

Support Vector Machine

Đình Bách, Khánh Linh, Tiến Sơn, Trường Trí

PiMA 2021



Trình bày: Nhóm 8, SVM

August 8, 2021

1 Phân loại nhị phân

- Giới thiệu
- Phân loại nhị phân có giám sát

2 Bài toán biên cứng SVM

- Phát biểu bài toán
- Bài toán gốc
- Bài toán đối ngẫu

3 Phân loại biên mềm SVM

- Phát biểu bài toán
- Bài toán gốc
- Bài toán đối ngẫu

4 Kernel SVM

- Phát biểu bài toán
- Biến đổi dữ liệu ban đầu
- Kernel SVM

5 Thực hành

6 Tổng kết

Contents

1 Phân loại nhị phân

■ Giới thiệu

- Phân loại nhị phân có giám sát

2 Bài toán biên cứng SVM

- Phát biểu bài toán
- Bài toán gốc
- Bài toán đối ngẫu

3 Phân loại biên mềm SVM

- Phát biểu bài toán

- Bài toán gốc

- Bài toán đối ngẫu

4 Kernel SVM

- Phát biểu bài toán

- Biến đổi dữ liệu ban đầu

- Kernel SVM

5 Thực hành

6 Tổng kết



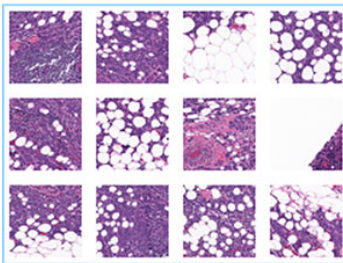
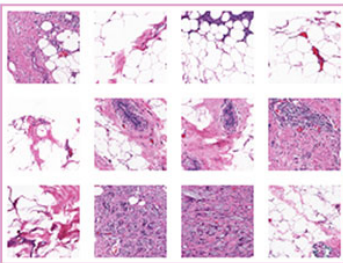
Là bài toán phân loại dữ liệu đã cho vào hai lớp riêng biệt. Ví dụ:

- Phân loại email quan trọng và email rác.



Là bài toán phân loại dữ liệu đã cho vào hai lớp riêng biệt. Ví dụ:

- Phát hiện và chẩn đoán ung thư.

Positive**Negative**

Contents

1 Phân loại nhị phân

- Giới thiệu
- Phân loại nhị phân có giám sát

2 Bài toán biên cứng SVM

- Phát biểu bài toán
- Bài toán gốc
- Bài toán đối ngẫu

3 Phân loại biên mềm SVM

- Phát biểu bài toán
- Bài toán gốc
- Bài toán đối ngẫu

4 Kernel SVM

- Phát biểu bài toán
- Biến đổi dữ liệu ban đầu
- Kernel SVM

5 Thực hành

6 Tổng kết



...

Bài toán phân loại thuộc lớp các bài toán học có giám sát.

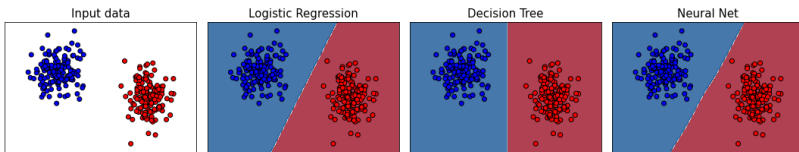
- Xây dựng mô hình: từ tập huấn luyện $(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots, (\mathbf{x}_N, \mathbf{y}_N)$ và từ một họ hàm số được tham số hóa bởi θ , ta đi tìm ta đi tìm tham số θ^* sao cho:

$$f(\mathbf{x}_n, \theta^*) = y_n \quad \text{với mọi } n = 1, \dots, N. \quad (1)$$

- Kiểm chứng: hoàn thiện mô hình thông qua tập kiểm thử. Khi có một dữ liệu \mathbf{x}_n mới, ta có thể tìm nhãn dán tương ứng qua $\mathbf{y}_n = f(\mathbf{x}_n)$



Một số thuật toán



Contents

1 Phân loại nhị phân

- Giới thiệu
- Phân loại nhị phân có giám sát

2 Bài toán biên cứng SVM

- Phát biểu bài toán
- Bài toán gốc
- Bài toán đối ngẫu

3 Phân loại biên mềm SVM

- Phát biểu bài toán

- Bài toán gốc

- Bài toán đối ngẫu

4 Kernel SVM

- Phát biểu bài toán
- Biến đổi dữ liệu ban đầu
- Kernel SVM

5 Thực hành

6 Tổng kết



Huấn luyện

■ Đầu vào:

- Đặc trưng: $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)$ với $\mathbf{x}_i \in \mathbb{R}^d$
- Nhãn: $y_i \in \{-1, 1\}$

