

ĐẠI HỌC HUẾ
TRƯỜNG ĐẠI HỌC KHOA HỌC

NGUYỄN THẾ CƯỜNG

NÂNG CAO HIỆU NĂNG PHÂN LỚP DỮ LIỆU
TRÊN CƠ SỞ CẢI TIẾN THUẬT TOÁN SVM

NGÀNH: KHOA HỌC MÁY TÍNH

MÃ SỐ: 9.48.01.01

TÓM TẮT LUẬN ÁN TIẾN SĨ
KHOA HỌC MÁY TÍNH

HUẾ - NĂM 2023

Công trình được hoàn thành tại Khoa Công nghệ Thông tin,
Trường Đại học Khoa học, Đại học Huế.

Người hướng dẫn khoa học: **PGS. TS. Huỳnh Thế Phùng**

Phản biện 1: **PGS. TS. Trần Văn Lăng**

Phản biện 2: **PGS. TS. Hoàng Hữu Hạnh**

Luận án được bảo vệ trước hội đồng cấp: Cơ sở
vào lúc 14 giờ ngày 29 tháng 11 năm 2022

Có thể tìm hiểu luận án tại thư viện

MỞ ĐẦU

1. Lý do chọn đề tài

Để giải quyết bài toán phân loại mẫu, nhiều thuật toán đã được xây dựng để nhận diện các mẫu khác nhau trên cơ sở các mẫu thử đã được huấn luyện. Một kỹ thuật phân loại có giám sát nổi tiếng là thuật toán SVM (SVM). SVM được vận dụng vào các bài toán như: nhận dạng hình ảnh, chữ viết, âm thanh, sắc thái giọng nói... Nhận thấy SVM vẫn đang là một vấn đề thời sự của cộng đồng nghiên cứu học thuật vì vậy chúng tôi chọn đề tài “*Nâng cao hiệu năng phân lớp dữ liệu trên cơ sở cải tiến thuật toán SVM*” để nghiên cứu.

2. Động lực nghiên cứu

Trong quá trình nghiên cứu SVM và các hướng phát triển, có thể kể đến một vài biến thể tiêu biểu của SVM như: SVM xấp xỉ (PSVM), SVM xấp xỉ thông qua trị riêng suy rộng (GEPSVM), SVM song sinh (TSVM), SVM song sinh có cấu trúc (S-TSVM), SVM song sinh dùng bình phương tối thiểu (LSTSVM).

Đối với dạng dữ liệu có cấu trúc phức tạp, nơi mà mỗi lớp gồm nhiều cụm, mỗi cụm có xu hướng phân phối riêng biệt. SVM và các biến thể chưa khai thác đầy đủ các thông tin về số lượng điểm dữ liệu trong mỗi cụm, thông tin cấu trúc của từng cụm. Điều này có thể ảnh hưởng đến hiệu năng (độ chính xác, thời gian) phân lớp dữ liệu. Đó chính là độc lực để luận án tập trung nghiên cứu và đề xuất mới các giải pháp nâng cao hiệu năng phân lớp dữ liệu đối với dạng dữ liệu có cấu trúc phức tạp.

3. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu là các thuật toán học máy, bài toán phân lớp dữ liệu.

Phạm vi nghiên cứu là học máy có giám sát, cải tiến SVM đối với loại dữ liệu có cấu trúc phức tạp.

4. Mục tiêu của luận án

Đề xuất các phương pháp mới nhằm nâng cao hiệu năng phân lớp dữ liệu đối với dạng dữ liệu có cấu trúc phức tạp, trên cơ sở cải tiến thuật toán SVM. Khai thác được thông tin về cấu trúc của từng cụm và thông tin về số lượng điểm dữ liệu của mỗi cụm trong các lớp.

5. Phương pháp nghiên cứu và giải quyết

- Các phương pháp toán học

Phương pháp nhân tử Lagrange, hệ KKT (Karush - Kuhn - Tucker).

Phương pháp dùng bình phương tối thiểu.

- Các phương pháp xử lý với dữ liệu có nhiều cụm

Khai thác lớp-đối-cụm.

Khai thác cụm-đối-lớp.

- *Phương pháp thực nghiệm khoa học.*

6. Ý nghĩa khoa học và thực tiễn

Ý nghĩa khoa học

Những đóng góp chính của luận án về khoa học:

- Đề xuất các thuật toán phân lớp nhị phân với dữ liệu có cấu trúc phức tạp, sử dụng chiến lược lớp-đối-cụm.

- Đề xuất các thuật toán phân lớp nhị phân với dữ liệu có cấu trúc phức tạp, sử dụng chiến lược cụm-đối-lớp.

Ý nghĩa thực tiễn

- Có thể giải quyết được các bài toán phân lớp với dữ liệu có cấu trúc phức tạp hoặc dữ liệu không cân bằng.
- Luận án có thể được sử dụng làm tài liệu tham khảo cho cộng đồng nghiên cứu đề tài về phân lớp dữ liệu.

7. Bố cục của luận án

Ngoài phần mở đầu và kết luận, luận án gồm 4 chương:

- Chương 1: là kiến thức bổ trợ về bài toán Quy hoạch toàn phương QP và cơ sở toán học của thuật toán SVM.
- Chương 2: là các cải tiến tiêu biểu của SVM được trình bày một cách ngắn gọn, các kết quả có cách tiếp cận sử dụng hai siêu phẳng để phân loại hai lớp dữ liệu.
- Chương 3: là phương pháp lớp-đối-cụm với hai thuật toán mới: SVM có cấu trúc có trọng số (được gọi là WS-SVM) và Cải tiến của SVM dùng bình phương tối thiểu (được gọi là ILS-SVM).
- Chương 4: là chiến lược cụm-đối-lớp với thuật toán mới: SVM dùng bình phương tối thiểu có trọng số (được gọi là WLS-SVM).

Các kết quả của luận án được công bố trong 05 công trình khoa học được đăng trong các hội nghị và tạp chí chuyên ngành trong và ngoài nước. Trong đó có 01 bài đăng trong chuyên san hội thảo quốc gia, 01 bài đăng ở hội thảo quốc tế, 01 bài đăng ở tạp chí Khoa học và Công nghệ, Đại học Khoa học Huế, 01 bài đăng ở tạp chí Kỹ thuật và Công nghệ, Đại học Huế, 01 bài đăng ở tạp chí Tin học và Điều khiển.

CHƯƠNG 1.

CƠ SỞ TOÁN HỌC CỦA SVM

1.1. Hàm toàn phương.

Hàm toàn phương dưới dạng chuẩn

$$Q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{G}\mathbf{x} + \mathbf{g}^T \mathbf{x} + \alpha; \mathbf{x} \in \mathbb{R}^n,$$

Hàm Q lồi khi và chỉ khi \mathbf{G} là ma trận nửa xác định dương. Hơn nữa, khi \mathbf{G} xác định dương thì Q là hàm lồi chặt.

1.2. Bài toán quy hoạch toàn phương.

Dạng ma trận tổng quát

$$(QP) : \begin{cases} Q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{G}\mathbf{x} + \mathbf{g}^T \mathbf{x} + \alpha \longrightarrow \min, \\ \mathbf{A}\mathbf{x} \geq \mathbf{b}, \\ \mathbf{C}\mathbf{x} = \mathbf{d}. \end{cases}$$

Khi hàm mục tiêu Q lồi, ta có bài toán quy hoạch toàn phương lồi.

1.3. Điều kiện tối ưu của bài toán QP

Định lý 1.1 (Điều kiện tối ưu).

(a) Giả sử \mathbf{x}^* là nghiệm của bài toán QP. Khi đó tồn tại các bộ hệ số $\boldsymbol{\lambda}^* = (\lambda_1^*, \dots, \lambda_m^*) \in \mathbb{R}^m$, $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_k^*) \in \mathbb{R}^k$ thỏa mãn:

$$\begin{cases} \mathbf{G}\mathbf{x}^* + \mathbf{g} = \sum_{i=1}^m \lambda_i^* \mathbf{a}_i + \sum_{j=1}^k \mu_j^* \mathbf{c}_j, \\ \mathbf{a}_i^T \mathbf{x}^* \geq b_i, \lambda_i^* \geq 0, & i \in I, \\ \lambda_i^* (\mathbf{a}_i^T \mathbf{x}^* - b_i) = 0, & i \in I, \\ \mathbf{c}_j^T \mathbf{x}^* = d_j, & j \in J. \end{cases}$$

Hệ trên được gọi là hệ KKT (Karush – Kuhn – Tucker) của bài toán QP, \mathbf{x}^* được gọi là điểm KKT, và các hệ số $\boldsymbol{\lambda}^*$, $\boldsymbol{\mu}^*$ được gọi là các nhân tử Lagrange tương ứng với \mathbf{x}^* .

(b) Nếu bài toán QP là lồi, \mathbf{x}^* là một điểm KKT cùng với các nhân tử Lagrange $\boldsymbol{\lambda}^*$, $\boldsymbol{\mu}^*$, thì \mathbf{x}^* cũng là nghiệm của bài toán QP.

Hệ KKT có thể viết lại dưới dạng ma trận như sau:

$$\begin{cases} \mathbf{G}\mathbf{x}^* + \mathbf{g} = \mathbf{A}^T \boldsymbol{\lambda}^* + \mathbf{C}^T \boldsymbol{\mu}^*, \\ \mathbf{A}\mathbf{x}^* \geq \mathbf{b}, \boldsymbol{\lambda}^* \geq \mathbf{0}, \\ \boldsymbol{\lambda}^{*T} (\mathbf{A}\mathbf{x}^* - \mathbf{b}) = \mathbf{0}, \\ \mathbf{C}\mathbf{x}^* = \mathbf{d}. \end{cases}$$

1.4. Bài toán đối ngẫu

Giả sử bài toán QP lồi, ta có hàm Lagrange của bài toán là

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \frac{1}{2} \mathbf{x}^T \mathbf{G}\mathbf{x} + \mathbf{g}^T \mathbf{x} - \boldsymbol{\lambda}^T (\mathbf{A}\mathbf{x} - \mathbf{b}) - \boldsymbol{\mu}^T (\mathbf{C}\mathbf{x} - \mathbf{d}),$$

với các biến $(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}^k$. Từ phương trình dừng

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{G}\mathbf{x} + \mathbf{g} - \mathbf{A}^T \boldsymbol{\lambda} - \mathbf{C}^T \boldsymbol{\mu} = \mathbf{0},$$

ta có bài toán đối ngẫu:

$$\begin{cases} -\frac{1}{2} (\mathbf{A}^T \boldsymbol{\lambda} + \mathbf{C}^T \boldsymbol{\mu} - \mathbf{g})^T \mathbf{G}^{-1} (\mathbf{A}^T \boldsymbol{\lambda} + \mathbf{C}^T \boldsymbol{\mu} - \mathbf{g}) + \boldsymbol{\lambda}^T \mathbf{b} + \boldsymbol{\mu}^T \mathbf{d} \longrightarrow \max, \\ \boldsymbol{\lambda} \in \mathbb{R}_+^m; \boldsymbol{\mu} \in \mathbb{R}^k. \end{cases}$$

1.5. Bài toán phân lớp dữ liệu

Xét bài toán phân loại nhị phân có tập dữ liệu được kí hiệu bởi ma trận $\mathbf{C} \subset \mathbb{R}^{m \times n}$, bao gồm m điểm (mỗi điểm là một hàng của \mathbf{C}) $\mathbf{x}_j^T \in \mathbb{R}^n$, $1 \leq j \leq m$. Giả sử rằng, $y_j \in \Omega = \{-1, 1\}$ là nhãn của điểm dữ liệu \mathbf{x}_j . Lớp $\{+\}$ gồm m_A điểm và được kí hiệu bởi ma trận $\mathbf{A} \subset \mathbb{R}^{m_A \times n}$, lớp $\{-\}$ gồm m_B điểm được kí hiệu bởi ma trận $\mathbf{B} \subset \mathbb{R}^{m_B \times n}$.

