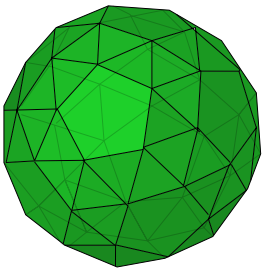
Projects in Mathematics and Applications

SUPPORT VECTOR MACHINE

Phạm Quốc Việt *, Nguyễn Hà Minh †
Nguyễn Đình Bảo Ngọc ‡, Nguyễn Huy Anh §

Ngày 27 tháng 8 năm 2018



* Trường THPT Chuyên Amsterdam
† Trường THPT Nguyễn Bỉnh Khiêm
‡ Trường Phổ Thông Năng Khiếu - ĐHQG TP.HCM
§ Campbell County Comprehensive High School

Lời cảm ơn

Chúng em xin chân thành cảm ơn Ban tổ chức PiMA vì đã mang đến một trại hè Toán học vô cùng thú vị và đầy ý nghĩa cho chúng em cũng như là các bạn trại sinh còn lại. Không những chúng em được tìm hiểu sâu và rộng hơn về Toán, chúng em còn được làm quen và học hỏi từ các bạn có cùng chung đam mê dành cho môn Toán từ khắp nơi trên mọi miền đất nước. Tham gia PiMA, chúng em được giới thiệu với các mảng Toán ở Đại học, được học lập trình, nghiên cứu và viết báo cáo, và trên hơn cả là rèn luyện kĩ năng làm việc nhóm. Chúng em rất trân trọng trải nghiệm 10 ngày vừa qua và hy vọng rằng PiMA sẽ tiếp tục truyền cảm hứng và nuôi giữ ngọn lửa đam mê cho các bạn học sinh tham gia như PiMA đã thực hiện thành công trong 3 năm vừa qua.

Chúng em cũng xin gửi lời cảm ơn đến Ban Giám Đốc Đại Học Khoa Học Tự Nhiên TP.HCM và các nhà tài trợ đã tạo điều kiện để trại hè PiMA 2018 được diễn ra thành công.

Chúng em xin cảm ơn đến các thầy diễn giả: GS.TS Hồ Tú Bảo, PGS.TS Đinh Điền, TS. Nguyễn Đức Hoàng Hạ, TS. Nguyễn Minh Triết, PGS.TS Lý Quốc Ngọc. Qua những lời chia sẻ nhiệt huyết từ các thầy, chúng em và các bạn trại sinh còn lại đã được mở mang và tiếp xúc với nhiều khía cạnh khác nhau trong môn Toán, đặc biệt là đối với mảng máy học.

Chúng em xin chân thành gửi lời cảm ơn đến sự hỗ trợ của anh Vũ Lê Thế Anh, là người đã trực tiếp hướng dẫn tụi em trong quá trình tìm hiểu và chuẩn bị bài báo cáo này, cùng sự góp ý của anh Dương Đình Trọng cho bài báo cáo này được hoàn hảo hơn.

Và xin cảm ơn đến các bạn trại sinh còn lại vì đã đồng hành cùng chúng mình trong 10 ngày vừa qua.

Tóm tắt nội dung

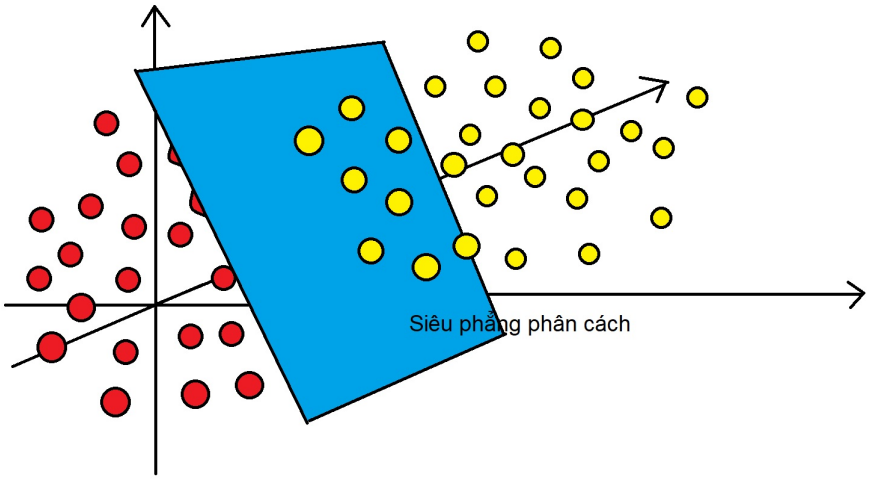
Trong khuôn khổ bài báo cáo này, chúng em sẽ trình bày về thuật toán Support Vector Machine và ứng dụng của nó trong các bài toán, đặc biệt là trong hệ thống phân lớp nhị phân. Bài báo cáo được xây dựng trên cơ sở toán học, Soft-margin và Hard-margin SVM, cách tối ưu hệ thống, và mô phỏng thuật toán này trên mô hình dự đoán giao thức tín dụng.

Mục lục

1 Tổng quan 1
2 Một số định nghĩa và tính chất 1
3 Phát biểu bài toán 3
4 Các phương pháp giải bài toán SVM 4
4.1 Tập dữ liệu có thể phân cách tuyến tính: Hard-margin SVM 4
4.2 Trường hợp tập dữ liệu không thể phân cách tuyến tính: Soft-margin SVM và Kernel Trick 6
5 Áp dụng vào mô hình thực tế 12
5.1 Mô tả dữ liệu 12
5.2 Các thuật ngữ quan trọng 12
5.3 Tham số C và kernel trong thuật toán SVM 13
5.4 Đánh giá mô hình thực tế 14
6 Phụ lục 15

1 Tổng quan

Thuật toán Support Vector Machine chính thức được giới thiệu bởi Bernhard E. Boser, Isabelle M. Guyon và Vladimir N. Vapnik vào năm 1992. Thuật toán Support Vector Machine bắt đầu trở nên nổi tiếng khi được áp dụng vào kĩ thuật nhận dạng con số viết bằng tay. Ngày nay, Support Vector Machine có vai trò quan trọng trong máy học và được biết đến như là một ví dụ điển hình của phương pháp kernel.
Support Vector Machine giúp hỗ trợ giải quyết một trong những nhiệm vụ phổ biến của máy học: phân cách dữ liệu. Cụ thể hơn, Support Vector Machine giúp ta phân loại một điểm dữ liệu mới vào một trong hai lớp dữ liệu đã được hình thành từ những điểm dữ liệu cũ bằng cách gắn với mỗi điểm dữ liệu đó một vector p chiều và đi tìm một siêu phẳng $p-1$ chiều để phân cách các vector trên sao cho khoảng cách từ lề đến siêu phẳng đó là lớn nhất.
Hình sau là một ví dụ của siêu phẳng phân cách trong không gian ba chiều.



