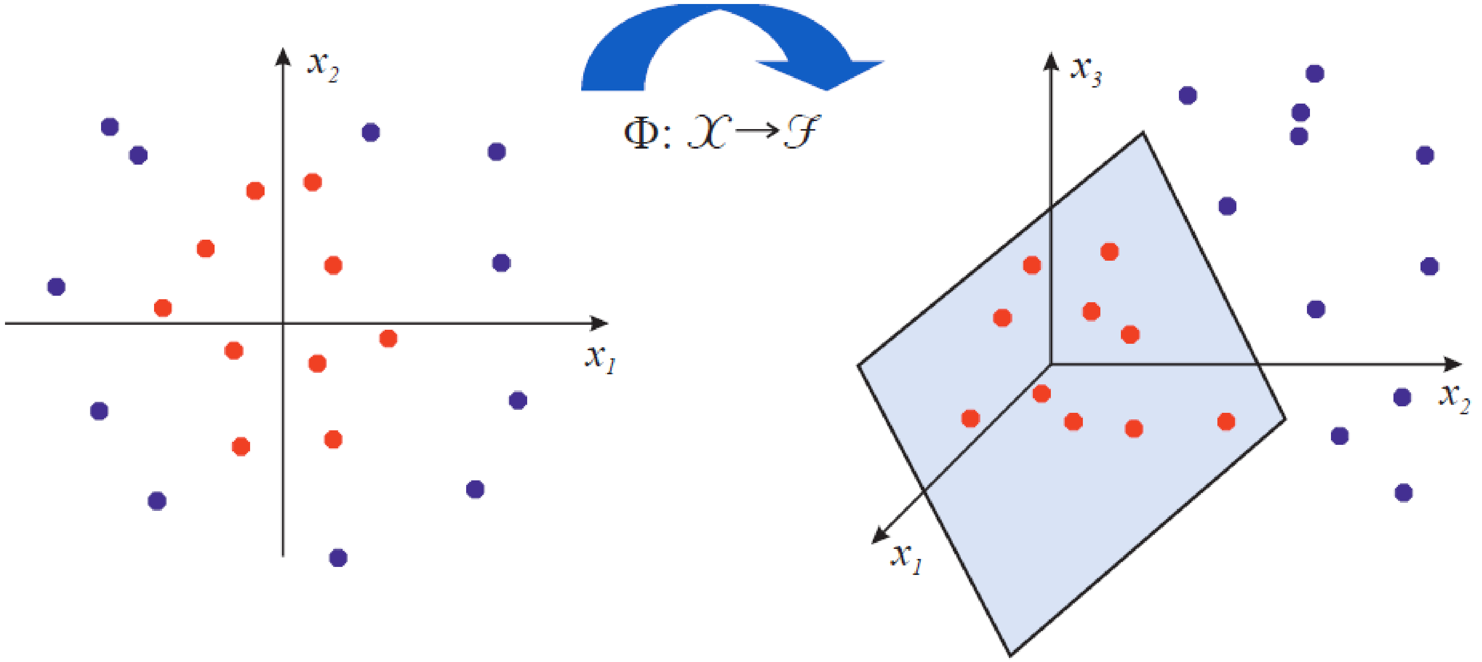


- Giải pháp: ánh xạ dữ liệu sang không gian cao chiều hơn



Phân lớp phi tuyến

- Giải pháp: ánh xạ dữ liệu sang không gian cao chiều hơn



Ánh xạ dữ liệu được thực hiện như thế nào với SVM?