Điểm dữ liệu \mathbf{x}_i được xếp vào lớp \mathbf{A} nếu tương ứng ta có $y_i = 1$, và được xếp vào lớp \mathbf{B} nếu $y_i = -1$. Bài toán đặt ra là: cần tìm một hàm phân lớp

$$f : \mathbb{R}^n \rightarrow \{-1, 1\} \text{ thỏa mãn: } f(\mathbf{x}_i) = y_i, \forall i \in Q := \{1, 2, \dots, m\}$$

1.6. Hàm phân lớp tuyến tính

Khi dữ liệu hai lớp là tách được tuyến tính ta giải bài toán tối ưu sau:

$$\begin{cases} \min_{(\mathbf{w}, b)} & \frac{1}{2} \|\mathbf{w}\|^2, \\ \text{s.t.} & 1 - y_i(\mathbf{w}^T \mathbf{x}_i + b) \leq 0, \forall i \in Q. \end{cases}$$

1.7. Siêu phẳng lề mềm

Khi dữ liệu hai lớp bị chồng lấn một phần, ta giải bài toán sau:

$$\begin{cases} \min_{(\mathbf{w}, b, \xi)} & \frac{1}{2} \|\mathbf{w}\|^2 + c \sum_{i=1}^m \xi_i, \\ \text{s.t.} & 1 - y_i(\mathbf{w}^T \mathbf{x}_i + b) - \xi_i \leq 0, \xi_i \geq 0, \forall i \in Q. \end{cases}$$

1.8. Hàm phân lớp phi tuyến

Cho ánh xạ $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ sao cho:

$$\Phi(\mathbf{x}) = (a_1 \varphi_1(\mathbf{x}), a_2 \varphi_2(\mathbf{x}), \dots, a_m \varphi_m(\mathbf{x})) \in \mathbb{R}^m.$$

Ta cần tìm ánh xạ Φ sao cho hai tập $\Phi(\mathbf{A})$ và $\Phi(\mathbf{B})$ (trong \mathbb{R}^m) là tách được tuyến tính. Nghĩa là tồn tại cặp $(\mathbf{w}, b) \in \mathbb{R}^m \times \mathbb{R}$ sao cho:

$$y_i(\mathbf{w}^T \Phi(\mathbf{x}_i) + b) \geq 1, \forall i \in Q.$$

Áp dụng kĩ thuật phân lớp tuyến tính trong không gian \mathbb{R}^m bằng phương pháp hàm Lagrange và đưa về bài toán đối ngẫu:

$$\begin{cases} \max_{(\boldsymbol{\lambda})} & -\frac{1}{2} \boldsymbol{\lambda}^T \mathbf{D} \boldsymbol{\lambda} + \boldsymbol{\lambda}^T \mathbf{e}, \\ \text{s.t.} & \boldsymbol{\lambda} \geq 0, \quad \boldsymbol{\lambda}^T \mathbf{y} = 0, \end{cases}$$

trong đó \mathbf{D} là ma trận vuông với $\mathbf{D}_{ij} = y_i y_j \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j)$, $\mathbf{e} \in \mathbb{R}^m$ là véc-tơ với tất cả các thành phần bằng 1.

1.9. Hàm phân lớp có trọng số

Ta sẽ đưa thêm các biến phụ ξ_i và giải bài toán tối ưu sau:

$$\begin{cases} \min_{(\mathbf{w}, b, \boldsymbol{\xi})} & \frac{1}{2} \|\mathbf{w}\|^2 + \sum_{\mathbf{x}_i \in \mathbf{A}} \delta^+ \xi_i + \sum_{\mathbf{x}_i \in \mathbf{B}} \delta^- \xi_i, \\ \text{s.t.} & y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \forall i \in Q. \end{cases}$$

Tương tự bài toán lẻ mềm, bằng phương pháp nhân tử Lagrange, ta giải bài toán bằng cách đưa về bài toán đối ngẫu.

1.10. Tiểu kết chương

Như vậy, tư tưởng toán học của SVM thực chất là tìm cách tách các lớp dữ liệu bởi một siêu phẳng có khoảng cách đến các tập dữ liệu là lớn nhất. Lời giải cho các trường hợp từ đơn giản đến phức tạp đã được trình bày bằng một phương pháp nhất quán là sử dụng quy tắc nhân tử Lagrange. Trong chương tiếp theo, luận án cung cấp ngắn gọn về các phương pháp sử dụng hai siêu phẳng, song song hoặc không nhất thiết song song để phân lớp dữ liệu.

CHƯƠNG 2.

CÁC BIẾN THỂ CỦA SVM

2.1. SVM xấp xỉ (PSVM)

PSVM (Proximal Support Vector Machine) tách hai lớp dữ liệu bằng cách giải bài toán tối ưu

$$\begin{cases} \min_{\mathbf{w}, b, \xi} & c \frac{1}{2} \|\xi\|^2 + \frac{1}{2} (\mathbf{w}^T \mathbf{w} + b^2), \\ \text{s.t} & \mathbf{D}(\mathbf{C}\mathbf{w} + \mathbf{e}b) + \xi = \mathbf{e}. \end{cases}$$

Với $\mathbf{D} \in \mathbb{R}^{m \times m}$ là ma trận đường chéo nhận giá trị 1 hoặc -1 tương ứng với nhãn của điểm dữ liệu $\mathbf{x}_i, i = 1, \dots, m$.

2.2. PSVM thông qua các trị riêng suy rộng (GEP SVM)

GEP SVM (Proximal Support Vector Machine Via Generalized Eigenvalues) tìm hai siêu phẳng không nhất thiết song song:

- $f_+(\mathbf{x}) (= \mathbf{w}_+^T \mathbf{x} + b_+) = 0$ là gần với lớp $\{+\}$ và cách xa lớp $\{-\}$,
- $f_-(\mathbf{x}) (= \mathbf{w}_-^T \mathbf{x} + b_-) = 0$ là gần với lớp $\{-\}$ và cách xa lớp $\{+\}$. Việc tìm siêu phẳng $f_+(\mathbf{x}) = 0$ tương đương với bài toán tối ưu

$$\min_{(\mathbf{w}_+, b_+) \neq 0} \frac{\|\mathbf{A}\mathbf{w}_+ + \mathbf{e}_A b_+\|^2 / \|(\mathbf{w}_+^T, b_+)\|^2}{\|\mathbf{B}\mathbf{w}_+ + \mathbf{e}_B b_+\|^2 / \|(\mathbf{w}_+^T, b_+)\|^2}$$

2.3. SVM song sinh (TSVM)

TSVM (Twin Support Vector Machine) tìm hai siêu phẳng:

$f_+(\mathbf{x})(= \mathbf{w}_+^T \mathbf{x} + b_+) = 0$ là gần với lớp $\{+\}$ và để lớp $\{-\}$ về một phía,
 $f_-(\mathbf{x})(= \mathbf{w}_-^T \mathbf{x} + b_-) = 0$ là gần với lớp $\{-\}$ và để lớp $\{+\}$ về một phía.

2.3.1. Trường hợp tuyến tính

Trong trường hợp hai lớp dữ liệu là tách được tuyến tính, TSVM tìm hai siêu phẳng bằng cách giải hai bài toán QP lồi như sau:

$$\begin{cases} \min_{\mathbf{w}_+, b_+, \xi} & \frac{1}{2} \|\mathbf{A}\mathbf{w}_+ + \mathbf{e}_A b_+\|^2 + c_+ \mathbf{e}_B^T \xi, \\ \text{s.t.} & -(\mathbf{B}\mathbf{w}_+ + \mathbf{e}_B b_+) + \xi \geq \mathbf{e}_B, \quad \xi \geq \mathbf{0}, \end{cases}$$

$$\begin{cases} \min_{\mathbf{w}_-, b_-, \eta} & \frac{1}{2} \|\mathbf{B}\mathbf{w}_- + \mathbf{e}_B b_-\|^2 + c_- \mathbf{e}_A^T \eta, \\ \text{s.t.} & (\mathbf{A}\mathbf{w}_- + \mathbf{e}_A b_-) + \eta \geq \mathbf{e}_A, \quad \eta \geq \mathbf{0}. \end{cases}$$

2.3.2. Trường hợp phi tuyến

Đặt $\Phi : \mathbb{R}^n \rightarrow \mathbb{S} = \text{span}(\Phi(\mathbf{C}^T))$. Trong \mathbb{S} , siêu phẳng $\Phi(\mathbf{x}^T)\mathbf{h} + b = 0$ có thể được viết lại dưới dạng $\Phi(\mathbf{x}^T)\Phi(\mathbf{C}^T)\mathbf{u} + b = 0$, $\mathbf{u} \in \mathbb{R}^m$. Định nghĩa $\Phi(\mathbf{x}^T)\Phi(\mathbf{C}^T) = K(\mathbf{x}^T, \mathbf{C}^T)$, K là một kernel xác định trước. TSVM xác định 2 siêu phẳng: $K(\mathbf{x}^T, \mathbf{C}^T)\mathbf{u}_+ + b_+ = 0$ và $K(\mathbf{x}^T, \mathbf{C}^T)\mathbf{u}_- + b_- = 0$ bằng cách giải hai bài toán QP lồi:

$$\begin{cases} \min_{\mathbf{u}_+, b_+, \xi} & \frac{1}{2} \|K(\mathbf{A}, \mathbf{C}^T)\mathbf{u}_+ + \mathbf{e}_A b_+\|^2 + c_+ \mathbf{e}_B^T \xi, \\ \text{s.t.} & -(K(\mathbf{B}, \mathbf{C}^T)\mathbf{u}_+ + \mathbf{e}_B b_+) + \xi \geq \mathbf{e}_B, \quad \xi \geq \mathbf{0}, \end{cases}$$

$$\begin{cases} \min_{\mathbf{u}_-, b_-, \eta} & \frac{1}{2} \|K(\mathbf{B}, \mathbf{C}^T)\mathbf{u}_- + \mathbf{e}_B b_-\|^2 + c_- \mathbf{e}_A^T \eta, \\ \text{s.t.} & (K(\mathbf{A}, \mathbf{C}^T)\mathbf{u}_- + \mathbf{e}_A b_-) + \eta \geq \mathbf{e}_A, \quad \eta \geq \mathbf{0}, \end{cases}$$

2.4. TSVM dùng bình phương tối thiểu (LST SVM)

LST SVM (Least Squares Twin Support Vector Machine) cũng tìm hai siêu phẳng bằng cách giải hai bài toán QP lồi với ràng buộc đẳng thức

$$\begin{cases} \min_{\mathbf{w}_+, b_+} & \frac{1}{2} \|\mathbf{A}\mathbf{w}_+ + \mathbf{e}_A b_+\|^2 + \frac{1}{2} c_+ \boldsymbol{\xi}^T \boldsymbol{\xi}, \\ \text{s.t.} & -(\mathbf{B}\mathbf{w}_+ + \mathbf{e}_B b_+) + \boldsymbol{\xi} = \mathbf{e}_B, \end{cases}$$

$$\begin{cases} \min_{\mathbf{w}_-, b_-} & \frac{1}{2} \|\mathbf{B}\mathbf{w}_- + \mathbf{e}_B b_-\|^2 + \frac{1}{2} c_- \boldsymbol{\eta}^T \boldsymbol{\eta}, \\ \text{s.t.} & (\mathbf{A}\mathbf{w}_- + \mathbf{e}_A b_-) + \boldsymbol{\eta} = \mathbf{e}_A. \end{cases}$$

LSTSVM được mở rộng cho trường hợp phi tuyến như TSVM.

2.5. SVM song sinh có cấu trúc (S-TSVM)

Trở lại bài toán phân loại ở Mục 1.5, giả sử có k cụm trong lớp $\{+\}$, cụm thứ i gồm m_{A_i} điểm được kí hiệu bởi ma trận $\mathbf{A}_i \in \mathbb{R}^{m_{A_i} \times n}$, có l cụm trong lớp $\{-\}$, cụm thứ j gồm m_{B_j} điểm được kí hiệu bởi ma trận $\mathbf{B}_j \in \mathbb{R}^{m_{B_j} \times n}$. S-TSVM (Structural Twin Support Vector Machine) tìm hai siêu phẳng bằng cách giải hai bài toán

$$\begin{cases} \min_{\mathbf{w}_+, b_+, \boldsymbol{\xi}} & \frac{1}{2} \|\mathbf{A}\mathbf{w}_+ + \mathbf{e}_A b_+\|^2 + c_+ \mathbf{e}_B^T \boldsymbol{\xi} + \frac{1}{2} \mu_+ (\|\mathbf{w}_+\|^2 + b_+^2) + \frac{1}{2} \lambda_+ \mathbf{w}_+^T \Sigma_+ \mathbf{w}_+, \\ \text{s.t.} & -(\mathbf{B}\mathbf{w}_+ + \mathbf{e}_B b_+) + \boldsymbol{\xi} \geq \mathbf{e}_B, \boldsymbol{\xi} \geq \mathbf{0}, \end{cases}$$

$$\begin{cases} \min_{\mathbf{w}_-, b_-, \boldsymbol{\eta}} & \frac{1}{2} \|\mathbf{B}\mathbf{w}_- + \mathbf{e}_B b_-\|^2 + c_- \mathbf{e}_A^T \boldsymbol{\eta} + \frac{1}{2} \mu_- (\|\mathbf{w}_-\|^2 + b_-^2) + \frac{1}{2} \lambda_- \mathbf{w}_-^T \Sigma_- \mathbf{w}_-, \\ \text{s.t.} & (\mathbf{A}\mathbf{w}_- + \mathbf{e}_A b_-) + \boldsymbol{\eta} \geq \mathbf{e}_A, \boldsymbol{\eta} \geq \mathbf{0}. \end{cases}$$

S-TSVM cũng dễ dàng được mở rộng cho trường hợp phi tuyến.

2.6. Tiểu kết chương

Như vậy, các biến thể của SVM đều dùng hai siêu phẳng để phân hai lớp dữ liệu. Các thuật toán trên chưa khai thác được thông tin cấu trúc và số lượng điểm dữ liệu trong mỗi cụm. Trong chương tiếp theo là các kết quả đã đạt được từ việc khai thác các thông tin trên vào huấn luyện mô hình, với chiến lược lớp-đối-cụm.

CHƯƠNG 3.

PHƯƠNG PHÁP LỚP ĐỐI CỤM

Chương 3 và Chương 4 là những công trình chính tập trung vào phân lớp dữ liệu có cấu trúc, tức là mỗi lớp có nhiều cụm dữ liệu, mỗi cụm có thể có số lượng khác nhau và cấu trúc khác nhau. Chương này là hai thuật toán với chiến lược lớp-đối-cụm: SVM Có Cấu Trúc Có Trọng Số (WS-SVM, công trình 2) và Cải Tiến của SVM Dừng Bình Phương Tối Thiểu (ILS-SVM, công trình 3).