2 Một số định nghĩa và tính chất

Định nghĩa 2.1. Một tập hợp A được gọi là lồi nếu với hai phần tử x_i và x_j bất kì thuộc A, phần tử $x_k = "x_i + (1-")x_j$ cũng thuộc $A, \forall " \in [0,1]$.

Tính chất 2.2.

- Một siêu phẳng là một tập lồi.
- Giao của các tập lồi là một tập lồi.

Định nghĩa 2.3. Một hàm số $f : \mathbf{R}^n \to \mathbf{R}$ là hàm lồi nếu tập xác định (dom) D là một tập lồi và:

$$f("x_i + (1-")x_j) \leq "f(x_i) + (1-")f(x_j), \forall x_i, x_j \in D, 0 \leq " \leq 1.$$

Tính chất 2.4. Hàm số $f : \mathbf{R}^n \to \mathbf{R}$ là một hàm lõm nếu $-f$ là một hàm lồi.

Tính chất 2.5. Hàm norm là một hàm lồi.

Định nghĩa 2.6. Một hàm số $f : \mathbf{R}^n \to \mathbf{R}$ là hàm lồi chặt nếu f là hàm lồi và dấu bằng không xảy ra $\forall x_i \neq x_j$.

Định nghĩa 2.7. Một bài toán tối ưu lồi là một bài toán tối ưu có dạng:

$$x = \underset{x}{\operatorname{argmin}} f_0, \tag{1}$$

với

$$\begin{aligned} & f_i(x) \leq 0 && \forall i = 1, 2, :::, n, \\ & h_j(x) = a_j^T x - b_j = 0 && \forall j = 1, 2, :::, m, \\ & x \in \mathcal{D} = (\bigcap_{i=0}^{n} \mathrm{dom} f_i) \cap (\bigcap_{j=1}^{n} \mathrm{dom} h_j). \end{aligned}$$

Trong đó, f_i là các hàm lồi $\forall i = 0, 1, ..., n$.

Định nghĩa 2.8. Gọi x_0 là nghiệm toàn cục của bài toán tối ưu lồi (1) nếu:

$$x_0 = \underset{x}{\operatorname{argmin}} f_0,$$

$$\text{và} \quad f_i(x_0) \leq 0 \quad \forall i = 1, 2, ...n, \quad h_j(x_0) = 0, \quad \forall j = 1, 2, :::, m.$$

Định nghĩa 2.9. Gọi x_0 là nghiệm cục bộ lân cận ε của bài toán tối ưu lồi (1) nếu:

$$x_0 = \underset{x}{\operatorname{argmin}} f_0,$$

$$\text{và} \quad f_i(x_0) \leq 0 \quad \forall i = 1, 2, ...n, \quad h_j(x_0) = 0 \quad \forall j = 1, ..., m, \ \|x - x_0\|_2 \leq \varepsilon.$$

Tính chất 2.10. Nghiệm cục bộ lân cận của một bài toán tối ưu lồi cũng chính là nghiệm toàn cục của bài toán đó.

Chứng minh bằng phản chứng: Giả sử x_0 là một nghiệm cục bộ lân cận ε nhưng không là cực trị toàn cục. Gọi y_0 là cực trị toàn cục, tức là $f(y_0) \leq f(x_0)$, $\|y_0 - x_0\|_2 > \varepsilon$.
Mặt khác, luôn tồn tại z_0 sao cho:

$$\begin{cases} z_0 = \theta y_0 + (1 - \theta) x_0 \text{ với } \theta \in [0, 1] \text{ đủ nhỏ}, \\ \|z_0 - x_0\|_2 < \varepsilon, \\ z_0 \text{ thỏa mãn các hàm điều kiện ràng buộc.} \end{cases}$$

Khi đó:

$$\begin{aligned} f_0(z_0) &= f_0(\theta y_0 + (1 - \theta) x_0) \\ &\leq \theta f_0(y_0) + (1 - \theta) f_0(x_0) \\ &< \theta f_0(x_0) + (1 - \theta) f_0(x_0) \\ &= f_0(x_0). \end{aligned}$$

Điều này mâu thuẫn do x_0 là nghiệm cục bộ lân cận ε.

Định nghĩa 2.11. Hàm Lagrangian của bài toán tối ưu lồi (1) là hàm:

$$\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{n} \lambda_i f_i(x) + \sum_{j=1}^{m} \nu_j h_j(x),$$

với,

$$\begin{aligned} & x \in \mathcal{D}, \\ & \lambda = [\lambda_1, \lambda_2, ..., \lambda_n]^T, \quad \lambda \succeq 0, \\ & \nu = [\nu_1, \nu_2, ..., \nu_n]^T. \end{aligned}$$

Định nghĩa 2.12. Hàm đối ngẫu của bài toán tối ưu lồi (1) với $(\ ,\ )$ là hàm:

$$g(\ ,\ ) = \inf_{x \in D} \mathcal{L}(x,\ ,\ ).$$

Nhận xét 2.13. *Hàm đối ngẫu trên là một hàm lõm.*

Tính chất 2.14. $g(\ ,\ ) \leq f_0(x_0)$ với x_0 là nghiệm toàn cục của bài toán tối ưu lồi (1).