Thuật toán học

$$\min_{\theta} \frac{1}{2} \sum_{j=1}^{n} \theta_j^2$$

$$\text{s.t.} \quad \theta^T x^{(i)} \geq 1 \quad \text{if} \quad y^{(i)} = 1$$

$$\theta^T x^{(i)} \leq -1 \quad \text{if} \quad y^{(i)} = 0$$

Đặt lại: y = -1 cho lớp âm

➡ $$\min_{\theta} \frac{1}{2} \sum_{j=1}^{n} \theta_j^2$$

$$\text{s.t.} \quad y^{(i)}(\theta^T x^{(i)}) \geq 1, i = 1, 2, ..., m$$

Thuật toán học

- Đặt $g_i(\theta) = -y^{(i)}(\theta^T x^{(i)}) + 1$

➡ $$\min_{\theta} \frac{1}{2}\sum_{j=1}^{n}\theta_j^2$$

$$\text{s.t.} \quad g_i(\theta) = -y^{(i)}(\theta^T x^{(i)}) + 1 \le 0$$

Chuyển về bài toán nhân tử Lagrange:

$$\mathcal{L}(\theta, \alpha) = \frac{1}{2}\sum_{j=1}^{n}\theta_j^2 - \sum_{i=1}^{m}\alpha_i[y^{(i)}(\theta^T x^{(i)}) - 1]$$

Mục tiêu: $\min\limits_{\theta,\alpha} \mathcal{L}(\theta, \alpha)$

Thuật toán học

$$\mathcal{L}(\theta, \alpha) = \frac{1}{2}\sum_{j=1}^{n}\theta_j^2 - \sum_{i=1}^{m}\alpha_i[y^{(i)}(\theta^T x^{(i)}) - 1]$$

Mục tiêu: $\min_{\theta,\alpha} \mathcal{L}(\theta, \alpha)$

Điều kiện Karush-Kuhn-Tucker (KKT)

Giả sử: θ^*, α^* là nghiệm

(1) $\frac{\partial}{\partial \theta_j} \mathcal{L}(\theta^*, \alpha^*) = 0, j = 0, ..., n$

(2) $\alpha_i^* g_i(\theta^*) = 0, i = 1, ..., m$

(3) $g_i(\theta^*) \leq 0, i = 1, ..., m$

(4) $\alpha^* \geq 0, i = 1, ..., m$

Thuật toán học

$$\mathcal{L}(\theta, \alpha) = \frac{1}{2}\sum_{j=1}^{n}\theta_j^2 - \sum_{i=1}^{m}\alpha_i[y^{(i)}(\theta^T x^{(i)}) - 1]$$

Mục tiêu: $\min\limits_{\theta,\alpha} \mathcal{L}(\theta, \alpha)$

Tìm theta:

(1) $$\nabla_\theta \mathcal{L}(\theta, \alpha) = \theta - \sum_{i=1}^{m}\alpha_i y^{(i)} x^{(i)} = 0$$

➡ $$\theta = \sum_{i=1}^{m}\alpha_i y^{(i)} x^{(i)}$$

Support vector

Điều kiện Karush-Kuhn-Tucker (KKT)

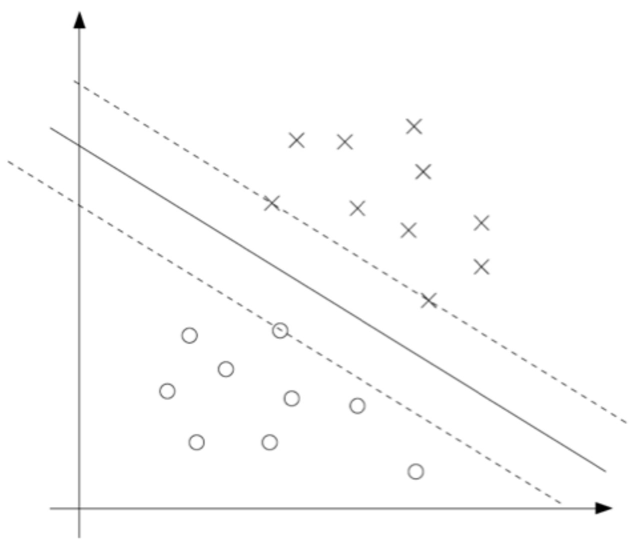
Giả sử: θ^*, α^* là nghiệm

(1) $$\frac{\partial}{\partial \theta_j} \mathcal{L}(\theta^*, \alpha^*) = 0, j = 0, ..., n$$