3.1. SVM có cấu trúc có trọng số (WS-SVM)

Xét bài toán phân loại hai lớp như ở Mục 2.5. WS-SVM (Weighted Least Squares Support Vector Machine) xác định $(l + k)$ siêu phẳng, mỗi trong chúng là gần với một lớp và cách xa một cụm trong lớp khác. Cụ thể, tìm l siêu phẳng sao cho siêu phẳng thứ j , $f_{j+}(\mathbf{x}) = \mathbf{w}_{j+}^T \mathbf{x} + b_{j+} = 0$ là gần với lớp $\{+\}$ và cách xa cụm \mathbf{B}_j của lớp $\{-\}$; tìm k siêu phẳng sao cho siêu phẳng thứ i , $f_{i-}(\mathbf{x}) = \mathbf{w}_{i-}^T \mathbf{x} + b_{i-} = 0$ là gần với lớp $\{-\}$ và cách xa cụm \mathbf{A}_i của lớp $\{+\}$. Ở đây $\mathbf{w}_{j+}, \mathbf{w}_{i-} \in \mathbb{R}^n, b_{j+}, b_{i-} \in \mathbb{R}, i = 1, \dots, k; j = 1, \dots, l$.

Bộ phân loại được chọn là:

$$f(\mathbf{x}) = \underset{+, -}{\operatorname{argmin}}(f_+(\mathbf{x}), f_-(\mathbf{x})),$$

$$f_+(\mathbf{x}) = \sum_{j=1}^l \frac{m_{Bj}}{m_B} |f_{j+}(\mathbf{x})|, \quad f_-(\mathbf{x}) = \sum_{i=1}^k \frac{m_{Ai}}{m_A} |f_{i-}(\mathbf{x})|.$$

Một điểm dữ liệu mới \mathbf{x} được phân loại vào lớp $\{+\}$ hoặc lớp $\{-\}$ phụ thuộc vào $f_+(\mathbf{x})$ là bé hơn hay lớn hơn $f_-(\mathbf{x})$.

3.1.1. Trường hợp tuyến tính

Khi dữ liệu hai lớp là tách được tuyến tính, WS-SVM xác định $(l+k)$ siêu phẳng bằng cách giải $(l+k)$ bài toán QP như sau:

$$\begin{cases} \min_{\mathbf{w}_{j+}, b_{j+}, \boldsymbol{\xi}_j} \frac{1}{2} \|\mathbf{A}\mathbf{w}_{j+} + \mathbf{e}_A b_{j+}\|^2 + c_+ \mathbf{e}_{Bj}^T \boldsymbol{\xi}_j + \frac{\mu_+}{2} (\|\mathbf{w}_{j+}\|^2 + b_{j+}^2) + \frac{\lambda_+}{2} \mathbf{w}_{j+}^T \boldsymbol{\Sigma}_+ \mathbf{w}_{j+}, \\ \text{s.t. } -(\mathbf{B}_j \mathbf{w}_{j+} + \mathbf{e}_{Bj} b_{j+}) + \boldsymbol{\xi}_j \geq \mathbf{e}_{Bj}; \boldsymbol{\xi}_j \geq \mathbf{0}, \end{cases}$$

$$\begin{cases} \min_{\mathbf{w}_{i-}, b_{i-}, \boldsymbol{\eta}_i} \frac{1}{2} \|\mathbf{B}\mathbf{w}_{i-} + \mathbf{e}_B b_{i-}\|^2 + c_- \mathbf{e}_{Ai}^T \boldsymbol{\eta}_i + \frac{\mu_-}{2} (\|\mathbf{w}_{i-}\|^2 + b_{i-}^2) + \frac{\lambda_-}{2} \mathbf{w}_{i-}^T \boldsymbol{\Sigma}_- \mathbf{w}_{i-}, \\ \text{s.t. } (\mathbf{A}_i \mathbf{w}_{i-} + \mathbf{e}_{Ai} b_{i-}) + \boldsymbol{\eta}_i \geq \mathbf{e}_{Ai}; \boldsymbol{\eta}_i \geq \mathbf{0}. \end{cases}$$

$j = 1, \dots, l$, và $i = 1, \dots, k$.

Bằng phương pháp nhân tử Lagrange ta có nghiệm

$$\mathbf{z}_{j+} = [\mathbf{w}_{j+}^T, b_{j+}]^T = -[\mathbf{H}^T \mathbf{H} + \mu_+ \mathbf{I} + \lambda_+ \mathbf{F}_+]^{-1} \mathbf{G}_j^T \boldsymbol{\alpha}_j, \quad j = 1, \dots, l,$$

$$\mathbf{z}_{i-} = [\mathbf{w}_{i-}^T, b_{i-}]^T = [\mathbf{G}^T \mathbf{G} + \mu_- \mathbf{I} + \lambda_- \mathbf{F}_-]^{-1} \mathbf{H}_i^T \boldsymbol{\gamma}_i, \quad i = 1, \dots, k,$$

trong đó $\boldsymbol{\alpha}_j, \boldsymbol{\gamma}_i$ là nghiệm của các bài toán đối ngẫu

$$\begin{cases} \max_{\boldsymbol{\alpha}_j} \mathbf{e}_{Bj}^T \boldsymbol{\alpha}_j - \frac{1}{2} \boldsymbol{\alpha}_j^T \mathbf{G}_j [\mathbf{H}^T \mathbf{H} + \mu_+ \mathbf{I} + \lambda_+ \mathbf{F}_+]^{-1} \mathbf{G}_j^T \boldsymbol{\alpha}_j, \\ \text{s.t. } \mathbf{0} \leq \boldsymbol{\alpha}_j \leq c_+ \mathbf{e}_{Bj}, \end{cases}$$

$$\begin{cases} \max_{\boldsymbol{\gamma}_i} \mathbf{e}_{Ai}^T \boldsymbol{\gamma}_i - \frac{1}{2} \boldsymbol{\gamma}_i^T \mathbf{H}_i (\mathbf{G}^T \mathbf{G} + \mu_- \mathbf{I} + \lambda_- \mathbf{F}_-)^{-1} \mathbf{H}_i^T \boldsymbol{\gamma}_i, \\ \text{s.t. } \mathbf{0} \leq \boldsymbol{\gamma}_i \leq c_- \mathbf{e}_{Ai}. \end{cases}$$

$$\text{Với } \mathbf{H} = [\mathbf{A}, \mathbf{e}_A], \mathbf{F}_+ = \begin{bmatrix} \boldsymbol{\Sigma}_+ & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}, \mathbf{G}_j = [\mathbf{B}_j, \mathbf{e}_{Bj}], \mathbf{G} = [\mathbf{B}, \mathbf{e}_B], \mathbf{F}_- =$$

$$\begin{bmatrix} \boldsymbol{\Sigma}_- & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}, \mathbf{H}_i = [\mathbf{A}_i, \mathbf{e}_{Ai}] \text{ và } \mathbf{I} \text{ là ma trận đơn vị bậc } (n+1).$$

3.1.2. Trường hợp phi tuyến

WS-SVM xác định l siêu phẳng: $K(\mathbf{x}^T, \mathbf{C}^T)\mathbf{u}_{j+} + b_{j+} = 0$ là gần với lớp $\Phi(\mathbf{A})$ và cách xa cụm $\Phi(\mathbf{B}_j)$. Xác định k siêu phẳng: $K(\mathbf{x}^T, \mathbf{C}^T)\mathbf{u}_{i-} + b_{i-} = 0$ là gần với lớp $\Phi(\mathbf{B})$ và cách xa cụm $\Phi(\mathbf{A}_i)$ bằng cách giải $(l + k)$ bài toán QP như sau:

$$\left\{ \begin{array}{l} \min_{\mathbf{u}_{j+}, b_{j+}, \boldsymbol{\xi}_j} \frac{1}{2} \|K(\mathbf{A}, \mathbf{C}^T)\mathbf{u}_{j+} + \mathbf{e}_A b_{j+}\|^2 + c_+ \mathbf{e}_{B_j}^T \boldsymbol{\xi}_j + \frac{\mu_+}{2} \|[\mathbf{u}_{j+}, b_{j+}]\|^2 \\ \quad + \frac{\lambda_+}{2} \mathbf{u}_{j+}^T \Phi(\mathbf{C}) \Sigma_+^{\Phi} \Phi(\mathbf{C})^T \mathbf{u}_{j+}, \\ \text{s.t.} \quad - (K(\mathbf{B}_j, \mathbf{C}^T)\mathbf{u}_{j+} + \mathbf{e}_{B_j} b_{j+}) + \boldsymbol{\xi}_j \geq \mathbf{e}_{B_j}, \quad \boldsymbol{\xi}_j \geq \mathbf{0}, \end{array} \right.$$

$$\left\{ \begin{array}{l} \min_{\mathbf{u}_{i-}, b_{i-}, \boldsymbol{\eta}_i} \frac{1}{2} \|K(\mathbf{B}, \mathbf{C}^T)\mathbf{u}_{i-} + \mathbf{e}_B b_{i-}\|^2 + c_- \mathbf{e}_{A_i}^T \boldsymbol{\eta}_i + \frac{\mu_-}{2} \|[\mathbf{u}_{i-}, b_{i-}]\|^2 \\ \quad + \frac{\lambda_-}{2} \mathbf{u}_{i-}^T \Phi(\mathbf{C}) \Sigma_-^{\Phi} \Phi(\mathbf{C})^T \mathbf{u}_{i-}, \\ \text{s.t.} \quad (K(\mathbf{A}_i, \mathbf{C}^T)\mathbf{u}_{i-} + \mathbf{e}_{A_i} b_{i-}) + \boldsymbol{\eta}_i \geq \mathbf{e}_{A_i}, \quad \boldsymbol{\eta}_i \geq \mathbf{0}, \end{array} \right.$$

3.1.3. Thực nghiệm

3.1.3.1. Tập dữ liệu giả 2 chiều

Thực hiện các thuật toán trên các tập dữ liệu giả có số lượng lớn để so sánh về thời gian huấn luyện giữa các thuật toán.

3.1.3.2. Các tập dữ liệu của UCI

So sánh về thời gian huấn luyện, độ chính xác kiểm thử, độ chính xác thẩm định chéo 10-lần trên các tập dữ liệu của UCI.

3.2. Cải tiến LSSVM (ILS-SVM)

ILS-SVM (Improvement Least Squares Support Vector Machine) (công trình 3) sử dụng chiến lược lớp-đối-cụm và các ràng buộc đẳng thức, giải

bài toán trực tiếp bằng cách dùng bình phương tối thiểu.

Bộ phân loại được chọn là

$$f(\mathbf{x}) = \underset{+, -}{\operatorname{argmin}}(f_+(\mathbf{x}), f_-(\mathbf{x})),$$

$$f_+(\mathbf{x}) = \sum_{j=1}^l \frac{m_{Bj}}{m_B} f_{j+}(\mathbf{x}); \quad f_-(\mathbf{x}) = \sum_{i=1}^k \frac{m_{Ai}}{m_A} f_{i-}(\mathbf{x}).$$

3.2.1. Trường hợp tuyến tính

ILS-SVM xác định $(l + k)$ siêu phẳng bằng cách giải $(l + k)$ bài toán

$$\begin{cases} \min_{\mathbf{w}_{j+}, b_{j+}, \xi_j} & \frac{1}{2} \|\mathbf{A}\mathbf{w}_{j+} + \mathbf{e}_A b_{j+}\|^2 + \frac{c_+}{2} \xi_j^T \xi_j + \frac{\mu_+}{2} (\|\mathbf{w}_{j+}\|^2 + b_{j+}^2), \\ \text{s.t.} & (\mathbf{B}_j \mathbf{w}_{j+} + \mathbf{e}_{Bj} b_{j+}) + \xi_j = \mathbf{e}_{Bj}, \end{cases}$$

$$\begin{cases} \min_{\mathbf{w}_{i-}, b_{i-}, \eta_i} & \frac{1}{2} \|\mathbf{B}\mathbf{w}_{i-} + \mathbf{e}_B b_{i-}\|^2 + \frac{c_-}{2} \eta_i^T \eta_i + \frac{\mu_-}{2} (\|\mathbf{w}_{i-}\|^2 + b_{i-}^2), \\ \text{s.t.} & (\mathbf{A}_i \mathbf{w}_{i-} + \mathbf{e}_{Ai} b_{i-}) + \eta_i = \mathbf{e}_{Ai}, \end{cases}$$

$j = 1, \dots, l$ và $i = 1, \dots, k$.

Bằng cách thay thế các ràng buộc đẳng thức vào hàm mục tiêu, giải các phương trình đạo hàm bằng 0 ta có các nghiệm

$$\mathbf{z}_{j+} = [\mathbf{w}_{j+}^T, b_{j+}]^T = \left[\frac{1}{c_+} \mathbf{H}^T \mathbf{H} + \mathbf{G}_j^T \mathbf{G}_j + \frac{\mu_+}{c_+} \mathbf{I} \right]^{-1} \mathbf{G}_j^T \mathbf{e}_{Bj},$$

$$\mathbf{z}_{i-} = [\mathbf{w}_{i-}^T, b_{i-}]^T = \left[\frac{1}{c_-} \mathbf{G}^T \mathbf{G} + \mathbf{H}_i^T \mathbf{H}_i + \frac{\mu_-}{c_-} \mathbf{I} \right]^{-1} \mathbf{H}_i^T \mathbf{e}_{Ai}.$$

Trong đó $\mathbf{H} = [\mathbf{A}, \mathbf{e}_A]$, $\mathbf{G}_j = [\mathbf{B}_j, \mathbf{e}_{Bj}]$, $j = 1, \dots, l$, $\mathbf{G} = [\mathbf{B}, \mathbf{e}_B]$, $\mathbf{H}_i = [\mathbf{A}_i, \mathbf{e}_{Ai}]$, $i = 1, \dots, k$, và \mathbf{I} là ma trận đơn vị bậc $(n + 1)$.