Định nghĩa 2.15. Bài toán đối ngẫu Lagrange của bài toán (1) là:

$$(\ _0,\ _0) = \operatorname*{argmax}_{,} g(\ ,\ ).$$

Định nghĩa 2.16. Đối ngẫu chặt xảy ra khi $g(\ _0,\ _0) = f_0(x_0)$.

Nhận xét 2.17. *Khi đối ngẫu chặt xảy ra, nghiệm của bài toán đối ngẫu chính là nghiệm của bài toán tối ưu gốc.*

Định lý 2.18 (Tiêu chuẩn Slater)**.** *Khi bài toán tối ưu gốc là lồi, và tồn tại* $x \in \mathcal{D}$ *sao cho:* $f_i(x\ ) < 0\ \ \forall i = 1, 2, ..., n,\ \ h_j(x\ ) = 0\ \ \forall j = 1, 2, ..., m$ *thì đối ngẫu chặt xảy ra.*

Định lý 2.19 (Hệ điều kiện KKT)**.** *Điều kiện cần và đủ KKT cho nghiệm* $(x_0,\ _0,\ _0)$ *bài toán đối ngẫu là:*

$$\begin{aligned}
f_i(x_0) &\leq 0, \forall i = 1, 2, ..., n, \\
h_j(x_0) &= 0, \forall j = 1, 2, ..., m, \\
{}^0_i &\geq 0, \forall i = 1, 2, ..., n, \\
{}^0_i f_i(x_0) &= 0, \forall i = 1, 2, ..., n, \\
\frac{@f_0}{@x_0} + \sum_{i=1}^{n} {}^0_i \frac{@f_i}{@x_0} + \sum_{j=1}^{m} v_j^0 \frac{@h_j}{@x_0} &= 0.
\end{aligned}$$

3 Phát biểu bài toán

Xét tập dữ liệu $D = \{(x_i, y_i); x_i \in \mathrm{R}^m; y_i \in \{1, -1\} \mid i = 1, 2, ..., n\}$ với n là số cặp dữ liệu và m là số chiều của dữ liệu (số lượng đặc trưng). Trong đó:

$$y_i = \begin{array}{ll} 1 & \text{nếu cặp dữ liệu thứ } i \text{ thuộc nhóm 1,} \\ -1 & \text{nếu cặp dữ liệu thứ } i \text{ thuộc nhóm 2.} \end{array}$$

Ta cần xác định siêu phẳng $w^T x + b = 0$ phân chia 2 nhóm sao cho lề là lớn nhất, với định nghĩa lề trong bài toán SVM là giá trị của:

$$\min_{i=1,2,...n} \frac{|w^T x_i + b|}{\|w\|_2}$$

Bài toán trên tương đương với bài toán tối ưu sau:

$$(w, b) = \operatorname*{argmax}_{w,b} \ \min_i \frac{|w^T x_i + b|}{\|w\|_2} \tag{2}$$

Nhận xét 3.1. *Ta luôn chọn được* (w, b) *sao cho* y_i *và* $(w^T x_i + b)$ *cùng dấu vì nếu không chỉ cần đổi dấu của* $y_i, \forall i = 1, 2, ..., n$.

Khi đó, (2) tương đương với tìm:

$$(w, b) = \operatorname*{argmax}_{w,b} \ \min_i \frac{y_i(w^T x_i + b)}{\|w\|_2} \tag{3}$$

4 C¡c ph÷ìng ph¡p gi£i b i to¡n SVM

4.1 T³p dú li»u câ thº ph¥n c¡ch tuy¸n t½nh: Hard-margin SVM

4.1.1 B i to¡n t÷ìng ÷ìng

Ta câ: luæn tçn t¤i si¶u ph¯ngw^{T}x + b = 0 ph¥n chia ho n to n hai nhâm. Gåi(x_0, y_0) l :

$$(x_0, y_0) = \underset{(x_i, y_i)}{\text{argmin}} \frac{y_i(w^T x_i + b)}{\|w\|_2}$$

Nh³n x²t 4.1. N¸u thay $(w, b) = (\|w\|, \|b\|)$ th¼$\frac{y_i(w^T x_i + b)}{\|w\|_2}$ khæng êi,8i = 1, 2, ..., n. Tø â, khæng m§t t½nh têng qu¡t, gi£ sû$y_0(w^T x_0 + b) = 1$.

Khi â (3) trð th nh t¼m:

$$(w, b) = \underset{w,b}{\text{argmax}} \frac{1}{\|w\|_2}, \tag{4}$$

$$\text{vîi } y_i(w^T x_i) \quad y_0(w^T x_0 + b) = 1, \ 8i = 1, 2, ..., n.$$

(4) t÷ìng ÷ìng vîi t¼m:

$$(w, b) = \underset{w,b}{\text{argmax}} \frac{1}{2}\|w\|_2^2, \tag{5}$$

$$\text{vîi } 1 \quad y_i(w^T x_i) \quad 0, \ 8i = 1, 2, ..., n.$$

Nh÷ vªy, sau khi x¡c ành ÷ñc si¶u ph¯ngw^{T}x + b = 0, nhâm cõax$_i$ 2 R^m b§t k¼ tròng vîi d§u cõaw^{T}x$_i$ + b = 0.

4.1.2 B i to¡n èi ng¨u Lagrange

Ki¬m tra ti¶u chu©n Slater : V¼ hai lîp dú li»u l câ thº ph¥n c¡ch tuy¸n t½nh, tùc b i to¡n câ nghi»m, n¶n luæn tçn t¤i c°p(w_o, b_o) sao cho:

$$\begin{aligned} 1 \quad & y_i(w_o^T x_i + b_o) \quad 0, 8i = 1, 2, ..., n, \\ , \quad 2 \quad & y_i(2w_o^T x_i + 2b_o) \quad 0, 8i = 1, 2, ..., n. \end{aligned}$$

Chån$w_1 = 2w_o$ v $b_1 = 2b_o$, suy ra:

$$1 \quad y_i(w_1^T x_i + b_1) \quad 1 < 0, 8i = 1, 2, ..., n$$

Vªy ti¶u chu©n Slater thäa m¢n.

