Kernel Support Vector Machine

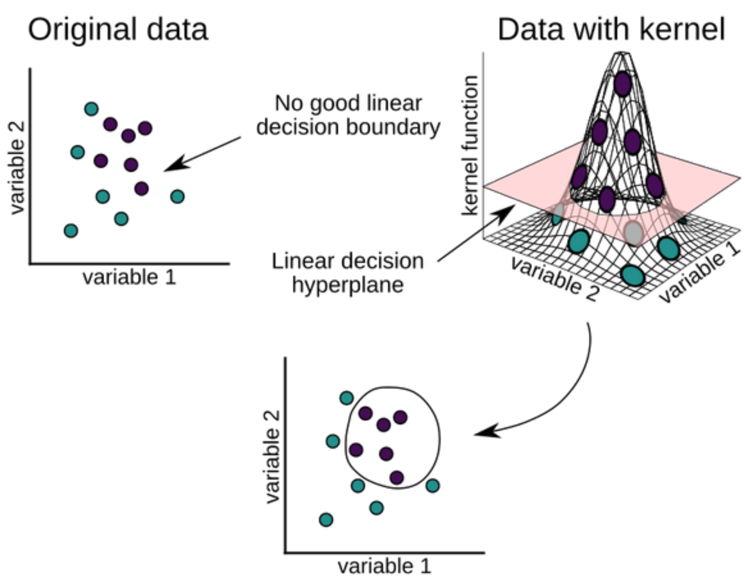
Trần Phú Vinh

June 2020

1 Đặt vấn đề

Các điểm dữ liệu trong thực tế rất khó xảy ra trường hợp linearly separable. Một cách giải quyết trong bài trước là chấp nhận đánh đổi sai sót để tìm được các nghiệm của thuật toán SVM tốt hơn (soft margin)

Một cách giải quyết khác được giới thiệu trong bài này là tìm cách ánh xạ các điểm dữ liệu đã có lên một không gian mới có số chiều lớn hơn (ví dụ từ không gian 2 chiều lên 3 chiều hoặc hơn 3 chiều), mà tại không gian này, các điểm dữ liệu có tính chất linearly separable, hoặc gần linearly separable. Sau đó sử dụng soft (hoặc hard) margin SVM để tìm siêu phẳng phân chia như bài trước.



Hình 1: Ảnh minh họa ý tưởng của kernel SVM

2 Kernel trick

Giả sử ta đã tìm được một hàm số $f(x)$ sao cho sau khi được biến đổi sang không gian mới, mỗi điểm dữ liệu x sẽ trở thành $f(x)$ và các điểm dữ liệu mới trở nên (gần) linearly separable. Khi đó nhãn của dữ liệu mới sẽ có dạng:

$$sgn(w^T f(x) + b)$$

Vấn đề là khi nâng lên số chiều càng lớn, số phép tính toán càng trở nên phức tạp, tiêu tốn nhiều tài nguyên. Để có thể chuyển hết tất cả các điểm dữ liệu ban đầu thành các điểm dữ liệu mới ta cần mất nhiều công sức, thậm chí với số chiều không gian mới gần như là vô hạn, việc tính toán càng trở nên bất khả thi

Một kỹ thuật giải quyết là thay vì phải ánh xạ toàn bộ điểm dữ liệu lên một không gian mới, ta chỉ cần biết được quan hệ giữa hai điểm dữ liệu bất kỳ trong không gian mới. Kỹ thuật này được xây dựng dựa trên quan sát về bài toán đối ngẫu của SVM.

Theo quan sát rút ra từ bài toán đối ngẫu, giả sử tồn tại một hàm $f(x)$ có thể ánh xạ các điểm dữ liệu lên một không gian mới có tính chất (gần) linearly separable, ta không cần trực tiếp tính $f(x)$ đối với mọi điểm dữ liệu x. Ta chỉ cần tính $f(x)^T f(z)$ với x và z là hai điểm dữ liệu bất kỳ. Kỹ thuật này được gọi là kernel trick

Hàm $k(x, z) = f(x)^T f(z)$ được gọi là hàm kernel

3 Một số tính chất hàm kernel

1. Hàm kernel là một hàm số liên tục
2. Hàm kernel có tính chất đối xứng: k(a,b) = k(b,a). Bởi vì hàm kernel có giá trị là kết quả của một phép tích vô hướng. Phép tích vô hướng của hai vecto có tính chất đối xứng
3. Về lý thuyết, hàm kernel phải thỏa mãn điều kiện Mercer Link.