■ Đầu ra: Hệ số \mathbf{w} và b của siêu phẳng

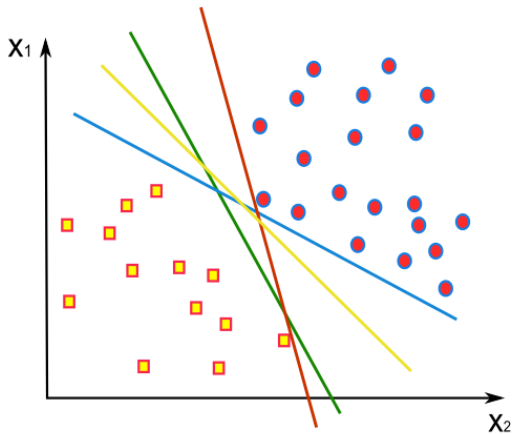
$$\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$$

phân dữ liệu thành hai lớp.

Truy vấn

- Đầu vào: Điểm dữ liệu mới $\mathbf{z} \in \mathbb{R}^d$
- Đầu ra: Dự đoán $f(\mathbf{z}) := \text{sgn}(\langle \mathbf{w}, \mathbf{z} \rangle + b)$





Contents

1 Phân loại nhị phân

- Giới thiệu
- Phân loại nhị phân có giám sát

2 Bài toán biên cứng SVM

- Phát biểu bài toán
- Bài toán gốc
- Bài toán đối ngẫu

3 Phân loại biên mềm SVM

- Phát biểu bài toán
- Bài toán gốc
- Bài toán đối ngẫu

4 Kernel SVM

- Phát biểu bài toán
- Biến đổi dữ liệu ban đầu
- Kernel SVM

5 Thực hành

6 Tổng kết



Định nghĩa Lề

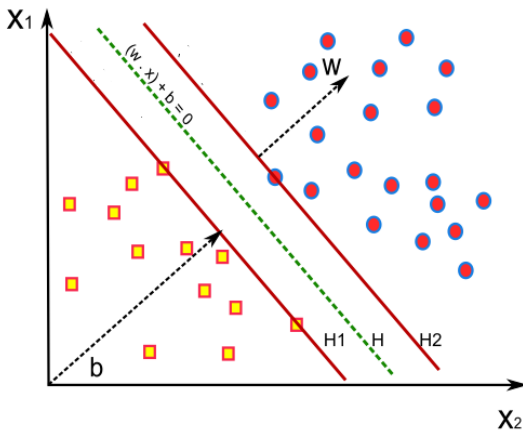
Định nghĩa

Lề là khoảng cách từ siêu phẳng phân tách tới điểm dữ liệu gần nhất, với giả thiết bộ dữ liệu có tính tách biệt tuyến tính. Nếu ta giả sử x_a là điểm dữ liệu gần với siêu mặt phẳng phân cách nhất thì ta có lề r được tính như sau :

$$r = \frac{|\langle \mathbf{w}, \mathbf{x}_a \rangle + b|}{\|\mathbf{w}\|} \quad (2)$$



Cái lẽ lớn nhất



Ta chỉ quan tâm đến hướng của \mathbf{w} nên ta coi $\|\mathbf{w}\| = 1$. Đặt $\mathbf{w}' = \frac{\mathbf{w}}{r}$ và $b' = \frac{b}{r}$. Ta có hệ quả:

- 1 Lề của siêu mặt phẳng phân cách có công thức :

$$\|\mathbf{w}'\| = \frac{\|\mathbf{w}\|}{r} = \frac{1}{r} \Leftrightarrow r = \frac{1}{\|\mathbf{w}'\|} \quad (3)$$

- 2 Siêu mặt phẳng $\langle \mathbf{w}', \mathbf{x}_a \rangle + b' = 1$ đi qua \mathbf{x}_a :

$$r = \frac{\|\langle \mathbf{w}', \mathbf{x}_a \rangle + b'\|}{\|\mathbf{w}'\|} = \frac{1}{\|\mathbf{w}'\|} \Leftrightarrow y_n(\langle \mathbf{w}', \mathbf{x}_a \rangle + b') = 1$$

Vậy không làm thay đổi siêu mặt phẳng ta chuẩn hóa : $|\langle \mathbf{w}, \mathbf{x}_a \rangle + b| = 1$ với \mathbf{x}_a là điểm gần siêu mặt phẳng nhất. Ta có $r = \frac{1}{\|\mathbf{w}\|}$:

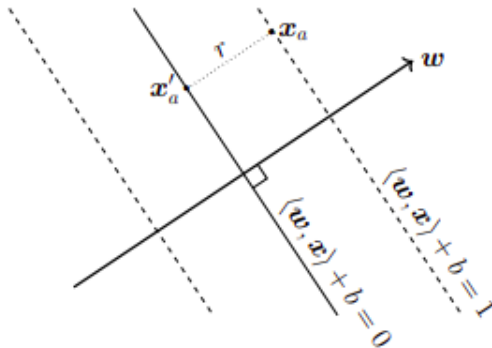


Suy ra các điểm dữ liệu còn lại phải cách siêu phẳng một khoảng lớn hơn hoặc bằng r :

$$\frac{y_n(\langle \mathbf{w}, \mathbf{x}_n \rangle + b)}{\|\mathbf{w}\|} \geq \frac{1}{\|\mathbf{w}\|} \Leftrightarrow y_n(\langle \mathbf{w}, \mathbf{x}_n \rangle + b) \geq 1. \quad (4)$$



Khoảng cách giữa hai siêu phẳng $\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$ và $\langle \mathbf{w}, \mathbf{x} \rangle + b = 1$ là lề cần tối ưu.



Bài toán tối ưu của chúng ta trở thành:

$$\max_{\mathbf{w}, b} \frac{1}{\|\mathbf{w}\|} \quad (5)$$

$$\text{subject to } y_n(\langle \mathbf{w}, \mathbf{x}_n \rangle + b) \geq 1 \quad \forall n = 1, \dots, N \quad (6)$$

Tương đương:

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \quad (7)$$

$$\text{subject to } y_n(\langle \mathbf{w}, \mathbf{x}_n \rangle + b) \geq 1 \quad \forall n = 1, \dots, N \quad (8)$$



Contents

1 Phân loại nhị phân

- Giới thiệu
- Phân loại nhị phân có giám sát

2 Bài toán biên cứng SVM

- Phát biểu bài toán
- Bài toán gốc
- Bài toán đối ngẫu

3 Phân loại biên mềm SVM

- Phát biểu bài toán
- Bài toán gốc
- Bài toán đối ngẫu

4 Kernel SVM

- Phát biểu bài toán
- Biến đổi dữ liệu ban đầu
- Kernel SVM

5 Thực hành

6 Tổng kết



Hàm Lagrangian

Hàm Lagrangian ứng với bài toán lề cứng là :

$$\mathcal{L}(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{n=1}^N \alpha_n (y_n (\langle \mathbf{w}, \mathbf{x}_n \rangle + b) - 1) \quad (9)$$

với $\alpha_i \geq 0 \ \forall \ i = 1, \dots, N$ là nhân tử Lagrange tương ứng với ràng buộc của bài toán gốc. Từ đây bài toán tối ưu của ta trở thành bài toán tối ưu mới.

Điều kiện KKT

Nghiệm của bài toán tối ưu mới thỏa mãn hệ điều kiện KKT:

$$1 - y_n(\langle \mathbf{w}, \mathbf{x}_n \rangle + b) \leq 0 \quad (10)$$

$$\alpha_n \geq 0 \quad (11)$$

$$\alpha_n(1 - y_n(\langle \mathbf{w}, \mathbf{x}_n \rangle + b)) = 0 \quad (12)$$

$$\mathbf{w} = \sum_{n=1}^N \alpha_n y_n \mathbf{x}_n \quad (13)$$

$$\sum_{n=1}^N \alpha_n y_n = 0 \quad (14)$$



Bài toán đối ngẫu

Sau khi thay các điều kiện của nghiệm vào hàm Lagrangian (9) ta có bài toán tối ưu trở thành :

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle + \sum_{i=1}^N \alpha_i \\ \text{subject to} \quad & \sum_{i=1}^N y_i \alpha_i = 0 \\ & 0 \leq \alpha_i \quad \forall i = 1, 2, \dots, N \end{aligned} \tag{15}$$

Tìm nghiệm bài toán gốc

Bài toán trên là bài toán có dạng quy hoạch toàn phương. Giải bài toán này ta tìm được α^* . Sau đó ta quay lại tìm \mathbf{w}^*, b^* theo công thức :

$$\mathbf{w}^* = \sum_{n=1}^N \alpha_n^* y_n \mathbf{x}_n \quad (16)$$

$$b^* = \frac{1}{|SV|} \sum_{n \in SV} (y_n - \sum_{m \in SV} \alpha_m^* y_m \langle \mathbf{x}_m, \mathbf{x}_n \rangle) \quad (17)$$



Nhận xét về kết quả của bài toán đối ngẫu

Phương trình (13) cho ta biết một kết quả thú vị là hoặc $\alpha_i = 0$ hoặc $1 - y_n(\langle \mathbf{w}, \mathbf{x}_n \rangle + b) = 0$.