H m Lagrangian cõa b i to¡n (5) l :

$$L(w, b, \ ) = \frac{1}{2}\|w\|_2^2 + \sum_{i=1}^{n} \ _i(1 \quad y_i(w^T x_i + b))$$

vîi = [$_1$, $_2$, ..., $_n$]^T v $_i$ 0, 8i = 1, 2, ..., n.

H m èi ng¨u Lagrange l

$$g(\ ) = \min_{w,b} L(w, b, \ ) \text{ vîi } \ 0.$$

X²t $(w_0, b_0, \ _0) = \arg\min_{w,b} L(w, b, \ )$. Khi â:

$$\frac{@L}{@w_0} = w_0 \ \sum_{i=1}^{n} \ _i y_i x_i = 0) \ w_0 = \sum_{i=1}^{n} \ _i y_i x_i .$$
$$\frac{@L}{@b} = \sum_{i=1}^{n} \ _i y_i = 0.$$

Thay v o biºu thùc cõa $L(w, b, \ )$, ta ÷ñc:

$$g(\ ) = \sum_{n=1}^{N} \ _n \ \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \ _n \ _m y_n y_m x_n^T x_m \qquad (6)$$

Tø â, k¸t hñp h m èi ng¨u Lagrange v c¡c i·u ki»n r ng buëc cõa , ta s³ thu ÷ñc b i to¡n èi ng¨u Lagrange cõa b i to¡n (26.3) câ d¤ng

$$= \arg\max g(\ )$$

vîi:

$$0$$
$$\sum_{n=1}^{N} \ _n y_n = 0$$

V¼ ¥y l mët b i to¡n lçi v èi ng¨u ch°t x£y ra, h» i·u ki»n KKT l i·u ki»n c¦n v õ cho bë $(w_0, b_0, \ _0)$:

$$\begin{aligned} 1 \ \ y_i(w_0^T x_i + b_0) \ \ & 0, 8i = 1, 2, ..., n \\ \ _i \ \ & 0, 8i = 1, 2, ..., n \\ \ _i(1 \ \ y_i(w_0^T x_i + b_0)) \ = \ & 0 \ 8i = 1, 2, ..., n \qquad (7) \\ w_0 \ = \ & \sum_{i=1}^{n} \ _i y_i x_i \\ \sum_{n=1}^{N} \ _n y_n \ = \ & 0 \qquad (8) \end{aligned}$$

Tø i·u ki»n (7) ta th§y ho°c $\ _i = 0$ ho°c $y_i(w_0^T x_i + b_0) = 1$. Khi â, hai ÷íng th¯ng $w_0^T x_i + b =$ 1 chia t¡ch ho n to n hai lîp dú li»u.
Tø ¥y ta câ ành ngh¾a sau:

ành ngh¾a 4.2. Hai ÷íng bi¶n cõa b i to¡n SVM l hai ÷íng th¬ng $w_0^T x_i + b =$ 1.

Nhªn x²t 4.3. Sè iºm n¬m tr¶n c£ hai si¶u ph¯ng l nhä hìn r§t nhi·u so vîi têng sæ iºm cõa t¤p hu§n luy»n.

°t S = f n : n ∈ 0g v N_S l sè ph¦n tû cõa tªp S
Theo (8), ta t½nhw:

$$w = \sum_{m \in S} {}_m y_m x_m,$$

$$b = \frac{1}{N_S} \sum_{n \in S} \left(y_n - \sum_{m \in S} {}_m y_m x_m^T x_n \right)$$

V· vi»c t½nh $_0$ th¸ n o, mæ h¼nh s³ ÷ñc düng b¬ng ngæn ngú lªp tr¼nh Python.

4.2 Tr÷íng hñp tªp dú li»u khæng thº ph¥n c¡ch tuy¸n t½nh: Soft-margin SVM v Kernel Trick

4.2.1 B i to¡n t÷ìng ÷ìng: Soft-margin SVM

Trong thüc t¸, tªp dú li»u hu§n luy»n ta câ g¦n nh÷ khæng thºph¥n c¡ch tuy¸n t½nhKhi â, n¸u dòng thuªt to¡n Hard-margin SVM s³ d¨n ¸n vi»c k¸t qu£ l· r§t nhä, thªm ch½ khæng t¼m ÷ñc si¶u ph¯ng ph¥n c¡ch.
Vîi suy ngh¾ nh÷ vªy, thuªt to¡nSoft-margin SVM ra íi. Lóc n y, º thu ÷ñc l· lîn hìn, ta ch§p nhªn mët v i iºm dú li»u bàsai rìi v o b¶n trong ho°c sang h¯n ph½a b¶n kia cõa D¹ th§y sü ch§p nhªnsai c¦n ÷ñc h¤n ch¸, n¸u khæng, ta câ thº t¤o ra mët l· r§t lîn b¬ng c¡ch ch§p nhªn h¦u h¸t c¡c iºm bà sai.
Tø â, h m möc ti¶u n¶n l mët sü k¸t hñp º tèi a l· công nh÷ tèi thiºu sü sai.
Nh÷ ¢ nâi trong hard-margin SVM, vi»c tèi a l· câ thº ÷a v· vi»c tèi thiºu $\|w\|_2^2$. º t½nh sü sai, ành ngh¾a th¶m mët bi¸n o ësai vîi méi iºm x_i trong tªp dú li»u hu§n luy»n.