4 Một số hàm kernel thông dụng trong sklearn

Theo document của sklearn. Biến kernel có các kiểu 'linear', 'poly', 'rbf', 'sigmoid', 'precomputed'

4.1 linear

Tương đương với Hard/Soft margin SVM. Hàm kernel có dạng:

$$k(a, b) = a^T b$$

4.2 Đa thức - poly

Hàm kernel có dạng:

$$k(a,b) = (r + \gamma a^T b)^d$$

Với r, γ, d là hằng số.

Cách hoạt động Xét trường hợp r=1/2, d=2, γ=1 Với a,b là các dữ liệu nằm trong không gian một chiều (nằm trên một trục số) Như vậy :

$$(ab + \frac{1}{2})^2 = a^2b^2 + ab + \frac{1}{4} = (a, a^2, \frac{1}{2}).(b, b^2, \frac{1}{2})$$

Kết quả của hàm kernel chính là một tích vô hướng của hai vector nằm trong không gian ba chiều với một điểm chỉ phụ thuộc vào a, và một điểm chỉ phụ thuộc vào b. Như vậy, tồn tại hàm $f(a)$ ánh xạ điểm a từ không gian một chiều lên một điểm trong không gian ba chiều (a,a^2,$\frac{1}{2}$). Tương tự đối với điểm b. Xem thêm tại video Youtube

4.3 Radial Basic Function - rbf

Hàm kernel có dạng:

$$k(a,b) = e^{-\gamma||a-b||_2^2}, \gamma > 0$$

Cách hoạt động: khai triển toán học ta có:

$$e^{-\gamma(a-b)^2} = e^{-\gamma(a^2+b^2)} * e^{\gamma 2ab}$$

Để đơn giản tính toán, xét trường hợp $\gamma = 1/2$ Dựa vào công thức khai triển Taylor ta phân tích:

$$e^{ab} = 1 + \frac{1}{1!}ab + \frac{1}{2!}(ab)^2 + \frac{1}{3!}(ab)^3 + ... + \frac{1}{\infty!}(ab)^\infty$$

$$e^{ab} = (1, \sqrt{\frac{1}{1!}}a, \sqrt{\frac{1}{2!}}a^2, \sqrt{\frac{1}{3!}}a^3, ..., \sqrt{\frac{1}{\infty!}}a^\infty).(1, \sqrt{\frac{1}{1!}}b, \sqrt{\frac{1}{2!}}b^2, \sqrt{\frac{1}{3!}}b^3, ..., \sqrt{\frac{1}{\infty!}}b^\infty)$$

$$e^{-\frac{1}{2}(a-b)^2} = e^{-\frac{1}{2}a^2}(1, \sqrt{\frac{1}{1!}}a, \sqrt{\frac{1}{2!}}a^2, \sqrt{\frac{1}{3!}}a^3, ..., \sqrt{\frac{1}{\infty!}}a^\infty).e^{-\frac{1}{2}b^2}(1, \sqrt{\frac{1}{1!}}a, \sqrt{\frac{1}{2!}}a^2, \sqrt{\frac{1}{3!}}a^3, ..., \sqrt{\frac{1}{\infty!}}a^\infty)$$

Kết quả của hàm kernel là phép tích vô hướng của hai vector có vô hạn chiều. Như vậy tồn tại một hàm $f(a)$ ánh xạ điểm a từ không gian một chiều lên một điểm trong không gian vô hạn chiều. Đây là hàm kernel được sử dụng phổ biến nhất trong hầu hết các trường hợp của SVM vì số chiều luôn là vô hạn, số với số chiều ban đầu là hữu hạn, nên khả năng để tìm được siêu phẳng phân chia sẽ cao hơn Xem thêm tại video Youtube

4.4 Sigmoid

Hàm kernel có dạng:

$$k(x, z) = tanh(\gamma x^T z + r)$$

Với r, γ là hằng số.

4.5 Bảng tóm tắt

Tên	Công thức	`kernel`	Thiết lập hệ số
linear	$\mathbf{x}^T\mathbf{z}$	`'linear'`	không có hệ số
polynomial	$(r + \gamma\mathbf{x}^T\mathbf{z})^d$	`'poly'`	d: `degree`, γ: `gamma`, r: `coef0`
sigmoid	$\tanh(\gamma\mathbf{x}^T\mathbf{z} + r)$	`'sigmoid'`	γ: `gamma`, r: `coef0`
rbf	$\exp(-\gamma\|\|\mathbf{x} - \mathbf{z}\|\|_2^2)$	`'rbf'`	$\gamma > 0$: `gamma`

Hình 2: Bảng tóm tắt các hàm kernel thông dụng trong sklearn