(2) $$\alpha_i^* g_i(\theta^*) = 0, i = 1, ..., m$$

(3) $$g_i(\theta^*) \leq 0, i = 1, ..., m$$

(4) $$\alpha_i^* \geq 0, i = 1, ..., m$$



Từ (2), (3), và (4):

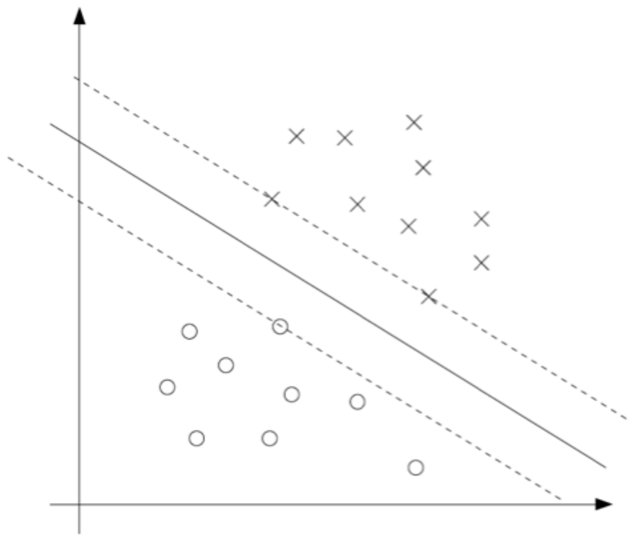
$$\text{if} \quad g_i(\theta^*) < 0, \text{then} \quad \alpha_i^* = 0, i = 1, ..., m$$

Những mẫu nằm ngoài margin: $\alpha_i^* = 0$

Support vector

Những mẫu nằm ngoài margin: $\alpha_i^* = 0$

$$\theta = \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)}$$



Predict:

$$\theta^T x = \left(\sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)}\right)^T x$$

$$= \sum_{i=1}^{m} \alpha_i y^{(i)} \langle x^{(i)}, x \rangle$$

Chỉ cần tính với những điểm nằm trên margin ➡ support vector

Ánh xạ đặc trưng

- Predict

$$\theta^T x = (\sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)})^T x$$
$$= \sum_{i=1}^{m} \alpha_i y^{(i)} \langle x^{(i)}, x \rangle$$

  - Có thể xem $\langle x^{(i)}, x \rangle$ là độ giống nhau

- Định nghĩa hàm

$$\phi(x) = [x \quad x^2 \quad x^3]^T$$

Thay $\langle x^{(i)}, x \rangle$ bằng $\langle \phi(x^{(i)}), \phi(x) \rangle$

Kernel

- Định nghĩa

$$K(x^{(i)}, x) = \langle \phi(x^{(i)}), \phi(x) \rangle$$

Predict:

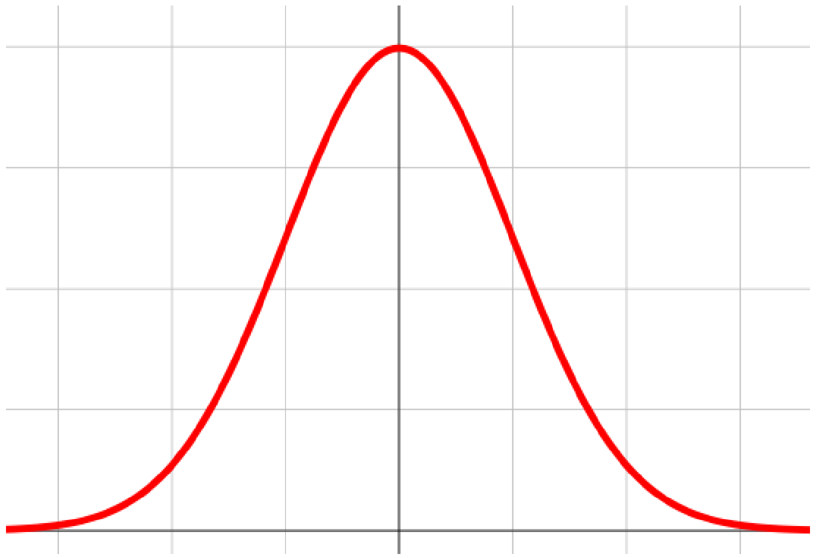
$$\begin{aligned} \theta^T x &= (\sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)})^T x \\ &= \sum_{i=1}^{m} \alpha_i y^{(i)} \langle x^{(i)}, x \rangle \end{aligned}$$