ành ngh¾a 4.4. $_i$ l bi¸n sai cõa bi¸n x_i n¸u nh÷:

$$_i = \begin{cases} 0 & n¸u\ iºm\ dú\ li»u\ x_i\ n¬m\ trong\ bi¶n\ cõa\ nâ, \\ |w^T x_i + b - y_i| & n¸u\ iºm\ dú\ li»u\ x_i\ khæng\ n¬m\ trong\ bi¶n\ cõa\ nâ \end{cases}$$

Nhªn x²t 4.5. Ta câ thº ành ngh¾a bi¸n sai nh÷ tr¶n do¸ ch¿ c¦n t¿ l» vîi kho£ng c¡ch tø x_i ¸n si¶u ph¯ng

Vîi soft-margin SVM, h m möc ti¶u s³ câ th¶m mët sè h¤ng núa gióp tèi thiºu têng ësai.
Khi â, h m möc ti¶u mîi l :

$$\frac{1}{2}\|w\|_2^2 + C \sum_{n=1}^{N} {}_n,$$

vîi C l mët h¬ng sè d÷ìng.
Nhªn th§y $|y_i| = 1, 8i = 1, 2, ..., n$, ta câ:

$$\begin{aligned} {}_i &= |w^T x_i + b - y_i| \\ &= |w^T x_i + b - y_i| \cdot |y_i| \\ &= |y_i (w^T x_i + b) - 1| \\ &\geq 1 - y_i (w^T x_i + b) \end{aligned}$$

Tø â ta thu ÷ñc i·u ki»n r ng buëc mîi sau, $8i = 1, 2, ..., n$:

$$1 - {}_i - y_i (w^T x_i + b) \leq 0,$$

v $_i \geq 0$.

Nh÷ vªy, b i to¡n tèi ÷u gèc cho Soft-margin SVM ph¡t biºu nh÷ sau, 8i = 1, 2, ..., n:

$$(w_0, b_0, \quad _0) = \underset{w,b,}{\operatorname{argmin}} \frac{1}{2} kwk_2^2 + C \sum_{i=1}^{n} \quad _i,$$
$$\text{vîi: } 1 \quad _i \quad y_i(w^T x_i + b) \quad 0, \tag{9}$$
$$_n \quad 0.$$

4.2.2 H÷îng gi£i quy¸t thù nh§t: B i to¡n èi ng¨u Lagrange cõa Soft-margin SVM

T÷ìng tü b i to¡n Hard-margin SVM, ta quan t¥m ¸n b i to¡n èi ng¨u cõa nâ.
Ki ºm tra ti¶u chu©n Slater : 8i = 1, 2, ..., n v 8(w, b), luæn tçn t¤i c¡c sè d÷ìng $_i$, õ lîn sao cho:

$$y_i(w^T x_i + b) + \quad _i > 1.$$

V¼ vªy, b i to¡n tèi ÷u n y thäa m¢n ti¶u chu©n Slater.
H m Lagrangian cõa b i to¡n Soft-margin SVM l :

$$L(w, b, \quad , \quad) = \frac{1}{2} kwk_2^2 + C \sum_{i=1}^{n} \quad _i + \sum_{i=1}^{n} \quad _i(1 \quad _i \quad y_i(w^T x_i + b)) \quad \sum_{i=1}^{n} \quad _i \quad _i, \tag{10}$$
$$\text{vîi} \quad = [\quad _1, \quad _2, ..., \quad _n]^T \quad 0 \text{ v} \quad = [\quad _1, \quad _2, ..., \quad _n]^T \quad 0.$$

B i to¡n èi ng¨u cõa Soft-margin SVM l :

$$g(\quad , \quad) = \min_{w,b,} L(w, b, \quad , \quad , \quad)$$

º gi£i b i to¡n tr¶n, vîi méi c°p (,), x²t bë $(w_0, b, \quad _0)$ thäa m¢n:

$$\frac{@}{@w_0} = 0, \quad w_0 = \sum_{i=1}^{n} \quad _i y_i x_i \tag{11}$$
$$\frac{@}{@b} = 0, \quad \sum_{i=1}^{n} \quad _i y_i = 0 \tag{12}$$
$$\frac{@}{@ \quad _0} = 0, \quad \quad _i = C \quad \quad _i \; 8i = 1, 2, ..., n \tag{13}$$

Tø (13) suy ra ch¿ c¦n quan t¥m tîi c¡c c°p (,) thäa m¢n $_i = C \quad _i$.
Tø ¥y ta công suy ra $0 \quad _i, \quad _i \quad C, 8i = 1, 2, ..., n$.
Thay v o biºu thùc Lagrangian, ta thu ÷ñc h m möc ti¶u cõa b i to¡n èi ng¨u:

$$g(\quad , \quad) = \sum_{i=1}^{n} \quad _i \quad \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \quad _i \quad _j y_i y_j x_i^T x_j. \tag{14}$$

Nhªn x²t 4.6. H m n y khæng phö thuëc v o v tø (13) suy ra $0 \quad _i \quad C, 8i = 1, 2, ..., n$

Khi â, b i to¡n èi ng¨u trð th nh t¼m:

$$_0 = \operatorname{argmax} g(\quad)$$

$$\text{vîi: } \sum_{i=1}^{n} \quad _i y_i = 0, \tag{15}$$
$$0 \quad _i \quad C, 8i = 1, 2, ..., n. \tag{16}$$

H» i·u ki»n KKT cõa b i to¡n tèi ÷u Soft-margin SVM l , 8i = 1, 2, ..., n:

$$1 \quad {}^{0}_{i} \quad y_i(w_0^T x_i + b_0) \quad 0, \tag{17}$$
$$ {}^{0}_{i} \quad 0, \tag{18}$$
$$ {}^{0}_{i} \quad 0, \tag{19}$$
$$ {}^{0}_{i} \quad 0, \tag{20}$$
$${}^{0}_{i}(1 \quad {}^{0}_{i} \quad y_i(w_0^T x_i + b_0)) = 0, \tag{21}$$
$${}^{0}_{i} \; {}^{0}_{i} = 0, \tag{22}$$