3.2.2. Trường hợp phi tuyến

Khi dữ liệu phi tuyến, tương tự như Mục (3.1.2), ILS-SVM xác định $(l + k)$ siêu phẳng bằng cách giải $(l + k)$ bài toán QP lồi chặt như sau:

$$\begin{cases} \min_{\mathbf{u}_{j+}, b_{j+}, \boldsymbol{\xi}_j} & \frac{1}{2} \|K(\mathbf{A}, \mathbf{C}^T) \mathbf{u}_{j+} + \mathbf{e}_A b_{j+}\|^2 + \frac{c_+}{2} \boldsymbol{\xi}_j^T \boldsymbol{\xi}_j + \frac{\mu_+}{2} (\|\mathbf{u}_{j+}\|^2 + b_{j+}^2), \\ \text{s.t.} & (K(\mathbf{B}_j, \mathbf{C}^T) \mathbf{u}_{j+} + \mathbf{e}_{B_j} b_{j+}) + \boldsymbol{\xi}_j = \mathbf{e}_{B_j}; \end{cases}$$

$$\begin{cases} \min_{\mathbf{u}_{i-}, b_{i-}, \boldsymbol{\eta}_i} & \frac{1}{2} \|K(\mathbf{B}, \mathbf{C}^T) \mathbf{u}_{i-} + \mathbf{e}_B b_{i-}\|^2 + \frac{c_-}{2} \boldsymbol{\eta}_i^T \boldsymbol{\eta}_i + \frac{\mu_-}{2} (\|\mathbf{u}_{i-}\|^2 + b_{i-}^2), \\ \text{s.t.} & (K(\mathbf{A}_i, \mathbf{C}^T) \mathbf{u}_{i-} + \mathbf{e}_{A_i} b_{i-}) + \boldsymbol{\eta}_i = \mathbf{e}_{A_i}; \end{cases}$$

$\mathbf{u}_{j+} \in \mathbb{R}^m, j = 1, \dots, l$ và $\mathbf{u}_{i-} \in \mathbb{R}^m, i = 1, \dots, k$.

3.2.3. Thực nghiệm

3.2.3.1. Tập dữ liệu giả 2 chiều

Thực hiện các thuật toán trên các tập dữ liệu giả 2 chiều có số lượng lớn để so sánh về thời gian huấn luyện.

3.2.3.2. Các tập dữ liệu UCI

So sánh về thời gian huấn luyện, độ chính xác kiểm thử, độ chính xác thẩm định chéo 10-lần trên các tập dữ liệu của UCI.

3.3. Tiểu kết chương

Với chiến lược lớp-đối-cụm, WS-SVM (công trình 2) chia hai bài toán QP ban đầu trong S-TSVM thành $(l + k)$ bài toán QP có cỡ nhỏ hơn và được giải bằng phương pháp đối ngẫu. Cũng với chiến lược lớp-đối-cụm, ILS-SVM (công trình 3) đã khai thác thông tin về số lượng điểm dữ liệu của mỗi cụm vào huấn luyện mô hình. Thuật toán được thiết lập bởi các ràng buộc đẳng thức và được giải bằng cách sử dụng bình phương tối thiểu.

CHƯƠNG 4.

PHƯƠNG PHÁP CỤM ĐỐI LỚP

Trong chương này là thuật toán mới: SVM dùng bình phương tối thiểu có trọng số (được gọi là WLS-SVM, công trình 5), với chiến lược cụm-đối-lớp.

4.1. Biến đổi của S-TSVM

Sử dụng chiến lược cụm-đối-lớp, ta có thể biến đổi hai bài toán QP của S-TSVM trong Mục 2.5 thành các bài toán như sau (công trình 4).

$$\begin{cases} \min_{\mathbf{w}_{i+}, b_{i+}, \xi} & \frac{1}{2} \|\mathbf{A}_i \mathbf{w}_{i+} + \mathbf{e}_{A_i} b_{i+}\|^2 + c_+ \mathbf{e}_B^T \xi + \frac{1}{2} \lambda_+ \mathbf{w}_{i+}^T \Sigma_{i+} \mathbf{w}_{i+}, \\ \text{s.t.} & -(\mathbf{B} \mathbf{w}_{i+} + \mathbf{e}_B b_{i+}) + \xi \geq \mathbf{e}_B; \quad \xi \geq \mathbf{0}, \end{cases}$$

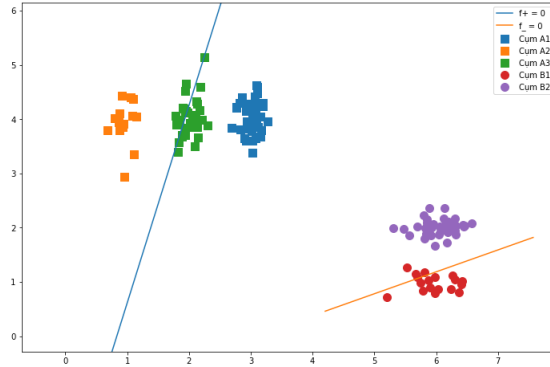
$$\begin{cases} \min_{\mathbf{w}_{j-}, b_{j-}, \eta} & \frac{1}{2} \|\mathbf{B}_j \mathbf{w}_{j-} + \mathbf{e}_{B_j} b_{j-}\|^2 + c_- \mathbf{e}_A^T \eta + \frac{1}{2} \lambda_- \mathbf{w}_{j-}^T \Sigma_{j-} \mathbf{w}_{j-}, \\ \text{s.t.} & (\mathbf{A} \mathbf{w}_{j-} + \mathbf{e}_A b_{j-}) + \eta \geq \mathbf{e}_A; \quad \eta \geq \mathbf{0}, \end{cases}$$

với $i = 1, \dots, k$ và $j = 1, \dots, l$.

Các bài toán này được giải bằng phương pháp nhân tử Lagrange. Ý tưởng của WLS-SVM (Weighted Least Squares Support Vector Machine) (công trình 5) xuất phát từ các bài toán này. Thấy rằng, các bài toán này mặc dù có thể giải được, tuy nhiên các ràng buộc bất đẳng thức đòi hỏi chúng ta phải đưa về giải các bài toán đối ngẫu, điều này vẫn tương đối phức tạp. Chúng ta có thể làm đơn giản bài toán hơn mà vẫn sử dụng chiến lược cụm-đối-lớp, bằng cách dùng bình phương tối thiểu.

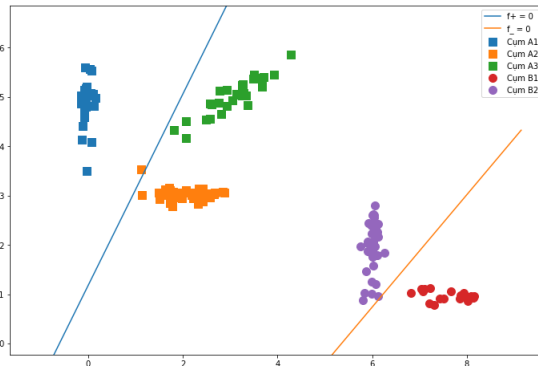
4.2. LSSVM có trọng số (WLS-SVM)

Quay trở lại bài toán phân loại ở Mục 1.5. Trong trường hợp đơn giản, S-TSVM tỏ ra hiệu quả trong mô phỏng xu hướng phân phối dữ liệu.



Hình 4.1: S-TSVM trường hợp dữ liệu có cấu trúc đơn giản

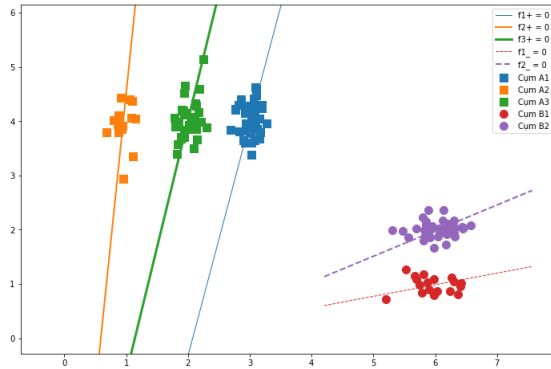
Trong trường hợp dữ liệu có cấu trúc phức tạp, S-TSVM chưa hiệu quả trong mô phỏng xu hướng dữ liệu. Hơn nữa, S-TSVM không khai thác thông tin về số lượng điểm dữ liệu trong mỗi cụm.



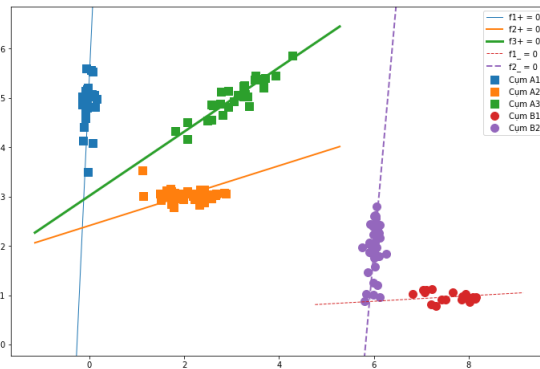
Hình 4.2: S-TSVM bị hạn chế khi dữ liệu có cấu trúc phức tạp

Để khắc phục hạn chế này, WLS-SVM sử dụng chiến lược cụm-đối-lớp và khai thác thông tin về số lượng điểm dữ liệu trong mỗi cụm để tìm

$(k + l)$ siêu phẳng, mỗi siêu phẳng là gần với một cụm của lớp này và để lớp còn lại về một phía. Cụ thể, tìm k siêu phẳng sao cho: siêu phẳng thứ i , $f_{i+}(\mathbf{x}) = \mathbf{w}_{i+}^T \mathbf{x} + b_{i+} = 0$ là gần với cụm \mathbf{A}_i và để lớp \mathbf{B} về một phía; tìm l siêu phẳng sao cho: siêu phẳng thứ j , $f_{j-}(\mathbf{x}) = \mathbf{w}_{j-}^T \mathbf{x} + b_{j-} = 0$ là gần với cụm \mathbf{B}_j và để lớp \mathbf{A} về một phía (xem Hình 4.3 và Hình 4.4).



Hình 4.3: WLS-SVM trong trường hợp dữ liệu có cấu trúc đơn giản.



Hình 4.4: WLS-SVM trong trường hợp dữ liệu có cấu trúc phức tạp

Bộ phân loại được chọn là:

$$f(\mathbf{x}) = \underset{+, -}{\operatorname{argmin}}(f_+(\mathbf{x}), f_-(\mathbf{x})),$$

$$f_+(\mathbf{x}) = \sum_{i=1}^k \frac{m_{A_i}}{m_A} f_{i+}(\mathbf{x}); \quad f_-(\mathbf{x}) = \sum_{j=1}^l \frac{m_{B_j}}{m_B} f_{j-}(\mathbf{x}).$$

Một điểm dữ liệu mới \mathbf{x} được phân loại vào lớp $\{+\}$ hay lớp $\{-\}$ phụ thuộc vào $f_+(\mathbf{x})$ là bé hơn hay lớn hơn $f_-(\mathbf{x})$.

4.2.1. Trường hợp tuyến tính

Chúng ta xác định $(k+l)$ siêu phẳng trong WLS-SVM bằng cách giải $(k+l)$ bài toán QP lồi chặt như sau

$$\begin{cases} \min_{\mathbf{w}_{i+}, b_{i+}, \xi} & \frac{1}{2} \|\mathbf{A}_i \mathbf{w}_{i+} + \mathbf{e}_{A_i} b_{i+}\|^2 + \frac{1}{2} c_+ \xi^T \xi + \frac{1}{2} \mu_+ (\|\mathbf{w}_{i+}\|^2 + b_{i+}^2), \\ \text{s.t.} & (\mathbf{B} \mathbf{w}_{i+} + \mathbf{e}_B b_{i+}) + \xi = \mathbf{e}_B, \end{cases}$$

$$\begin{cases} \min_{\mathbf{w}_{j-}, b_{j-}, \eta} & \frac{1}{2} \|\mathbf{B}_j \mathbf{w}_{j-} + \mathbf{e}_{B_j} b_{j-}\|^2 + \frac{1}{2} c_- \eta^T \eta + \frac{1}{2} \mu_- (\|\mathbf{w}_{j-}\|^2 + b_{j-}^2), \\ \text{s.t.} & (\mathbf{A} \mathbf{w}_{j-} + \mathbf{e}_A b_{j-}) + \eta = \mathbf{e}_A, \end{cases}$$

$i = 1, \dots, k$ và $j = 1, \dots, l$. Bằng cách thay thế các ràng buộc đẳng thức vào hàm mục tiêu, giải các đạo hàm bằng không ta có:

$$\mathbf{z}_{i+} = [\mathbf{w}_{i+}^T, b_{i+}]^T = \left[\frac{1}{c_+} \mathbf{H}_i^T \mathbf{H}_i + \mathbf{G}^T \mathbf{G} + \frac{\mu_+}{c_+} \mathbf{I} \right]^{-1} \mathbf{G}^T \mathbf{e}_B,$$

$$\mathbf{z}_{j-} = [\mathbf{w}_{j-}^T, b_{j-}]^T = \left[\frac{1}{c_-} \mathbf{G}_j^T \mathbf{G}_j + \mathbf{H}^T \mathbf{H} + \frac{\mu_-}{c_-} \mathbf{I} \right]^{-1} \mathbf{H}^T \mathbf{e}_A.$$

Trong đó $\mathbf{H}_i = [\mathbf{A}_i, \mathbf{e}_{A_i}]$, $\mathbf{G} = [\mathbf{B}, \mathbf{e}_B]$, $\mathbf{G}_j = [\mathbf{B}_j, \mathbf{e}_{B_j}]$, $\mathbf{H} = [\mathbf{A}, \mathbf{e}_A]$, \mathbf{I} là ma trận đơn vị cỡ $(n+1)$.