Thay $\langle x^{(i)}, x \rangle$ bằng $K(x^{(i)}, x)$

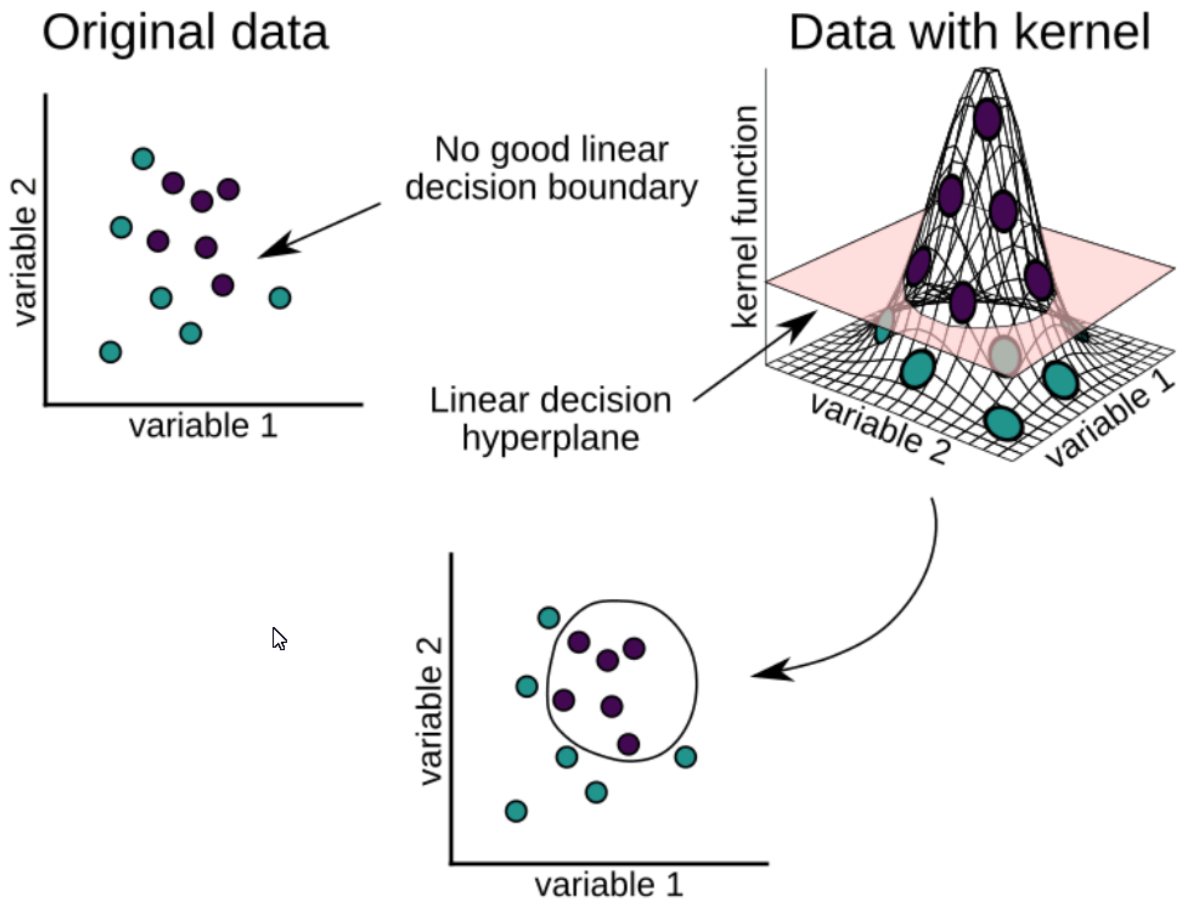
Kernel

- Gaussian (RBF) kernel

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



Kernel

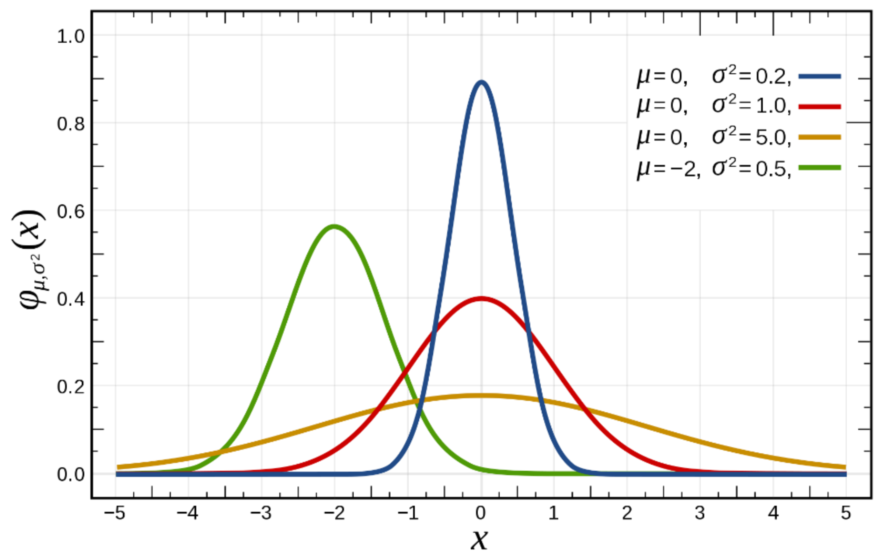


Original data
variable 2
variable 1
No good linear decision boundary
Data with kernel
kernel function
variable 2
variable 1
Linear decision hyperplane
variable 2
variable 1