- 1 Nếu $\alpha = 0$ là những điểm không đóng góp vào việc tính \mathbf{w}^*, b^*
- 2 Nếu $1 - y_n(\langle \mathbf{w}, \mathbf{x}_n \rangle + b) = 0 \Rightarrow \langle \mathbf{w}, \mathbf{x}_n \rangle + b = y_n$ tức là những điểm nằm trên lề và đóng góp vào việc tính toán \mathbf{w}^*, b^* do $\alpha \neq 0$.
- 3 ta thấy giá trị tối ưu \mathbf{w}^* là tổ hợp tuyến tính của các vectơ hỗ trợ:

$$\mathbf{w}^* = \sum \alpha_i^* y_i \mathbf{x}_i$$



Tại sao bài toán đối ngẫu lại thường được sử dụng?

- 1 Vai trò của các vectơ hỗ trợ được thấy rõ.
- 2 Nhân tử Lagrange gắn với số điểm dữ liệu \Rightarrow bài toán tối ưu có thể giải quyết các bài toán dữ liệu có số chiều lớn
- 3 Vectơ đối ngẫu thừa thớt nên cho phép một số phương pháp tìm nghiệm áp dụng để tìm ra kết quả rất nhanh.
- 4 Cả hàm mục tiêu và ràng buộc đều chỉ phụ thuộc vào tích vô hướng giữa điểm dữ liệu \Rightarrow áp dụng thủ thuật kernel



Contents

1 Phân loại nhị phân

- Giới thiệu
- Phân loại nhị phân có giám sát

2 Bài toán biên cứng SVM

- Phát biểu bài toán
- Bài toán gốc
- Bài toán đối ngẫu

3 Phân loại biên mềm SVM

■ Phát biểu bài toán

- Bài toán gốc
- Bài toán đối ngẫu

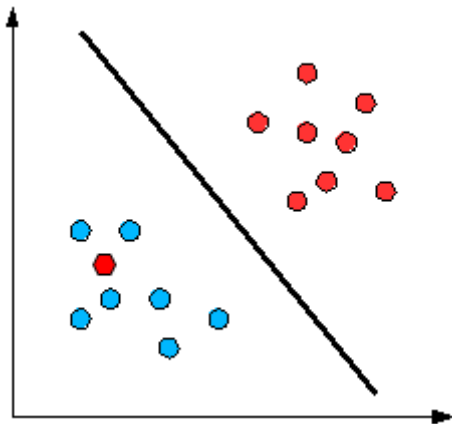
4 Kernel SVM

- Phát biểu bài toán
- Biến đổi dữ liệu ban đầu
- Kernel SVM

5 Thực hành

6 Tổng kết





Hình: Dữ liệu đầu vào *gần* tách biệt tuyến tính.



Contents

1 Phân loại nhị phân

- Giới thiệu
- Phân loại nhị phân có giám sát

2 Bài toán biên cứng SVM

- Phát biểu bài toán
- Bài toán gốc
- Bài toán đối ngẫu

3 Phân loại biên mềm SVM

- Phát biểu bài toán

■ Bài toán gốc

- Bài toán đối ngẫu

4 Kernel SVM

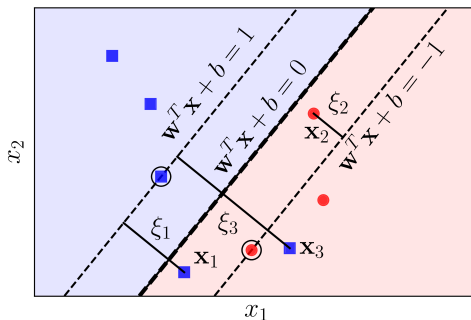
- Phát biểu bài toán
- Biến đổi dữ liệu ban đầu
- Kernel SVM

5 Thực hành

6 Tổng kết



Cho phép một số điểm được vượt biên



Hình: Với những điểm dữ liệu nằm phía bên kia đường biên lề đúng:

$$\xi_n = |\langle \mathbf{w}, \mathbf{x}_n \rangle + b - y_n|$$



Cho phép một số điểm được vượt biên

Ta sẽ mở rộng ràng buộc về lề bằng cách thêm cách biến bù $\xi_n (\geq 0)$ gắn với điểm dữ liệu \mathbf{x}_n :

$$\langle \mathbf{w}, \mathbf{x}_n \rangle + b \geq 1 - \xi_n \quad \text{với } y_n = 1 \quad (18)$$

$$\langle \mathbf{w}, \mathbf{x}_n \rangle + b \leq -1 + \xi_n \quad \text{với } y_n = -1 \quad (19)$$

- Với điểm nằm ở đúng phần không gian bên lề: $\xi_n = 0$
- Những điểm còn lại: $\xi_n = |\langle \mathbf{w}, \mathbf{x}_n \rangle + b - y_n|$