4.2.2. Trường hợp phi tuyến

WLS-SVM xác định $(k+l)$ siêu phẳng bằng cách giải $(k+l)$ bài toán QP lồi chặt như sau:

$$\begin{cases} \min_{\mathbf{u}_{i+}, b_{i+}, \xi} & \frac{1}{2} \|K(\mathbf{A}_i, \mathbf{C}^T) \mathbf{u}_{i+} + \mathbf{e}_{A_i} b_{i+}\|^2 + \frac{1}{2} c_+ \xi^T \xi + \frac{1}{2} \mu_+ (\|\mathbf{u}_{i+}\|^2 + b_{i+}^2), \\ \text{s.t.} & (K(\mathbf{B}, \mathbf{C}^T) \mathbf{u}_{i+} + \mathbf{e}_B b_{i+}) + \xi = \mathbf{e}_B; \end{cases}$$

$$\begin{cases} \min_{\mathbf{u}_{j-}, b_{j-}, \boldsymbol{\eta}} \frac{1}{2} \|K(\mathbf{B}_j, \mathbf{C}^T) \mathbf{u}_{j-} + \mathbf{e}_{B_j} b_{j-}\|^2 + \frac{1}{2} c_- \boldsymbol{\eta}^T \boldsymbol{\eta} + \frac{1}{2} \mu_- (\|\mathbf{u}_{j-}\|^2 + b_{j-}^2), \\ \text{s.t.} \quad (K(\mathbf{A}, \mathbf{C}^T) \mathbf{u}_{j-} + \mathbf{e}_A b_{j-}) + \boldsymbol{\eta} = \mathbf{e}_A; \end{cases}$$

Bằng cách làm tương tự trường hợp dữ liệu tuyến tính ta có:

$$\mathbf{z}_{i+} = [\mathbf{u}_{i+}^T, b_{i+}]^T = \left[\frac{1}{c_+} \mathbf{H}_i^T \mathbf{H}_i + \mathbf{G}^T \mathbf{G} + \frac{\mu_+}{c_+} \mathbf{I} \right]^{-1} \mathbf{G}^T \mathbf{e}_B,$$

$$\mathbf{z}_{j-} = [\mathbf{u}_{j-}^T, b_{j-}]^T = \left[\frac{1}{c_-} \mathbf{G}_j^T \mathbf{G}_j + \mathbf{H}^T \mathbf{H} + \frac{\mu_-}{c_-} \mathbf{I} \right]^{-1} \mathbf{H}^T \mathbf{e}_A.$$

Với $\mathbf{H}_i = [K(\mathbf{A}_i, \mathbf{C}^T), \mathbf{e}_{A_i}]$, $\mathbf{G} = [K(\mathbf{B}, \mathbf{C}^T), \mathbf{e}_B]$, \mathbf{I} là ma trận đơn vị cỡ $(m+1)$, $\mathbf{G}_j = [K(\mathbf{B}_j, \mathbf{C}^T), \mathbf{e}_{B_j}]$, $\mathbf{H} = [K(\mathbf{A}, \mathbf{C}^T), \mathbf{e}_A]$.

4.3. Thực nghiệm

4.3.1. Tập dữ liệu giả 2 chiều

Thực hiện các thuật toán TSVM, LSTSVM, S-TSVM và WLS-SVM trên các tập dữ liệu giả 2 chiều có số lượng lớn.

4.3.2. Các tập dữ liệu UCI

So sánh về thời gian huấn luyện, độ chính xác kiểm thử, độ chính xác thẩm định chéo 10-lần giữa các thuật toán WLS-SVM, S-TSVM, LSTSVM và TSVM.

4.4. Tiểu kết chương

Như vậy, với chiến lược cụm-đối-lớp, WLS-SVM (công trình 5) đã khai thác thông tin về số lượng điểm dữ liệu của mỗi cụm, thông tin cấu trúc của từng cụm trong huấn luyện mô hình. Thuật toán tỏ ra hiệu quả về mô phỏng xu hướng phân phối của các cụm đối với cả dữ liệu có cấu trúc đơn giản và phức tạp.

KẾT LUẬN

Trong lĩnh vực Học máy, phân loại mẫu có giám sát đã, đang và sẽ tiếp tục phát triển không ngừng. Trong đó, thuật toán SVM và các biến thể của nó vẫn tỏ ra hiệu quả so với các phương pháp học máy khác. Bởi lẽ, SVMs đưa bài toán phân loại mẫu về một bài toán tối ưu, cụ thể là bài toán QP lồi hoặc lồi chặt. Luận án đã làm sáng tỏ rằng: tư tưởng toán học của SVM thực chất là tìm cách tách các lớp dữ liệu bởi một siêu phẳng có khoảng cách đến các tập dữ liệu là lớn nhất, bằng một phương pháp nhất quán là sử dụng quy tắc nhân tử Lagrange.

Chương 1 là cung cấp các khái niệm và kết quả cơ bản về toán. Cụ thể đó là hàm toàn phương, bài toán QP, điều kiện tối ưu của bài toán QP, bài toán đối ngẫu của bài toán QP lồi. Tiếp đó là cơ sở toán học của SVM trong kỹ thuật phân lớp dữ liệu, cho các trường hợp khác nhau, từ đơn giản đến phức tạp. Trường hợp đơn giản nhất là hàm phân lớp tuyến tính, tiếp theo là kỹ thuật siêu phẳng lề mềm cho bài toán không tách được tuyến tính, trường hợp hàm phân lớp phi tuyến và cuối cùng là phân lớp có trọng số.

Chương 2 trình bày ngắn gọn một số biến thể của SVM. Đầu tiên là cách tiếp cận dùng hai siêu phẳng song song để phân lớp dữ liệu (PSVM), sau đó là các phương pháp phân lớp dữ liệu bằng cách sử dụng hai siêu phẳng không nhất thiết song song (GEPSVM, LSTSVM, S-TSVM). Chúng tôi đã chỉ ra ưu nhược của các phương pháp trên, khi dữ liệu hai lớp có cấu trúc phức tạp.

Chương 3 và chương 4 là các kết quả mới mà chúng tôi đã công bố. Cụ

thể, Chương 3 là hai thuật toán phân loại sử dụng chiến lược lớp-đối-cụm: SVM có cấu trúc có trọng số (được gọi là WS-SVM, công trình 2) và Cải tiến của SVM dùng bình phương tối thiểu (được gọi là ILS-SVM, công trình 3). Bộ phân loại của hai thuật toán này đều dựa vào trung bình có trọng số các khoảng cách từ một điểm dữ liệu đến các siêu phẳng gần với mỗi lớp. WS-SVM sử dụng thông tin cấu trúc theo cụm và được giải thông qua bài toán đối ngẫu, còn ILS-SVM sử dụng bình phương tối thiểu để tìm ra nghiệm của các bài toán QP.

Chương 4 là biến đổi của thuật toán S-TSVM (công trình 4) và thuật toán SVM dùng bình phương tối thiểu có trọng số (được gọi là WLS-SVM, công trình 5) cho bài toán phân loại với chiến lược cụm-đối-lớp. Bộ phân loại dựa vào trung bình có trọng số các khoảng cách từ một điểm tới các siêu phẳng gần với các cụm. Thuật toán WLS-SVM được giải bằng cách sử dụng bình phương tối thiểu. Các thuật toán đều gồm hai bước: Bước thứ nhất là phân cụm trong mỗi lớp bằng phương pháp liên kết của Ward; Bước thứ hai là huấn luyện mô hình.

Đối với các bài toán có dữ liệu lớn và mỗi lớp chứa nhiều cụm có xu hướng phân phối khác nhau, phương pháp cụm-đối-lớp tỏ ra hiệu quả hơn trong mô phỏng xu hướng phân phối của các cụm và do đó đạt được độ chính xác cao hơn trong phân loại. Có thể phương pháp này không phù hợp cho bài toán phân loại nhiều lớp, tuy nhiên nó có thể hiệu quả đối với bài toán phân loại nhị phân với dữ liệu không cân bằng. Kết hợp phương pháp lớp-đối-cụm và cụm-đối-lớp có thể giải quyết được bài toán phân loại nhiều lớp hay không. Đây cũng là một trong những hướng nghiên cứu đáng để cộng đồng học thuật quan tâm tới. Ngoài ra, có thể áp dụng một trong các phương pháp trên với các kỹ thuật xử lý tín hiệu âm thanh để xây dựng ứng dụng nhận dạng sắc thái giọng nói.

DANH MỤC CÁC CÔNG TRÌNH KHOA HỌC CỦA TÁC GIẢ LIÊN QUAN ĐẾN LUẬN ÁN

1. Nguyễn Thế Cường, Máy véc-tơ tựa song sinh và áp dụng, *Tap chí Khoa học và Công nghệ, Đại học Khoa học, Đại học Huế*, T. 17, Tr. 1-13, 2020.
2. Nguyen The Cuong and Huynh The Phung, Weighted structural support vector machine, *Journal of Computer Science and Cybernetics*, Vietnam, vol. 37, no. 1, pp. 43-56, 2021.
3. Nguyen The Cuong and Nguyen Thanh Vi, Improvement of least square - twin support vector machine, *Journal of Research and Development on Information and Communication Technology*, vol. 2021, no. 1, pp. 8-13, 2021.
4. Nguyen The Cuong, Hierarchical multi twin support vector machine, *Hue University Journal of Science: Techniques and Technology*, vol. 130, no. 2B, 2021.
5. Nguyen The Cuong and Huynh The Phung, Weighted least square - support vector machine, in *2021 RIVF International Conference on Computing and Communication Technologies (RIVF)*, Hanoi, ser. 15. IEEE, 2021.

**HUE UNIVERSITY
UNIVERSITY OF SCIENCES**

NGUYEN THE CUONG

**ENHANCING DATA CLASSIFICATION PERFORMANCE
BASED ON IMPROVING THE SVM ALGORITHM**

MAJOR: COMPUTER SCIENCE

CODE: 9.48.01.01

**PhD DISSERTATION ABSTRACT
COMPUTER SCIENCE**

HUE - 2023

The dissertation has been completed at Department of Information Technology, University of Sciences, Hue University.

Supervisor: **Assoc. Prof., Dr. Huynh The Phung**

Reviewer 1: **Assoc. Prof., Dr. Tran Van Lang**

Reviewer 2: **Assoc. Prof., Dr. Hoang Huu Hanh**

The dissertation is defended in front of level council: Foundation
at 14 hours date 29 month 11 year 2022

The dissertation will be saved in the library

INTRODUCTION

1. The reason for choosing the subject of dissertation

The objective of sample classification is to find the rules based on outside observation to assign one object into exactly some class. Many algorithms have been constructed to recognize different samples based on trained samples base of them. The famous supervised classification technique is the support vector machine algorithm (SVM). SVM is applied to problems such as image recognition, handwriting, sound, speech emotion... Since this is a topical problem of academic research community, so we choose the subject “*Enhancing data classification performance based on improving the SVM algorithm*” to research.

2. Research motivation

When researching SVM and development directions, there are some typical variations of SVM can be mentioned such as: Proximal SVM (PSVM), PSVM via generalized eigenvalues (GEPSVM), Twin SVM (TSVM), Structural TSVM (S-TSVM), Least Square TSVM (LST SVM). For complex structured data, where each class consists of many clusters, each cluster tends to have a distinct distribution. SVM and its variants have not fully exploited the information about the number of data points in each cluster, and the structure information of each cluster. This may affect the performance (accuracy, time) of data classification. That is the motivation for the dissertation to focus on research and propose new methods to improve data classification performance for complex structured data.

3. Object and scope of dissertation

The object of research is machine learning, the data classification problems.

The scope of research is supervised learning, improving SVM for complex structured data.

4. Purpose of dissertation

Propose the new methods to enhance data classification performance for complex structured data, based on improving SVM algorithm. Exploiting information about the structure of each cluster, and the number of data points of each cluster in classes.

5. Methods of dissertation

- **Methods of mathematical**

- The Lagrange multiplier method, KKT system (Karush - Kuhn - Tucker).*

- The least squares method.*

- Data processing method with many clusters

- Exploiting class-vs-cluster.*

- Exploiting cluster-vs-class.*

- *Scientific experimental method.*

6. Scientific and practical significance

Scientific significance

The main contributions of the dissertation on science:

- Proposing binary classification algorithms with complex structured data, using class-vs-cluster strategy.

- Proposing binary classification algorithms with complex structured data, using cluster-vs-class strategy.

Practical significance

- Can be solving the classification problem with complex structured data or unbalanced data.
- The dissertation can serve as a reference for academy research community to carry out the data classification topic.