Ngo i ra cán câ c¡c i·u ki»n (11), (12), (13).
Khi ${}^{0}_{i} > 0$, theo (21):

$$y_i(w_0^T x_i + b_0) = 1 \quad {}^{0}_{i}. \tag{23}$$

N¸u $0 < {}^{0}_{i} < C$, theo (13), ${}^{0}_{i} = C \quad {}^{0}_{i} > 0$; ngo i ra theo (22) ta thu ÷ñc ${}^{0}_{i} = 0$. Ti¸p töc theo (23), suy ra $y_i(w_0^T x_i + b_0) = 1$, hay:

$$w_0^T x_i + b_0 = y_i, 8i : 0 < {}^{0}_{i} < C.$$

Tâm l¤i, n¸u $0 < {}^{0}_{i} < C$, c¡c i°m x_i n¬m ch½nh x¡c tr¶n c¡c l·. Khi â,(w_0, b_0) ÷ñc t½nh theo:

$$\begin{aligned} w_0 &= \sum_{i=1}^{n} {}^{0}_{i} y_i x_i, \\ b_0 &= \frac{1}{N_S} \sum_{i2S} (y_i \quad w_0^T x_i) \end{aligned} \tag{24}$$

vîi $S = f m : 0 < {}_m < C g$ v N_S l sè ph¦n tû cõa S.
Khi â, nhâm cõa x ch½nh l d§u cõa:

$$w_0^T x \; + b_0 = \sum_{i2M} {}^{0}_{i} y_i x_i^T x \; + \frac{1}{N_S} \sum_{j2S} y_j \quad \sum_{i2M} {}^{0}_{j} y_j x_j^T x \; ! \tag{25}$$

Vîi $M = f n : 0 < {}_n g$ v N_M l sè ph¦n tû cõa M .

4.2.3 H÷îng gi£i quy¸t thù 2: B i to¡n tèi ÷u khæng r ng buëc cho soft-margin SVM

B i to¡n tèi ÷u khæng r ng buëc t÷ìng ÷ìng: X²t i·u ki»n r ng buëc ¦u ti¶n:

$$1 \quad {}_i \quad y_i(w^T x_i + b) \quad 0, \quad {}_i \quad 1 \quad y_i(w^T x_i + b)$$

M°t kh¡c, ${}_i \quad 0$ n¶n ta câ b i to¡n r ng buëc t÷ìng ÷ìng vîi b i to¡n (9) nh÷ sau:

$$\begin{aligned} (w_0, b_0, \; {}_0) &= \arg\min_{w,b,} \frac{1}{2}\|w\|_2^2 + C \sum_{i=1}^{n} {}_i \\ \text{thäa m¢n: } {}_i \quad &\max(0, 1 \quad y_i(w^T x_i + b)), 8i = 1, 2, ..., n \end{aligned} \tag{26}$$

M»nh · 4.7. N¸u $(w_0, b_0, \; {}_0)$ l nghi»m cõa b i to¡n tèi ÷u (26), th¼:

$${}^{0}_{i} = \max(0, 1 \quad y_i(w_0^T x_i + b_0)), 8i = 1, 2, ..., n \tag{27}$$

Chùng minh: Gi£ sû tçn t¤i i sao cho

$$_i > \max(0, 1 \quad y_i(w_0^T x_i + b_0)),$$

Khi â, chån $_i^0 = \max(0, 1 \quad y_i(w_0^T x_i + b_0))$, ta ÷ñc mët gi¡ trà nhä hìn cõa h m möc ti¶u, trong khi t§t c£ c¡c r ng buëc v¨n ÷ñc thäa m¢n (m¥u thu¨n).
V¾y m»nh · ÷ñc chùng minh.
Tø â, b i to¡n tèi ÷u trð th nh:

$$(w_0, b_0, \quad _0) = \arg\min_{w,b,} \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{n} \max(0, 1 \quad y_i(w^T x_i + b)) \qquad (28)$$
$$\text{thäa m¢n: } _i = \max(0, 1 \quad y_i(w^T x_i + b)), 8i = 1, 2, ..., n,$$

hay:

$$(w, b) = \arg\min_{w,b} \left(\frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{n} \max(0, 1 \quad y_i(w^T x_i + b)) \right) \qquad (29)$$

Nh¾n x²t 4.8. B i to¡n n y câ thº gi£i ÷ñc b¬ng c¡c ph÷ìng ph¡p gradient descent.

ành ngh¾a 4.9. H m Hinge loss câ d¤ng:

$$J_i(w, b) = \max(0, 1 \quad y_i(w^T x_i + b))$$

X¥y düng h m m§t m¡t

ành ngh¾a 4.10. H m Hinge loss trung b¼nh câ cæng thùc:

$$J(w, b) = \frac{1}{n}\sum_{i=1}^{n} J_i = \frac{1}{n}\sum_{i=1}^{n} \max(0, 1 \quad y_i(w^T x_i + b))$$

Nh¾n x²t 4.11. X²t (w_0, b_0) l nghi»m b i to¡n Khi t¤p dú li»u câ thº ph¥n c¡ch tuy¸n t½nh, gi¡ trà tèi ÷u cõa $J(w, b) = 0$. Khi â d¹ th§y $8c > 1$ l h¬ng sè, (cw_0, cb_0) công l nghi»m b i to¡n d¨n tîi nghi»m câ thº lîn tòy þ.

V¼ vªy, ta câ ành ngh¾a sau:

ành ngh¾a 4.12. H m m§t m¡t têng cëng câ cæng thùc l :

$$J(w, b) = L(w, b) + \quad R(w, b),$$

vîi > 0, $R(w, b) = \frac{1}{n}\frac{}{2}\|w\|_2^2$.

Nh¾n x²t 4.13. H m m§t m¡t têng cëng l ët h m lçi theo (w, b).

Tèi ÷u h m m§t m¡t: V¼ vi»c tèi ÷u h m m§t m¡t düa tr¶n gradient descent, vi»c ch½nh cõa möc n y l t½nh ¤o h m cõa h m m§t m¡t theo w v b.

Nh¾n x²t 4.14. ¤o h m cõa h m m§t m¡t têng cëng khæng qu¡ phùc t¤p.
V¼ th¸ ta dòng Gradient Descent theow v b º t½nh gi¡ trà tèi ÷u h m m§t m¡t.