Kết hợp điều kiện 18 và 19, ta có điều kiện ràng buộc mềm:

$$y_n(\langle \mathbf{w}, \mathbf{x}_n \rangle + b) \geq 1 - \xi_n \quad \forall n = 1, \dots, N \quad (20)$$

$$\xi_n \geq 0 \quad (21)$$



Tối thiểu phân loại lỗi

Ta thêm thông tin về những điểm nằm sai phía vào hàm mục tiêu:

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{n=1}^N \xi_n \quad (22)$$

Trong đó, $C > 0$ là hằng số phạt:

- C thể hiện mức độ ưu tiên giữa tối đa lề của siêu phẳng phân tách về tối thiểu các biến bù.
- C càng lớn thì lỗi phân loại trong huấn luyện càng bị phạt nặng.



Bài toán tối ưu

Phát biểu bài toán tối ưu của phân loại biên mềm SVM:

$$\min_{\mathbf{w}, b, \xi} \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{n=1}^N \xi_n \quad (23)$$

$$\text{subject to} \quad y_n(\langle \mathbf{w}, \mathbf{x}_n \rangle + b) \geq 1 - \xi_n \quad (24)$$

$$\xi_n \geq 0 \quad (25)$$

- $N + D + 1$ ẩn, $2N$ ràng buộc.
- Bài toán tối ưu lồi với ràng buộc tuyến tính có thể giải bằng quy hoạch toàn phương.



Mềm tốt hơn cứng

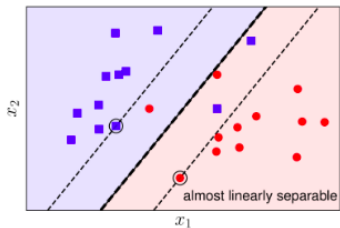
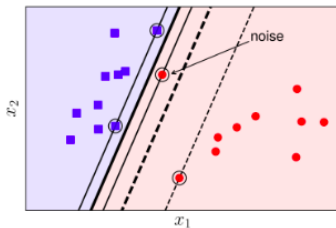
Khi dữ liệu huấn luyện không tách biệt tuyến tính...

- Phần lớn dữ liệu trong thực tế không có tính chất này.
- Bài toán tối ưu của Biên cứng SVM vô nghiệm.

và cả khi dữ liệu huấn luyện tách biệt tuyến tính.

- Biên mềm SVM giảm bớt sự tác động của các điểm ngoại lai đến vị trí của đường biên quyết định.
- Mô hình phân loại của khả năng tổng quát lớn hơn.

Mềm tốt hơn cứng



Hình: Đường biên quyết định của mô hình phân loại biên cứng SVM (nét liền) và biên mềm SVM (nét gạch).



Contents

1 Phân loại nhị phân

- Giới thiệu
- Phân loại nhị phân có giám sát

2 Bài toán biên cứng SVM

- Phát biểu bài toán
- Bài toán gốc
- Bài toán đối ngẫu

3 Phân loại biên mềm SVM

- Phát biểu bài toán

- Bài toán gốc

- Bài toán đối ngẫu

4 Kernel SVM

- Phát biểu bài toán
- Biến đổi dữ liệu ban đầu
- Kernel SVM

5 Thực hành

6 Tổng kết



Hàm Lagrangian

$$\mathcal{L}(\mathbf{w}, b, \xi, \alpha, \gamma) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{n=1}^N \xi_n \quad (26)$$
$$- \sum_{n=1}^N \alpha_n (y_n (\langle \mathbf{w}, \mathbf{x}_n \rangle + b) - 1 + \xi_n) - \sum_{n=1}^N \gamma_n \xi_n$$

Trong đó $\alpha_n \geq 0$ và $\gamma_n \geq 0$ là các nhân tử Lagrange tương ứng với hai ràng buộc của bài toán tối ưu.



Bài toán đối ngẫu

$$\max_{\alpha} \quad -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle + \sum_{i=1}^N \alpha_i \quad (27)$$

$$\text{subject to} \quad \sum_{i=1}^N y_i \alpha_i = 0 \quad (28)$$

$$0 \leq \alpha_i \leq C \quad \forall i = 1, 2, \dots, N \quad (29)$$

- Ràng buộc 29 khiến bài toán giải hiệu quả hơn.



Nhận xét về kết quả bài toán đối ngẫu

Từ điều kiện KKT

- 1 Nếu $\alpha_n = 0$ thì $y_i(\langle w^*, \mathbf{x}_n \rangle + b) > 1$.
- 2 Nếu $0 < \alpha_n < C$ thì $y_i(\langle w^*, \mathbf{x}_n \rangle + b) = 1$.
- 3 Nếu $\alpha_n = C$ thì $y_i(\langle w^*, \mathbf{x}_n \rangle + b) \leq 1$.