7. Structures of dissertation

The introduction, conclusion, and four chapters:

- Chapter 1: is a foundation about Quadratic Programming (QP) problem and the mathematical basis of the SVM algorithm.
- Chapter 2: is typical improvements of SVM are briefly presented, the results having an approach using two hyperplanes to classify two data classes.
- Chapter 3: is class-vs-cluster strategy and two new methods: Weighted structural SVM (called WS-SVM) and Improvement of least squares SVM (called ILS-SVM).
- Chapter 4: is cluster-vs-class strategy, together with a new algorithm: Weighted least square SVM (called WLS-SVM).

The results are published in 05 scientific works, which are published in domestic and foreign conferences, and specialized journals. Which, there is 01 article in the National conference journal, 01 article in the International conferences, 01 in the Journal of Science and Technology, Hue University of Sciences, 01 in the Journal of Techniques and Technology, Hue University, and 01 article in the Journal of Computer Science and Cybernetics.

CHAPTER 1.

MATHEMATICAL BASIS OF SVM

1.1. Quadratic function.

The standard form of quadratic function is as follows:

$$Q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{G}\mathbf{x} + \mathbf{g}^T \mathbf{x} + \alpha; \mathbf{x} \in \mathbb{R}^n,$$

The function Q is convex if and only if \mathbf{G} is a positive semi-definite matrix. Moreover, if \mathbf{G} is positive definite then Q is strictly convex.

1.2. The quadratic programming problem.

The general matrix form of the problem is as follows:

$$(\text{QP}) : \begin{cases} Q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{G}\mathbf{x} + \mathbf{g}^T \mathbf{x} + \alpha \longrightarrow \min, \\ \mathbf{A}\mathbf{x} \geq \mathbf{b}, \\ \mathbf{C}\mathbf{x} = \mathbf{d}. \end{cases}$$

When the objective function Q is convex, QP is a convex quadratic programming problem.

1.3. The optimal conditions of QP problem

Theorem 1.1 (Optimal conditions).

(a) Suppose \mathbf{x}^* is the solution of QP problem. Then \exists coefficients $\boldsymbol{\lambda}^* = (\lambda_1^*, \dots, \lambda_m^*) \in \mathbb{R}^m$, $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_k^*) \in \mathbb{R}^k$ satisfying the following

conditions:

$$\begin{cases} \mathbf{G}\mathbf{x}^* + \mathbf{g} = \sum_{i=1}^m \lambda_i^* \mathbf{a}_i + \sum_{j=1}^k \mu_j^* \mathbf{c}_j, \\ \mathbf{a}_i^T \mathbf{x}^* \geq b_i, \lambda_i^* \geq 0, & i \in I, \\ \lambda_i^* (\mathbf{a}_i^T \mathbf{x}^* - b_i) = 0, & i \in I, \\ \mathbf{c}_j^T \mathbf{x}^* = d_j, & j \in J. \end{cases}$$

The above conditions is called the KKT system (Karush–Kuhn–Tucker) of QP problem, \mathbf{x}^* is called a KKT point, and the coefficients $\boldsymbol{\lambda}^*$, $\boldsymbol{\mu}^*$ are called Lagrange multiplier corresponding to \mathbf{x}^* .

(b) If the QP problem is convex, and \mathbf{x}^* is a KKT point with Lagrange multipliers $\boldsymbol{\lambda}^*$, $\boldsymbol{\mu}^*$, then \mathbf{x}^* is also a solution of QP problem.

The KKT system can be rewritten as a matrix as follows:

$$\begin{cases} \mathbf{G}\mathbf{x}^* + \mathbf{g} = \mathbf{A}^T \boldsymbol{\lambda}^* + \mathbf{C}^T \boldsymbol{\mu}^*, \\ \mathbf{A}\mathbf{x}^* \geq \mathbf{b}, \boldsymbol{\lambda}^* \geq \mathbf{0}, \\ \boldsymbol{\lambda}^{*T} (\mathbf{A}\mathbf{x}^* - \mathbf{b}) = \mathbf{0}, \\ \mathbf{C}\mathbf{x}^* = \mathbf{d}. \end{cases}$$

1.4. The dual problems

Suppose that the QP problem is convex, the Lagrange function of it is: $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \frac{1}{2} \mathbf{x}^T \mathbf{G}\mathbf{x} + \mathbf{g}^T \mathbf{x} - \boldsymbol{\lambda}^T (\mathbf{A}\mathbf{x} - \mathbf{b}) - \boldsymbol{\mu}^T (\mathbf{C}\mathbf{x} - \mathbf{d})$, with the variables $(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}^k$. From the equation

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{G}\mathbf{x} + \mathbf{g} - \mathbf{A}^T \boldsymbol{\lambda} - \mathbf{C}^T \boldsymbol{\mu} = \mathbf{0},$$

we have a dual problem as follows:

$$\begin{cases} -\frac{1}{2} (\mathbf{A}^T \boldsymbol{\lambda} + \mathbf{C}^T \boldsymbol{\mu} - \mathbf{g})^T \mathbf{G}^{-1} (\mathbf{A}^T \boldsymbol{\lambda} + \mathbf{C}^T \boldsymbol{\mu} - \mathbf{g}) + \boldsymbol{\lambda}^T \mathbf{b} + \boldsymbol{\mu}^T \mathbf{d} \longrightarrow \max, \\ \boldsymbol{\lambda} \in \mathbb{R}_+^m; \boldsymbol{\mu} \in \mathbb{R}^k. \end{cases}$$

1.5. The data classification problems

Consider a binary classification problem with the dataset, denoted by a matrix $\mathbf{C} \subset \mathbb{R}^{m \times n}$, consisting of m point (each point is a row of \mathbf{C}) $\mathbf{x}_j^T \in \mathbb{R}^n$, $1 \leq j \leq m$. Suppose that, $y_j \in \Omega = \{-1, 1\}$ is the label of data point \mathbf{x}_j . Class $\{+\}$ consists of m_A points denoted by a matrix $\mathbf{A} \subset \mathbb{R}^{m_A \times n}$, class $\{-\}$ consists of m_B points denoted by a matrix $\mathbf{B} \subset \mathbb{R}^{m_B \times n}$.

The data point \mathbf{x}_i is assigned into class \mathbf{A} if $y_i = 1$, and is assigned into class \mathbf{B} if $y_i = -1$ corresponding. The problem is: need to find a classification function

$$f : \mathbb{R}^n \rightarrow \{-1, 1\} \text{ such that: } f(\mathbf{x}_i) = y_i, \forall i \in Q := \{1, 2, \dots, m\}$$

1.6. The linear classifier

To find the canonical decision surface with a maximum margin we need to solve the following optimization problem:

$$\begin{cases} \min_{(\mathbf{w}, b)} & \frac{1}{2} \|\mathbf{w}\|^2, \\ \text{s.t.} & 1 - y_i(\mathbf{w}^T \mathbf{x}_i + b) \leq 0, \forall i \in Q. \end{cases}$$

1.7. The soft margin hyperplane

We solve the following problem:

$$\begin{cases} \min_{(\mathbf{w}, b, \xi)} & \frac{1}{2} \|\mathbf{w}\|^2 + c \sum_{i=1}^m \xi_i, \\ \text{s.t.} & 1 - y_i(\mathbf{w}^T \mathbf{x}_i + b) - \xi_i \leq 0, \xi_i \geq 0, \forall i \in Q. \end{cases}$$

1.8. The nonlinear classifier

Let $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a mapping such that:

$$\Phi(\mathbf{x}) = (a_1 \varphi_1(\mathbf{x}), a_2 \varphi_2(\mathbf{x}), \dots, a_m \varphi_m(\mathbf{x})) \in \mathbb{R}^m.$$

We need to find Φ such that, $\Phi(\mathbf{A})$ and $\Phi(\mathbf{B})$ (in \mathbb{R}^m) is linear separable. That means, there is a pair of $(\mathbf{w}, b) \in \mathbb{R}^m \times \mathbb{R}$ such that:

$$y_i(\mathbf{w}^T \Phi(\mathbf{x}_i) + b) \geq 1, \forall i \in Q.$$

Apply linear technique in \mathbb{R}^m space and solve the dual problem

$$\begin{cases} \max_{(\boldsymbol{\lambda})} & -\frac{1}{2} \boldsymbol{\lambda}^T \mathbf{D} \boldsymbol{\lambda} + \boldsymbol{\lambda}^T \mathbf{e}, \\ \text{s.t.} & \boldsymbol{\lambda} \geq 0, \quad \boldsymbol{\lambda}^T \mathbf{y} = 0, \end{cases}$$

where \mathbf{D} is a square matrix with $\mathbf{D}_{ij} = y_i y_j \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j)$, $\mathbf{e} \in \mathbb{R}^m$ is a vector with all components equal to 1.

1.9. The weighted classifier

Add the slack variables ξ_i and solve the optimization problem

$$\begin{cases} \min_{(\mathbf{w}, b, \boldsymbol{\xi})} & \frac{1}{2} \|\mathbf{w}\|^2 + \sum_{\mathbf{x}_i \in \mathbf{A}} \delta^+ \xi_i + \sum_{\mathbf{x}_i \in \mathbf{B}} \delta^- \xi_i, \\ \text{s.t.} & y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \forall i \in Q. \end{cases}$$

1.10. Chapter sub-conclusion

Thus, the mathematical idea of SVM is essentially to find a way to separate the data classes by a hyperplane with the largest distance to the data sets. We have presented solutions for simple to complex cases using the Lagrange multiplier rule's consistent method. In the next chapter, the dissertation briefly provide the methods that using two parallel or nonparallel hyperplanes to classify the data.

CHAPTER 2.

THE VARIANTS OF SVM

2.1. Proximal support vector machine (PSVM)

PSVM separates two data classes by solving the problem

$$\begin{cases} \min_{\mathbf{w}, b, \boldsymbol{\xi}} & c \frac{1}{2} \|\boldsymbol{\xi}\|^2 + \frac{1}{2} (\mathbf{w}^T \mathbf{w} + b^2), \\ \text{s.t} & \mathbf{D}(\mathbf{C}\mathbf{w} + \mathbf{e}b) + \boldsymbol{\xi} = \mathbf{e}. \end{cases}$$

Where, $\mathbf{D} \in \mathbb{R}^{m \times m}$ is a diagonal matrix getting value 1 or -1 corresponding to the label of data point $\mathbf{x}_i, i = 1, \dots, m$.

2.2. PSVM via generalized eigenvalues (GEPSVM)

GEPSVM seeks two hyperplanes being not necessarily parallel: $f_+(\mathbf{x})(= \mathbf{w}_+^T \mathbf{x} + b_+) = 0$ is closer to class $\{+\}$ and far away from class $\{-\}$, $f_-(\mathbf{x})(= \mathbf{w}_-^T \mathbf{x} + b_-) = 0$ is closer to class $\{-\}$ and far away from class $\{+\}$. Seeking the hyperplane $f_+(\mathbf{x}) = 0 \Leftrightarrow$

the optimization problem:
$$\min_{(\mathbf{w}_+, b_+) \neq \mathbf{0}} \frac{\|\mathbf{A}\mathbf{w}_+ + \mathbf{e}_A b_+\|^2 / \|(\mathbf{w}_+^T, b_+)\|^2}{\|\mathbf{B}\mathbf{w}_+ + \mathbf{e}_B b_+\|^2 / \|(\mathbf{w}_+^T, b_+)\|^2}$$

2.3. Twin support vector machine (TSVM)

TSVM seeks two nonparallel hyperplanes: $f_+(\mathbf{x})(= \mathbf{w}_+^T \mathbf{x} + b_+) = 0$ is closer to class $\{+\}$ and put class $\{-\}$ to one side, $f_-(\mathbf{x})(= \mathbf{w}_-^T \mathbf{x} + b_-) = 0$ is closer to class $\{-\}$ and put class $\{+\}$ to one side.