Kernel

- Gaussian (RBF) kernel

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



- σ càng lớn, độ lớn càng nhỏ

Kernel

- Linear kernel

$$K(x,z) = x^T z$$

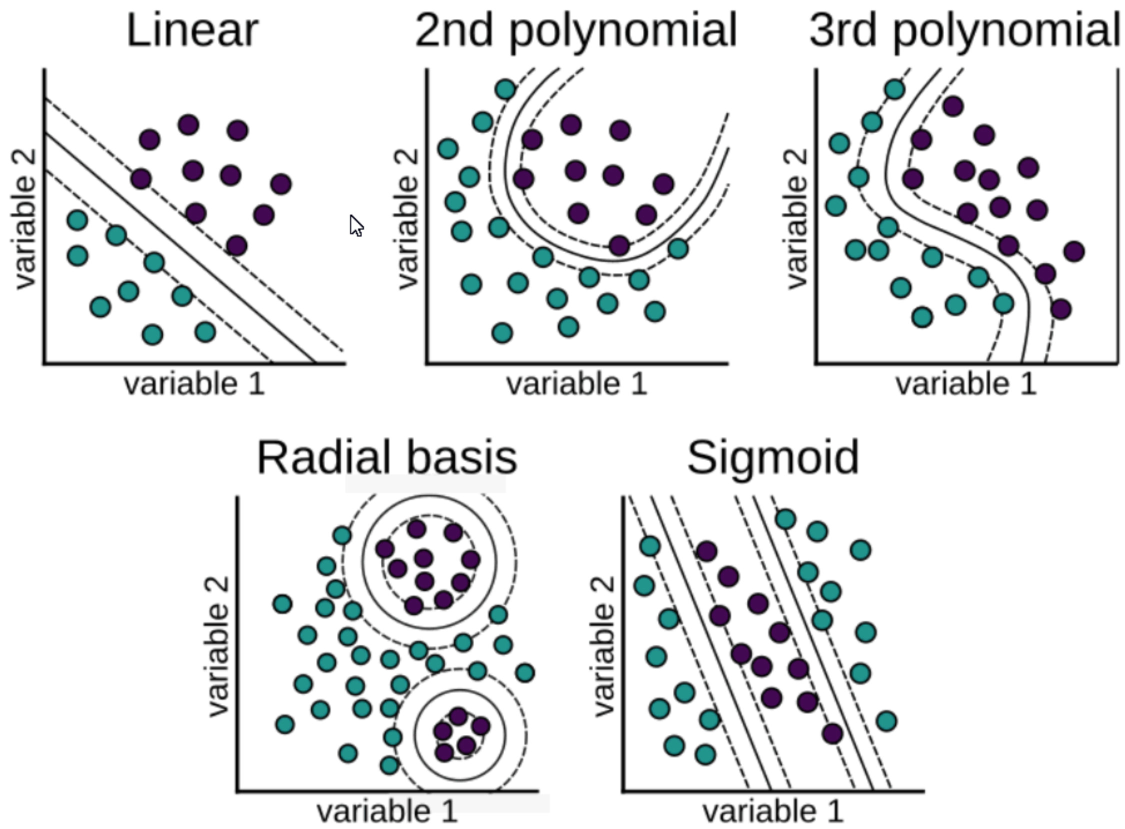
- Polynomial kernel

$$K(x,z) = (x^T z + c)^d$$

- Sigmoid kernel