Contents

1 Phân loại nhị phân

- Giới thiệu
- Phân loại nhị phân có giám sát

2 Bài toán biên cứng SVM

- Phát biểu bài toán
- Bài toán gốc
- Bài toán đối ngẫu

3 Phân loại biên mềm SVM

- Phát biểu bài toán
- Bài toán gốc
- Bài toán đối ngẫu

4 Kernel SVM

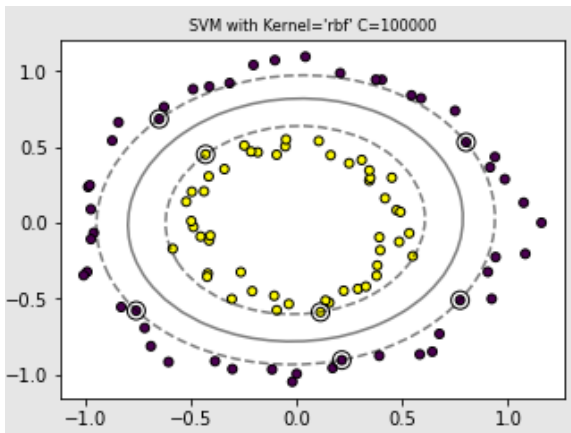
- Phát biểu bài toán
- Biến đổi dữ liệu ban đầu
- Kernel SVM

5 Thực hành

6 Tổng kết



Không còn điều kiện tách biệt tuyến tính

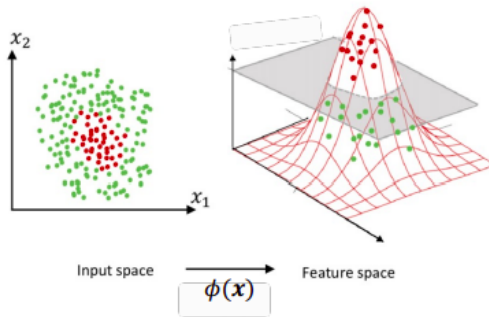


Hình: Sử dụng kernel SVM để giải quyết bài toán với dữ liệu không tách biệt tuyến tính.



Sử dụng mô hình tuyến tính

- Bước 1: Biến đổi dữ liệu ban đầu sang không gian khác, thường có nhiều chiều hơn.
- Bước 2: Sử dụng mô hình tuyến tính như biên cứng hay biên mềm SVM.



Contents

1 Phân loại nhị phân

- Giới thiệu
- Phân loại nhị phân có giám sát

2 Bài toán biên cứng SVM

- Phát biểu bài toán
- Bài toán gốc
- Bài toán đối ngẫu

3 Phân loại biên mềm SVM

- Phát biểu bài toán
- Bài toán gốc
- Bài toán đối ngẫu

4 Kernel SVM

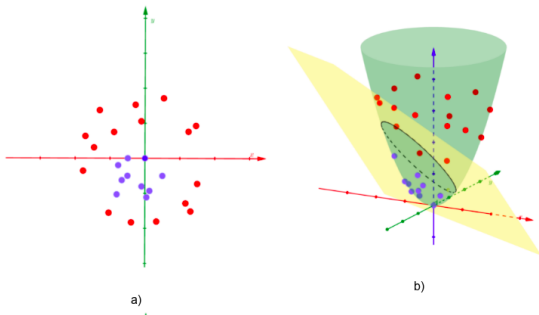
- Phát biểu bài toán
- Biến đổi dữ liệu ban đầu
- Kernel SVM

5 Thực hành

6 Tổng kết



Thêm đặc trưng



Hình: Thêm đặc trưng dẫn đến việc dữ liệu từ không tách biệt tuyến trở thành tách biệt tuyến tính (a) Tập dữ liệu (x, y) trong \mathbb{R}^2 (b) Tập dữ liệu $(x, y, x^2 + y^2)$ trong \mathbb{R}^3 .



Sử dụng mô hình tuyến tính

Bài toán gốc của biên mềm SVM:

$$\min_{\mathbf{w}, b, \xi} \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{n=1}^N \xi_n \quad (30)$$

$$\text{subject to} \quad y_n(\langle \mathbf{w}, \Phi(\mathbf{x}_n) \rangle + b) \geq 1 - \xi_n \quad (31)$$

$$\xi_n \geq 0 \quad (32)$$

Bài toán đối ngẫu của biên mềm SVM:

$$\max_{\alpha} \quad -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle + \sum_{i=1}^N \alpha_i$$