Ph¦n t½nh to¡n cö thº, s³ · cªp trong code Python. ¤o h m cõa ph¦n hinge losskhæng qu¡ phùc t¤p:

$$\frac{@(\max(0, 1 \quad y_i(w^T x_i + b)))}{@w} = \begin{cases} y_i x_i & n¸u\ 1 \quad y_i(w^T x_i + b) \quad 0 \\ 0 & n¸u\ 1 \quad y_i(w^T x_i + b) < 0 \end{cases}$$

$$\frac{@(\max(0, 1 \quad y_i(w^T x_i + b)))}{@b} = \begin{cases} y_i & n¸u\ 1 \quad y_i(w^T x_i + b) \quad 0 \\ 0 & n¸u\ 1 \quad y_i(w^T x_i + b) < 0 \end{cases}$$

Ph¦n regularizationcông câ ¤o h m t÷ìng èi ìn gi£n:

$$\frac{@\ \frac{1}{2}kwk_2^2}{@w} = \quad w$$

$$\frac{@\ \frac{1}{2}kwk_2^2}{@b} = 0$$

N¸u cªp nhªt b¬ng gradient descent thæng qua ch¿ mët iºm dú li»u(x_i, y_i) (stochastic gradient descent). N¸u $1 \quad y_i(w^T x_i + b) < 0$, ta khæng cªp nhªt g¼ v chuyºn sang iºm ti¸p theo. Ng÷ñc l¤i biºu thùc cªp nhªt cho w, b ÷ñc cho bði

$$w \quad w \quad (\ y_i x_i + \ w); \quad b \quad b + \ y_i \quad n¸u \quad 1 \quad y_i(w^T x_i + b) \quad 0$$
$$w \quad w \quad w; \quad b \quad b \quad n¸u \quad 1 \quad y_i(w^T x_i + b) < 0$$

vîi l learning rate

4.2.4 Kernel trick

º ¡p döng thuªt to¡n Support Vector Machine l¶n c¡c dú li»u khæng ph¥n bi»t tuy¸n t½nh, ng÷íi ta c¦n bi¸n c¡c iºm dú li»u n y l¶n khæng gian nhi·u chi·u sao cho nâ ph¥n bi»t tuy¸n t½nh ho°c g¦n nh÷ ph¥n bi»t tuy¸n t½nh tr¶n khæng gian â. Ð khæng gian n y, b i to¡n câ thº ÷ñc gi£i quy¸t b¬ng hard/soft-margin SVM. Düa tr¶n suy ngh¾ â, ta i t¼m mët h m sè (x) nh¬m bi¸n êi dú li»u x º thäa c¡c t½nh ch§t tr¶n. Nh¬m t«ng hi»u n«ng t½nh to¡n v ti¸t ki»m bë nhî, c¡c h m Kernel ÷ñc sû döng º mæ t£ mèi quan h» giúa hai iºm dú li»u b§t k¼ trong khæng gian mîi, thay v¼ t½nh to¡n trüc ti¸p bi¸n êi cõa chóng.
V½ dö:

Þ t÷ðng cho h m Kernel X²t h m sè () sao cho sau khi ÷ñc bi¸n êi sang khæng gian mîi, méi x trð th nh (x), v khi â tªp dú li»u mîi trð n¶n g¡n ph¥n bi»t tuy¸n t½nh. Khi â b i to¡n èi ng¨u cõa Soft-margin SVM trð th nh:

$$\begin{aligned} &= \arg\max \sum_{i=1}^{n} \; i \; - \sum_{i=1}^{n}\sum_{j=1}^{n} \; i \; j y_i y_j \; (x_i)^T \; (x_j), \\ \text{thäa m¢n: } &\sum_{i=1}^{n} \; i y_i = 0, \\ &0 \quad i \quad C, 8i = 1, 2, ..., n, \end{aligned} \tag{30}$$

v nhâm cõa mët iºm dú li»u mîi x ch½nh l d§u cõa biºu thùc:

$$w^T \; (x \;) + b = \sum_{j2S} \; j y_j \; (x_j)^T \; (x \;) + \frac{1}{N_M} \sum_{i2M} \; y_i \; - \sum_{j2S} \; j y_j \; (x_j)^T \; (x_i) \; ! \; . \tag{31}$$

Nhªn x²t 4.15. Trong (30) v (31), ta ch¿ c¦n t½nh $(x_i)^T \; (x_j)$ vîi hai iºm dú li»u x_i, x_j b§t k¼. V¼ th¸ ta ch¿ c¦n x¡c ành mët h m $k(x_i, x_j) = \; (x_i)^T \; (x_j)$.

Khi â, (30) v (31) trð th nh:

$$\begin{aligned} &= \arg\max \sum_{i=1}^{n} \; i \; - \sum_{i=1}^{n}\sum_{j=1}^{n} \; i \; j y_i y_j k(x_i, x_j) \\ \text{thäa m¢n: } &\sum_{i=1}^{n} \; i y_i = 0, \\ &0 \quad i \quad C, 8i = 1, 2, ..., n, \end{aligned} \tag{32}$$

v nhâm cõa mët iºm dú li»u mîi x ch½nh l d§u cõa biºu thùc:

$$w^T \; (x \;) + b = \sum_{j2S} \; j y_j k(x_i, x \;) + \frac{1}{N_M} \sum_{i2M} \; y_i \; - \sum_{j2S} \; j y_j k(x_i, x_j) \; ! \; . \tag{33}$$

T½nh ch§t cõa h m Kernel:
+ T½nh ch§t èi xùng.
+ Thäa m¢n i·u ki»n Mercer:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} k(x_i, j_n) c_i c_j \quad 0, 8c_k \; 2 \; R, k = 1, 2, ..., n. \tag{34}$$

Nhªn x²t 4.16. i·u ki»n Mercer gióp £m b£o b i to¡n èi ng¨u (33) l lçi.

Mët sè h m Kernel thæng döng

H m linear (tuy¸n t½nh): $k(x, z) = \; x^T z.$

H m RBF (radial basic function/Gaussian kernel) $k(x, z) = \; exp(\quad kx \quad zk_2^2), \; > \; 0$

H m sigmoid: $k(x, z) = \; tanh(\; x^T z + r)$

H m polynomial (a thùc): $k(x, z) = (\; r + \; x^T z)^d$

5 Áp dụng vào mô hình thực tế

Bài toán Support Vector Machine trên có thể áp dụng vào mô hình dự đoán lừa đảo qua thẻ tín dụng. Ta sẽ áp dụng thuật toán phân lớp nhị phân vào mô hình này. Mục đích của thuật toán là cho biết giao dịch là lừa đảo (**positive class**) hay tin cậy (**negative class**).
Phần code chi tiết thuật toán bằng Python được đề cập ở mục Phụ lục.