$$K(x,z) = \tanh(a x^T z + b)$$

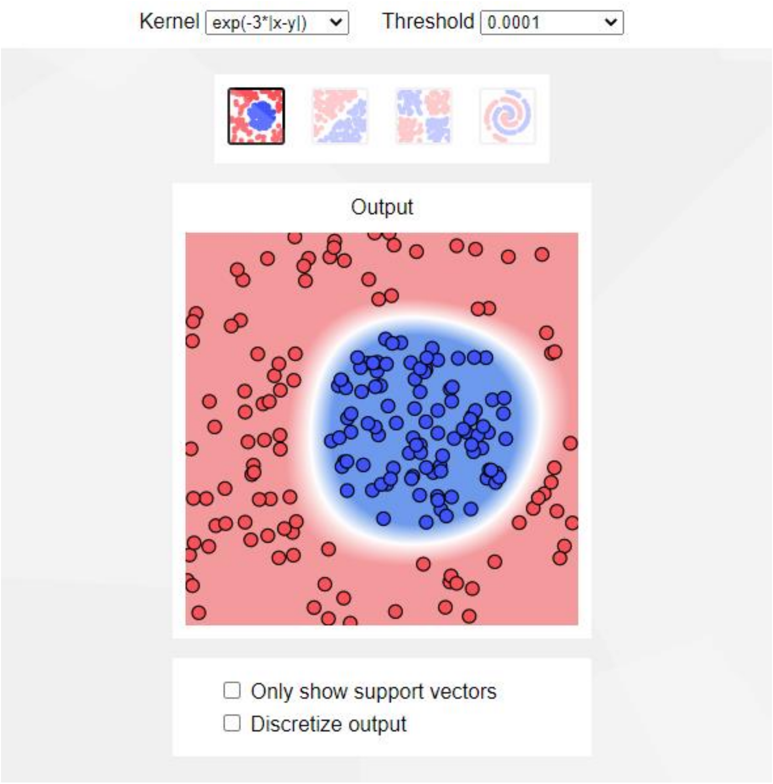
Kernel



Linear
2nd polynomial
3rd polynomial
Radial basis
Sigmoid
variable 1
variable 2

SVM Playground

- http://macheads101.com/demos/svm-playground



SVM Playground

- http://macheads101.com/demos/svm-playground