$$\text{subject to} \quad \sum_{i=1}^N y_i \alpha_i = 0 \quad (33)$$

$$0 \leq \alpha_i \leq C \quad \forall i = 1, 2, \dots, N$$



Contents

1 Phân loại nhị phân

- Giới thiệu
- Phân loại nhị phân có giám sát

2 Bài toán biên cứng SVM

- Phát biểu bài toán
- Bài toán gốc
- Bài toán đối ngẫu

3 Phân loại biên mềm SVM

- Phát biểu bài toán
- Bài toán gốc
- Bài toán đối ngẫu

4 Kernel SVM

- Phát biểu bài toán
- Biến đổi dữ liệu ban đầu
- Kernel SVM

5 Thực hành

6 Tổng kết



Kernelized method

Huấn luyện:

$$\begin{aligned}
 \max_{\alpha} \quad & -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle + \sum_{i=1}^N \alpha_i \\
 \text{subject to} \quad & \sum_{i=1}^N y_i \alpha_i = 0 \\
 & 0 \leq \alpha_i \leq C \quad \forall i = 1, 2, \dots, N
 \end{aligned} \tag{34}$$

Đưa ra dự đoán:

$$f(\mathbf{z}) = \langle \mathbf{w}^*, \Phi(\mathbf{z}) \rangle + b^* = \sum_{\mathbf{x}_i \in SV} \alpha_i^* y_i \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{z}) \rangle + b^* \tag{35}$$



Kernel Trick

Hàm nhân

Hàm nhân K ứng với $\Phi(\cdot) : X \rightarrow F$ là hàm sao cho $\mathbf{x}, \mathbf{z} \in X$:

$$K(\mathbf{x}, \mathbf{z}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{z}) \rangle$$

Thường có thể tính toán trực tiếp $K(\mathbf{x}, \mathbf{z})$ mà không cần tính $\Phi(\mathbf{x}), \Phi(\mathbf{z})$.

Examples

Ví dụ

$$\begin{aligned}\Phi(\mathbf{x})^T \Phi(\mathbf{z}) &= [x_1^2, \sqrt{2}x_1x_2, x_2^2][z_1^2, \sqrt{2}z_1z_2, z_2^2]^T \\ &= \langle \mathbf{x}, \mathbf{z} \rangle^2 = k(\mathbf{x}, \mathbf{z})\end{aligned}$$



Định lý Mercer

Hàm xác định dương

Một hàm nhân đối xứng $k : X \times X \rightarrow \mathbb{R}$ là bán xác định dương nếu với một tập hữu hạn $\{x_1, \dots, x_n\} \in X$, ma trận nhân của tập này:

$$K = (k(x_i, x_j))_{ij}$$

là bán xác định dương.

Định lý Mercer

Hàm đối xứng $k(w, x)$ có thể được biểu diễn dưới dạng một tích vô hướng:

$$k(w, x) = \langle \Phi(w), \Phi(x) \rangle$$

của một ánh xạ Φ khi và chỉ khi $k(w, x)$ là bán xác định dương.



Đưa ra dự đoán

$$\begin{aligned}
 f(\mathbf{z}) &= \langle \mathbf{w}^*, \Phi(\mathbf{z}) \rangle + b^* \\
 &= \sum_{\mathbf{x}_i \in SV} \alpha_i y_i \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{z}) \rangle + \frac{1}{|SV|} \sum_{\mathbf{x}_n \in SV} (y_n - \sum_{\mathbf{x}_m \in SV} \alpha_m^* y_m \langle \Phi(\mathbf{x}_m), \Phi(\mathbf{x}_n) \rangle) \\
 &= \sum_{\mathbf{x}_i \in SV} \alpha_i y_i k(\mathbf{x}_i, \mathbf{z}) + \frac{1}{|SV|} \sum_{\mathbf{x}_n \in SV} (y_n - \sum_{\mathbf{x}_m \in SV} \alpha_m^* y_m k(\mathbf{x}_m, \mathbf{x}_n))
 \end{aligned}$$



Các hàm nhân phổ biến

- Hàm nhân tuyến tính: $k(\mathbf{x}, \mathbf{z}) = \mathbf{x}^T \mathbf{z}$
- Hàm nhân đa thức: $k(\mathbf{x}, \mathbf{z}) = (r + \gamma \mathbf{x}^T \mathbf{z})^d$
- Hàm nhân Radial Basis Kernel (Gaussian):
 $k(\mathbf{x}, \mathbf{z}) = \exp(-\gamma \|\mathbf{x} - \mathbf{z}\|_2^2), \quad \gamma > 0$
- Hàm nhân Sigmoid: $k(\mathbf{x}, \mathbf{z}) = \tanh(\gamma \mathbf{x}^T \mathbf{z} + r)$