2.3.1. The linear case

In the case, two classes are linear separable, TSVM seeks two hyperplanes by solving two convex QP problems as follows:

$$\begin{cases} \min_{\mathbf{w}_+, b_+, \boldsymbol{\xi}} & \frac{1}{2} \|\mathbf{A}\mathbf{w}_+ + \mathbf{e}_A b_+\|^2 + c_+ \mathbf{e}_B^T \boldsymbol{\xi}, \\ \text{s.t.} & -(\mathbf{B}\mathbf{w}_+ + \mathbf{e}_B b_+) + \boldsymbol{\xi} \geq \mathbf{e}_B, \quad \boldsymbol{\xi} \geq \mathbf{0}, \end{cases}$$

$$\begin{cases} \min_{\mathbf{w}_-, b_-, \boldsymbol{\eta}} & \frac{1}{2} \|\mathbf{B}\mathbf{w}_- + \mathbf{e}_B b_-\|^2 + c_- \mathbf{e}_A^T \boldsymbol{\eta}, \\ \text{s.t.} & (\mathbf{A}\mathbf{w}_- + \mathbf{e}_A b_-) + \boldsymbol{\eta} \geq \mathbf{e}_A, \quad \boldsymbol{\eta} \geq \mathbf{0}. \end{cases}$$

2.3.2. The nonlinear case

Let $\Phi : \mathbb{R}^n \rightarrow \mathbb{S} = \text{span}(\Phi(\mathbf{C}^T))$. In \mathbb{S} , the hyperplane $\Phi(\mathbf{x}^T)\mathbf{h} + b = 0$ can be rewritten as the form: $\Phi(\mathbf{x}^T)\Phi(\mathbf{C}^T)\mathbf{u} + b = 0$, $\mathbf{u} \in \mathbb{R}^m$. Defining $\Phi(\mathbf{x}^T)\Phi(\mathbf{C}^T) = K(\mathbf{x}^T, \mathbf{C}^T)$, K is a predefined kernel. TSVM determines 2 hyperplanes: $K(\mathbf{x}^T, \mathbf{C}^T)\mathbf{u}_+ + b_+ = 0$ is closer to class $\Phi(\mathbf{A})$ and far away from class $\Phi(\mathbf{B})$; $K(\mathbf{x}^T, \mathbf{C}^T)\mathbf{u}_- + b_- = 0$ is closer to class $\Phi(\mathbf{B})$ and far away from class $\Phi(\mathbf{A})$.

$$\begin{cases} \min_{\mathbf{u}_+, b_+, \boldsymbol{\xi}} & \frac{1}{2} \|K(\mathbf{A}, \mathbf{C}^T)\mathbf{u}_+ + \mathbf{e}_A b_+\|^2 + c_+ \mathbf{e}_B^T \boldsymbol{\xi}, \\ \text{s.t.} & -(K(\mathbf{B}, \mathbf{C}^T)\mathbf{u}_+ + \mathbf{e}_B b_+) + \boldsymbol{\xi} \geq \mathbf{e}_B, \quad \boldsymbol{\xi} \geq \mathbf{0}, \end{cases}$$

$$\begin{cases} \min_{\mathbf{u}_-, b_-, \boldsymbol{\eta}} & \frac{1}{2} \|K(\mathbf{B}, \mathbf{C}^T)\mathbf{u}_- + \mathbf{e}_B b_-\|^2 + c_- \mathbf{e}_A^T \boldsymbol{\eta}, \\ \text{s.t.} & (K(\mathbf{A}, \mathbf{C}^T)\mathbf{u}_- + \mathbf{e}_A b_-) + \boldsymbol{\eta} \geq \mathbf{e}_A, \quad \boldsymbol{\eta} \geq \mathbf{0}, \end{cases}$$

2.4. Least squares TSVM (LST SVM)

LST SVM also seeks two hyperplanes $\mathbf{w}_+^T \mathbf{x} + b_+ = 0$ and $\mathbf{w}_-^T \mathbf{x} + b_- = 0$ by solving two QP problems with the equal constraints:

$$\begin{cases} \min_{\mathbf{w}_+, b_+} & \frac{1}{2} \|\mathbf{A}\mathbf{w}_+ + \mathbf{e}_A b_+\|^2 + \frac{1}{2} c_+ \boldsymbol{\xi}^T \boldsymbol{\xi}, \\ \text{s.t.} & -(\mathbf{B}\mathbf{w}_+ + \mathbf{e}_B b_+) + \boldsymbol{\xi} = \mathbf{e}_B, \end{cases}$$

$$\begin{cases} \min_{\mathbf{w}_-, b_-} & \frac{1}{2} \|\mathbf{B}\mathbf{w}_- + \mathbf{e}_B b_-\|^2 + \frac{1}{2} c_- \boldsymbol{\eta}^T \boldsymbol{\eta}, \\ \text{s.t.} & (\mathbf{A}\mathbf{w}_- + \mathbf{e}_A b_-) + \boldsymbol{\eta} = \mathbf{e}_A. \end{cases}$$

In the nonlinear case, LSTSVM has been generalized as TSVM.

2.5. Structural twin support vector machine (S-TSVM)

Come again with the classification problem in the Section 1.5, suppose that, there are k clusters in class $\{+\}$, the i -th cluster consists of m_{A_i} points and denoted by a matrix $\mathbf{A}_i \in \mathbb{R}^{m_{A_i} \times n}$, there are l clusters in class $\{-\}$, the j -th cluster consists of m_{B_j} points and denoted by a matrix $\mathbf{B}_j \in \mathbb{R}^{m_{B_j} \times n}$. S-TSVM determines two hyperplanes by solving two (QPP)

$$\begin{cases} \min_{\mathbf{w}_+, b_+, \boldsymbol{\xi}} & \frac{1}{2} \|\mathbf{A}\mathbf{w}_+ + \mathbf{e}_A b_+\|^2 + c_+ \mathbf{e}_B^T \boldsymbol{\xi} + \frac{1}{2} \mu_+ (\|\mathbf{w}_+\|^2 + b_+^2) + \frac{1}{2} \lambda_+ \mathbf{w}_+^T \Sigma_+ \mathbf{w}_+, \\ \text{s.t.} & -(\mathbf{B}\mathbf{w}_+ + \mathbf{e}_B b_+) + \boldsymbol{\xi} \geq \mathbf{e}_B, \quad \boldsymbol{\xi} \geq \mathbf{0}, \end{cases}$$

$$\begin{cases} \min_{\mathbf{w}_-, b_-, \boldsymbol{\eta}} & \frac{1}{2} \|\mathbf{B}\mathbf{w}_- + \mathbf{e}_B b_-\|^2 + c_- \mathbf{e}_A^T \boldsymbol{\eta} + \frac{1}{2} \mu_- (\|\mathbf{w}_-\|^2 + b_-^2) + \frac{1}{2} \lambda_- \mathbf{w}_-^T \Sigma_- \mathbf{w}_-, \\ \text{s.t.} & (\mathbf{A}\mathbf{w}_- + \mathbf{e}_A b_-) + \boldsymbol{\eta} \geq \mathbf{e}_A, \quad \boldsymbol{\eta} \geq \mathbf{0}. \end{cases}$$

2.6. Chapter sub-conclusion

Thus, the variants of SVM use two hyperplanes to separate two data classes. The above algorithms have not yet able to exploit the structural information and the number of data points in each cluster. In the next chapter, we present the results obtained from exploiting the above information to train the model, with the class-vs-cluster strategy.

CHAPTER 3.

THE CLASS-VS-CLUSTER METHOD

The Chapter 3, and 4 are the main works, focus on classification with the data that have structure, which each class consists of many clusters, may have different number, and different structure. This chapter presents two algorithms with the class-vs-cluster strategy: Weighted structural support vector machine (WS-SVM, work 2), and Improvement least squares support vector machine (ILS-SVM, work 3).

3.1. Weighted structural SVM (WS-SVM)

Consider the classification problem as in the Section 2.5. WS-SVM determines $(l + k)$ hyperplanes such that each of them is closer to one class and far away from a cluster of another class. Specifically, seeks l hyperplanes such that the j -th hyperplane, $f_{j+}(\mathbf{x}) = \mathbf{w}_{j+}^T \mathbf{x} + b_{j+} = 0$ is closer to class $\{+\}$ and far away from cluster \mathbf{B}_j of class $\{-\}$; the same, seeks k hyperplanes such that the i -th hyperplane, $f_{i-}(\mathbf{x}) = \mathbf{w}_{i-}^T \mathbf{x} + b_{i-} = 0$ is closer to class $\{-\}$ and far away from cluster \mathbf{A}_i of class $\{+\}$.

The classifier is now selected as:

$$f(\mathbf{x}) = \underset{+, -}{\operatorname{argmin}}(f_+(\mathbf{x}), f_-(\mathbf{x})),$$

$$f_+(\mathbf{x}) = \sum_{j=1}^l \frac{m_{B_j}}{m_B} |f_{j+}(\mathbf{x})|, \quad f_-(\mathbf{x}) = \sum_{i=1}^k \frac{m_{A_i}}{m_A} |f_{i-}(\mathbf{x})|.$$

A new data point \mathbf{x} is assigned into class $\{+\}$ or class $\{-\}$ depending on $f_+(\mathbf{x})$ is less than or greater than $f_-(\mathbf{x})$.

3.1.1. The linear case

WS-SVM determines $(l+k)$ hyperplanes by solving $(l+k)$ convex QP problems as follows:

$$\left\{ \begin{array}{l} \min_{\mathbf{w}_{j+}, b_{j+}, \boldsymbol{\xi}_j} \frac{1}{2} \|\mathbf{A}\mathbf{w}_{j+} + \mathbf{e}_A b_{j+}\|^2 + c_+ \mathbf{e}_{Bj}^T \boldsymbol{\xi}_j + \frac{\mu_+}{2} (\|\mathbf{w}_{j+}\|^2 + b_{j+}^2) + \frac{\lambda_+}{2} \mathbf{w}_{j+}^T \boldsymbol{\Sigma}_+ \mathbf{w}_{j+}, \\ \text{s.t. } -(\mathbf{B}_j \mathbf{w}_{j+} + \mathbf{e}_{Bj} b_{j+}) + \boldsymbol{\xi}_j \geq \mathbf{e}_{Bj}; \boldsymbol{\xi}_j \geq \mathbf{0}, \end{array} \right.$$

$$\left\{ \begin{array}{l} \min_{\mathbf{w}_{i-}, b_{i-}, \boldsymbol{\eta}_i} \frac{1}{2} \|\mathbf{B}\mathbf{w}_{i-} + \mathbf{e}_B b_{i-}\|^2 + c_- \mathbf{e}_{Ai}^T \boldsymbol{\eta}_i + \frac{\mu_-}{2} (\|\mathbf{w}_{i-}\|^2 + b_{i-}^2) + \frac{\lambda_-}{2} \mathbf{w}_{i-}^T \boldsymbol{\Sigma}_- \mathbf{w}_{i-}, \\ \text{s.t. } (\mathbf{A}_i \mathbf{w}_{i-} + \mathbf{e}_{Ai} b_{i-}) + \boldsymbol{\eta}_i \geq \mathbf{e}_{Ai}; \boldsymbol{\eta}_i \geq \mathbf{0}. \end{array} \right.$$

$j = 1, \dots, l$, v\`a $i = 1, \dots, k$.

By using the Lagrange function method we have the solutions

$$\mathbf{z}_{j+} = [\mathbf{w}_{j+}^T, b_{j+}]^T = -[\mathbf{H}^T \mathbf{H} + \mu_+ \mathbf{I} + \lambda_+ \mathbf{F}_+]^{-1} \mathbf{G}_j^T \boldsymbol{\alpha}_j, \quad j = 1, \dots, l,$$

$$\mathbf{z}_{i-} = [\mathbf{w}_{i-}^T, b_{i-}]^T = [\mathbf{G}^T \mathbf{G} + \mu_- \mathbf{I} + \lambda_- \mathbf{F}_-]^{-1} \mathbf{H}_i^T \boldsymbol{\gamma}_i, \quad i = 1, \dots, k,$$

where $\boldsymbol{\alpha}_j, \boldsymbol{\gamma}_i$ are the solutions of the dual problems

$$\left\{ \begin{array}{l} \max_{\boldsymbol{\alpha}_j} \quad \mathbf{e}_{Bj}^T \boldsymbol{\alpha}_j - \frac{1}{2} \boldsymbol{\alpha}_j^T \mathbf{G}_j [\mathbf{H}^T \mathbf{H} + \mu_+ \mathbf{I} + \lambda_+ \mathbf{F}_+]^{-1} \mathbf{G}_j^T \boldsymbol{\alpha}_j, \\ \text{s.t.} \quad \mathbf{0} \leq \boldsymbol{\alpha}_j \leq c_+ \mathbf{e}_{Bj}, \end{array} \right.$$

$$\left\{ \begin{array}{l} \max_{\boldsymbol{\gamma}_i} \quad \mathbf{e}_{Ai}^T \boldsymbol{\gamma}_i - \frac{1}{2} \boldsymbol{\gamma}_i^T \mathbf{H}_i (\mathbf{G}^T \mathbf{G} + \mu_- \mathbf{I} + \lambda_- \mathbf{F}_-)^{-1} \mathbf{H}_i^T \boldsymbol{\gamma}_i, \\ \text{s.t.} \quad \mathbf{0} \leq \boldsymbol{\gamma}_i \leq c_- \mathbf{e}_{Ai}. \end{array} \right.$$

$$\mathbf{H} = [\mathbf{A}, \mathbf{e}_A], \quad \mathbf{F}_+ = \begin{bmatrix} \boldsymbol{\Sigma}_+ & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}, \quad \mathbf{G}_j = [\mathbf{B}_j, \mathbf{e}_{Bj}], \quad \mathbf{G} = [\mathbf{B}, \mathbf{e}_B], \quad \mathbf{F}_- = \begin{bmatrix} \boldsymbol{\Sigma}_- & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}, \quad \mathbf{H}_i = [\mathbf{A}_i, \mathbf{e}_{Ai}], \quad \mathbf{I} \text{ being the unit matrix of order } (n+1).$$

3.1.2. The nonlinear case

WS-SVM determines l hyperplanes: $K(\mathbf{x}^T, \mathbf{C}^T)\mathbf{u}_{j+} + b_{j+} = 0$ is closer to class $\Phi(\mathbf{A})$ and far away from cluster $\Phi(\mathbf{B}_j)$; k hyperplanes: $K(\mathbf{x}^T, \mathbf{C}^T)\mathbf{u}_{i-} + b_{i-} = 0$ is closer to class $\Phi(\mathbf{B})$ and far away from cluster $\Phi(\mathbf{A}_i)$ by solving $(l + k)$ convex QP problems as follows:

$$\left\{ \begin{array}{l} \min_{\mathbf{u}_{j+}, b_{j+}, \boldsymbol{\xi}_j} \frac{1}{2} \|K(\mathbf{A}, \mathbf{C}^T)\mathbf{u}_{j+} + \mathbf{e}_A b_{j+}\|^2 + c_+ \mathbf{e}_{B_j}^T \boldsymbol{\xi}_j + \frac{\mu_+}{2} \|[\mathbf{u}_{j+}, b_{j+}]\|^2 \\ \quad + \frac{\lambda_+}{2} \mathbf{u}_{j+}^T \Phi(\mathbf{C}) \Sigma_+^{\Phi} \Phi(\mathbf{C})^T \mathbf{u}_{j+}, \\ \text{s.t.} \quad - (K(\mathbf{B}_j, \mathbf{C}^T)\mathbf{u}_{j+} + \mathbf{e}_{B_j} b_{j+}) + \boldsymbol{\xi}_j \geq \mathbf{e}_{B_j}, \quad \boldsymbol{\xi}_j \geq \mathbf{0}, \end{array} \right.$$

$$\left\{ \begin{array}{l} \min_{\mathbf{u}_{i-}, b_{i-}, \boldsymbol{\eta}_i} \frac{1}{2} \|K(\mathbf{B}, \mathbf{C}^T)\mathbf{u}_{i-} + \mathbf{e}_B b_{i-}\|^2 + c_- \mathbf{e}_{A_i}^T \boldsymbol{\eta}_i + \frac{\mu_-}{2} \|[\mathbf{u}_{i-}, b_{i-}]\|^2 \\ \quad + \frac{\lambda_-}{2} \mathbf{u}_{i-}^T \Phi(\mathbf{C}) \Sigma_-^{\Phi} \Phi(\mathbf{C})^T \mathbf{u}_{i-}, \\ \text{s.t.} \quad (K(\mathbf{A}_i, \mathbf{C}^T)\mathbf{u}_{i-} + \mathbf{e}_{A_i} b_{i-}) + \boldsymbol{\eta}_i \geq \mathbf{e}_{A_i}, \quad \boldsymbol{\eta}_i \geq \mathbf{0}, \end{array} \right.$$

3.1.3. Experiments

3.1.3.1. The 2-dimensional simulated data

Implement the algorithms on the large scale simulated data to compare the training time between all algorithms.