5.1 Mô tả dữ liệu

Dữ liệu đầu vào gồm các giao dịch qua thẻ tín dụng thu thập vào 09/2013 tại Châu Âu bao gồm:

- Có 492 giao dịch lừa đảo trong tổng số 284,807 (chiếm **0.172%**).
- 28 đặc trưng ($\mathbf{v} = v_1, v_2, ..., v_{28}$) đã được giảm chiều dữ liệu qua phương thức phân tích thành phần chính (Principal Component Analysis).
- 2 đặc trưng mô tả thời gian và số tiền chuyển giữa hai giao dịch kế nhau.
- Lớp của dữ liệu: là lừa đảo hay tin cậy.

Do dữ liệu về giao dịch lừa đảo quá ít so với tổng số dữ liệu, nên việc dự đoán các giao dịch lừa đảo không được chú trọng. Để giải quyết vấn đề này, ta chỉ giữ lại một số dữ liệu của các giao thức tin cậy sao cho tập dữ liệu cân bằng. Kĩ thuật này gọi là **Undersampling**.
Việc mô phỏng thuật toán SVM trên tập dữ liệu này được thực hiện trên môi trường Python với sự hỗ trợ của thư viện **scikit-learn**.

5.2 Các thuật ngữ quan trọng

Để đánh giá một hệ thống phân lớp, người ta thường phân tập dữ liệu thành 3 phần chính:

- Tập **training**: dùng để huấn luyện mô hình.
- Tập **validation**: tối ưu các tham số khác cho mô hình, sử dụng w được huấn luyện từ tập **training**.
- Tập **test**: dùng để đánh giá thuật toán sau khi đã chọn ra tham số mong muốn .

Việc này nhằm mục đích chọn các tham số tối ưu và đánh giá hệ thống trên một tập dữ liệu chưa thấy (**unseen data**).

Định nghĩa 5.1. Độ chính xác là tỉ lệ giữa số điểm dữ liệu đoán đúng trên tổng số dữ liệu.

Nhận xét 5.2. *Ta thấy rằng với các bộ dữ liệu phân lớp không cân bằng (dữ liệu chẩn đoán có bệnh, xác định phần tử khủng bố), độ chính xác không là thang đo phù hợp để đánh giá hệ thống (do sự thiếu hụt dữ liệu dẫn đến kết quả sai).*

Để khắc phục điều này, người ta đưa ra hai thang đo sau.

Định nghĩa 5.3. Precision là tỉ lệ số điểm được dự đoán đúng **positive class** trên tổng số các điểm được dự đoán **positive class**.

Định nghĩa 5.4. Recall là tỉ lệ số điểm được dự đoán đúng **positive class** trên tổng số các điểm thật sự thuộc **positive class**.

Nhận xét 5.5. *Precision cao đồng nghĩa với việc độ chính xác của các điểm được dự đoán* **positive class** *là cao. Recall cao đồng nghĩa với việc tỉ lệ bỏ sót các điểm thực sự thuộc* **positive class** *là thấp.*

5.3 Tham số C và kernel trong thuật toán SVM

Tham số C là tham số quyết định việc tổng quát hóa của thuật toán với dữ liệu. Ở góc nhìn khác, tham số C xác định việc xem xét các dữ liệu nhiễu gần siêu phẳng phân cách của thuật toán có đáng kể hay không.

Nhận xét 5.6. *Nếu C có giá trị đủ lớn, thuật toán Soft-margin SVM trở thành Hard-margin SVM.*

Thư viện **scikit-learn** cho phép ta cung cấp cụ thể các kernel khác nhau nhằm mục đích chọn ra kernel thích hợp để tối ưu hóa (ở bài toán này là ưu tiên tối đa giá trị recall để dự đoán không bỏ sót các giao dịch lừa đảo).
Quá trình chọn hai tham số trên được thực hiện lần lượt như sau:

- Thực hiện thuật toán trên tập **training** để chọn ra tham số w.
- Lần lượt thử w với các tham số C và kernel khác nhau trên tập **validation** và chọn bộ tham số tối ưu nhất.

Ở mô hình này ta chọn C $\in \{0.01, 1, 100\}$ và kernel $\in$ {'linear','rbf','sigmoid'}. Các kết quả đánh giá với các tham số khác nhau:

Kernel = 'linear'	0.01	1	100
A.Precision	0.983	0.985	0.981
A.Accuracy	0.898	0.898	0.901
A.Recall	0.81	0.808	0.819
Standard deviation of recall	0.0475	0.0413	0.049

Kernel = 'rbf'	0.01	1	100
A.Precision	0.326	0.61	0.607
A.Accuracy	0.483	0.598	0.606
A.Recall	0.666	0.577	0.625
Standard deviation of recall	0.471	0.0623	0.0327

Kernel = 'sigmoid'	0.01	1	100
A.Precision	0.326	0.493	0.462
A.Accuracy	0.483	0.483	0.488
A.Recall	0.666	0.666	0.67
Standard deviation of recall	0.471	0.468	0.465

Sau quá trình thử ta chọn được C = 100 với kernel là hàm tuyến tính (linear).

Nhận xét 5.7. *Ta thấy rằng việc dùng Hard-margin SVM là hợp lí, vì việc dự đoán sót các giao dịch lừa đảo nghiêm trọng hơn việc dự đoán nhầm các giao dịch tin cậy. Lí giải cho việc hàm đa thức và hàm sigmoid không thực hiện hiệu quả bằng hàm tuyến tính, có thể là do tình huống* **Overfitting** *trên tập* **training**.

5.4 Đánh giá mô hình thực tế

Thuật toán dự đoán các giao thức lừa đảo đạt được các kết quả sau trên tập **test**:

- Precision đạt được 0.88.
- Recall đạt được 1.00.

	Precision	Recall
Giao dịch lừa đảo	0.88	1.00
Giao dịch tin cậy	1.00	0.85
Độ chính xác	0.929	