Dữ liệu tự sinh

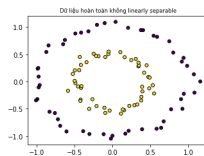
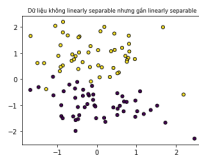
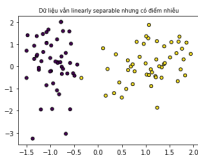
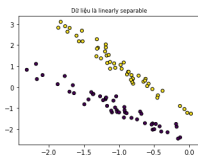
Ta sẽ chạy trên 4 bộ dữ liệu được sinh từ sklearn:

- Dữ liệu là linearly separable
- Dữ liệu vẫn linearly separable nhưng có điểm nhiễu
- Dữ liệu không linearly separable nhưng gần linearly separable
- Dữ liệu hoàn toàn không linearly separable



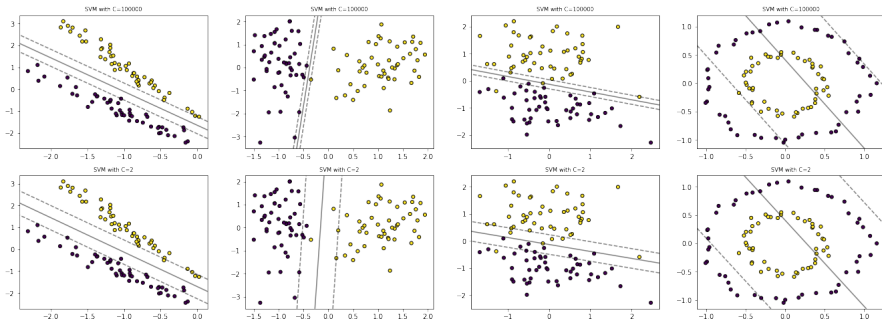
Thực hành trên dữ liệu tự sinh

Mỗi tập dữ liệu gồm 100 điểm dữ liệu, có hình dạng như sau:



Chạy mô hình với kernel='linear'

Chạy mô hình SVM thông thường lần lượt với $C=100000$ và $C=2$ thì ta được đường thẳng phân cách và margin như sau:



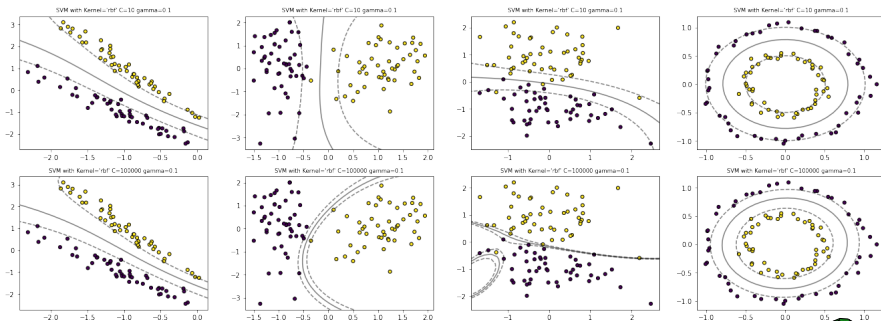
Chạy mô hình với kernel='linear'

- Với hằng số C nhỏ thì mô hình hoạt động tốt hơn khi dữ liệu có điểm nhiễu.
- SVM thông thường không cho kết quả đúng khi dữ liệu không linearly separable



Chạy mô hình với kernel='rbf'

Tiếp theo, chạy mô hình SVM với kernel rbf lần lượt với $C=10$ và $C=100000$, hằng số $\gamma=0.1$ thì ta được đường thẳng phân cách và lề như sau:



Chạy mô hình với kernel='rbf'

- Với kernel rbf thì mô hình chạy tốt và ổn định trên tập dữ liệu linearly separable và không linearly separable
- Mô hình hoạt động không ổn định với 2 tập dữ liệu còn lại



Nhận xét

- SVM soft margin hoạt động tốt khi dữ liệu có dạng linearly separable.
- Khi dữ liệu không có dạng linearly thì SVM với kernel='linear' không thể cho ra đáp án.
- Với SVM kernel mô hình chạy tốt với các dữ liệu không linearly separable, còn dữ liệu linearly separable vẫn chạy được nhưng phụ thuộc nhiều vào tham số.



Tổng kết

- Support Vector Machine là bài toán đi tìm mặt phân cách sao cho lề là lớn nhất.
- Bài toán tối ưu trong SVM là một bài toán lồi với hàm mục tiêu là lồi chặt, nghiệm của bài toán này là duy nhất.
- Để thuận tiện thường sẽ giải bài toán đối ngẫu thay vì bài toán tối ưu gốc.
- Với các bài toán mà dữ liệu gần tách biệt tuyến tính hoặc không tách biệt tuyến tính, có những cải tiến khác của SVM để thích nghi với dữ liệu đó.



Tài liệu tham khảo I

The Mathematics of Data Science