3.1.3.2. The UCI datasets

Compare the training time, testing and 10-fold Cross Validated (10-fold CV) accuracy on the UCI datasets.

3.2. Improvement least squares SVM (ILS-SVM)

ILS-SVM (work 3) uses the class-vs-cluster strategy, and equality constraints, solves the problem directly by using least squares. Specifi-

cally, ILS-SVM also determines $(l + k)$ hyperplanes such that each hyperplane is closer to one class and far away from a cluster of another class: $f_{j+}(\mathbf{x}) = \mathbf{w}_{j+}^T \mathbf{x} + b_{j+} = 0$ is closer to class **A** and far away from the cluster **B_j**; $f_{i-}(\mathbf{x}) = \mathbf{w}_{i-}^T \mathbf{x} + b_{i-} = 0$ is closer to class **B** and far away from the cluster **A_i**.

The classifier is selected as:

$$f(\mathbf{x}) = \underset{+, -}{\operatorname{argmin}}(f_+(\mathbf{x}), f_-(\mathbf{x})),$$

$$f_+(\mathbf{x}) = \sum_{j=1}^l \frac{m_{Bj}}{m_B} f_{j+}(\mathbf{x}); \quad f_-(\mathbf{x}) = \sum_{i=1}^k \frac{m_{Ai}}{m_A} f_{i-}(\mathbf{x}).$$

3.2.1. The linear case

ILS-SVM determines $(l + k)$ hyperplanes by solving $(l + k)$ convex QP problems as follows:

$$\begin{cases} \min_{\mathbf{w}_{j+}, b_{j+}, \boldsymbol{\xi}_j} & \frac{1}{2} \|\mathbf{A} \mathbf{w}_{j+} + \mathbf{e}_A b_{j+}\|^2 + \frac{c_+}{2} \boldsymbol{\xi}_j^T \boldsymbol{\xi}_j + \frac{\mu_+}{2} (\|\mathbf{w}_{j+}\|^2 + b_{j+}^2), \\ \text{s.t.} & (\mathbf{B}_j \mathbf{w}_{j+} + \mathbf{e}_{Bj} b_{j+}) + \boldsymbol{\xi}_j = \mathbf{e}_{Bj}, \end{cases}$$

$$\begin{cases} \min_{\mathbf{w}_{i-}, b_{i-}, \boldsymbol{\eta}_i} & \frac{1}{2} \|\mathbf{B} \mathbf{w}_{i-} + \mathbf{e}_B b_{i-}\|^2 + \frac{c_-}{2} \boldsymbol{\eta}_i^T \boldsymbol{\eta}_i + \frac{\mu_-}{2} (\|\mathbf{w}_{i-}\|^2 + b_{i-}^2), \\ \text{s.t.} & (\mathbf{A}_i \mathbf{w}_{i-} + \mathbf{e}_{Ai} b_{i-}) + \boldsymbol{\eta}_i = \mathbf{e}_{Ai}, \end{cases}$$

$j = 1, \dots, l$ and $i = 1, \dots, k$.

By substituting the equality constraints into the objective function, and solving derivative equations to zero, we have the solutions:

$$\mathbf{z}_{j+} = [\mathbf{w}_{j+}^T, b_{j+}]^T = \left[\frac{1}{c_+} \mathbf{H}^T \mathbf{H} + \mathbf{G}_j^T \mathbf{G}_j + \frac{\mu_+}{c_+} \mathbf{I} \right]^{-1} \mathbf{G}_j^T \mathbf{e}_{Bj},$$

$$\mathbf{z}_{i-} = [\mathbf{w}_{i-}^T, b_{i-}]^T = \left[\frac{1}{c_-} \mathbf{G}^T \mathbf{G} + \mathbf{H}_i^T \mathbf{H}_i + \frac{\mu_-}{c_-} \mathbf{I} \right]^{-1} \mathbf{H}_i^T \mathbf{e}_{Ai}.$$

Where $\mathbf{H} = [\mathbf{A}, \mathbf{e}_A]$, $\mathbf{G}_j = [\mathbf{B}_j, \mathbf{e}_{Bj}]$, $j = 1, \dots, l$, $\mathbf{G} = [\mathbf{B}, \mathbf{e}_B]$, $\mathbf{H}_i = [\mathbf{A}_i, \mathbf{e}_{Ai}]$, $i = 1, \dots, k$, and \mathbf{I} being the unit matrix of order $(n + 1)$.

3.2.2. The nonlinear case

When the data is nonlinear, by exactly similar way such as Subsection (3.1.2), ILS-SVM determines $(l + k)$ hyperplanes by solving $(l + k)$ convex QP problems as follows:

$$\left\{ \begin{array}{l} \min_{\mathbf{u}_{j+}, b_{j+}, \boldsymbol{\xi}_j} \quad \frac{1}{2} \|K(\mathbf{A}, \mathbf{C}^T) \mathbf{u}_{j+} + \mathbf{e}_A b_{j+}\|^2 + \frac{c_+}{2} \boldsymbol{\xi}_j^T \boldsymbol{\xi}_j + \frac{\mu_+}{2} (\|\mathbf{u}_{j+}\|^2 + b_{j+}^2), \\ \text{s.t.} \quad (K(\mathbf{B}_j, \mathbf{C}^T) \mathbf{u}_{j+} + \mathbf{e}_{B_j} b_{j+}) + \boldsymbol{\xi}_j = \mathbf{e}_{B_j}; \end{array} \right.$$

$$\left\{ \begin{array}{l} \min_{\mathbf{u}_{i-}, b_{i-}, \boldsymbol{\eta}_i} \quad \frac{1}{2} \|K(\mathbf{B}, \mathbf{C}^T) \mathbf{u}_{i-} + \mathbf{e}_B b_{i-}\|^2 + \frac{c_-}{2} \boldsymbol{\eta}_i^T \boldsymbol{\eta}_i + \frac{\mu_-}{2} (\|\mathbf{u}_{i-}\|^2 + b_{i-}^2), \\ \text{s.t.} \quad (K(\mathbf{A}_i, \mathbf{C}^T) \mathbf{u}_{i-} + \mathbf{e}_{A_i} b_{i-}) + \boldsymbol{\eta}_i = \mathbf{e}_{A_i}; \end{array} \right.$$

$\mathbf{u}_{j+} \in \mathbb{R}^m, j = 1, \dots, l$ v\`a $\mathbf{u}_{i-} \in \mathbb{R}^m, i = 1, \dots, k$.

3.2.3. Experiments

3.2.3.1. The 2-dimensional simulated data

Implement the algorithms on the large scale simulated data.

3.2.3.2. The UCI datasets

Compare the training time, testing and 10-fold cross validated accuracy on the UCI datasets.

3.3. Chapter sub-conclusion

With the class-vs-cluster strategy, WS-SVM (works 2) splits the two problems QP originally in S-TSVM into $(l + k)$ QP has a smaller size and is solved by the dual method. With also the class-vs-cluster strategy, ILS-SVM (works 3) exploited information about the number of data points of each cluster in model learning. The algorithm is established by equality constraints and is solved by using least squares.

CHAPTER 4.

THE CLUSTER-VS-CLASS METHOD

In this chapter, we present a new algorithm: Weighted least squares support vector machine (called WLS-SVM, work 5), with the cluster-vs-class strategy.

4.1. Transforming of S-TSVM

Use the cluster-vs-class strategy, two problems of S-TSVM in the Section 2.5 can be transformed into QP problems as follows (work 4).

$$\left\{ \begin{array}{ll} \min_{\mathbf{w}_{i+}, b_{i+}, \boldsymbol{\xi}} & \frac{1}{2} \|\mathbf{A}_i \mathbf{w}_{i+} + \mathbf{e}_{A_i} b_{i+}\|^2 + c_+ \mathbf{e}_B^T \boldsymbol{\xi} + \frac{1}{2} \lambda_+ \mathbf{w}_{i+}^T \Sigma_{i+} \mathbf{w}_{i+}, \\ \text{s.t.} & -(\mathbf{B} \mathbf{w}_{i+} + \mathbf{e}_B b_{i+}) + \boldsymbol{\xi} \geq \mathbf{e}_B; \boldsymbol{\xi} \geq \mathbf{0}, \end{array} \right.$$

$$\left\{ \begin{array}{ll} \min_{\mathbf{w}_{j-}, b_{j-}, \boldsymbol{\eta}} & \frac{1}{2} \|\mathbf{B}_j \mathbf{w}_{j-} + \mathbf{e}_{B_j} b_{j-}\|^2 + c_- \mathbf{e}_A^T \boldsymbol{\eta} + \frac{1}{2} \lambda_- \mathbf{w}_{j-}^T \Sigma_{j-} \mathbf{w}_{j-}, \\ \text{s.t.} & (\mathbf{A} \mathbf{w}_{j-} + \mathbf{e}_A b_{j-}) + \boldsymbol{\eta} \geq \mathbf{e}_A; \boldsymbol{\eta} \geq \mathbf{0}, \end{array} \right.$$

with $i = 1, \dots, k$ and $j = 1, \dots, l$.

These problems can be solved by Lagrange multipliers method. The main idea of WLS-SVM (Weighted Least Squares Support Vector Machine, work 5) comes from these transformed problems. Seeing that, although these problems can be solved, inequality constraints require us to solve the dual problems, which are still complex. We can simplify the problem and still use the cluster-vs-class strategy, by using the least squares method.

4.2. Weighted least squares SVM (WLS-SVM)

Come again to the problem in Section 1.5. In the simple case, S-TSVM proved effective in simulating the data distribution trends. In the

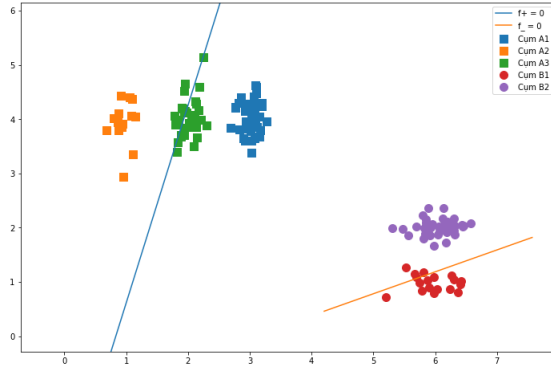


Figure 4.1: S-TSVM in the case of simple structured data

case of complex structured data, S-TSVM is not yet effective in simulating the data trends. Furthermore, S-TSVM does not exploit the information about the number of data points in each cluster.

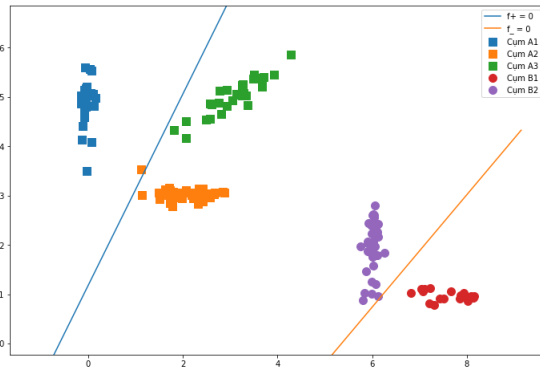


Figure 4.2: S-TSVM is limited when data is complex structured

To overcome this limitation, WLS-SVM uses a cluster-vs-class strategy and exploits information about the number of data points in each class

to find $(k + l)$ hyperplanes such that: each hyperplane is closer to a cluster of one class and put the remaining class to one side. Specifically, seeks k hyperplanes such that: the i -th hyperplane, $f_{i+}(\mathbf{x}) = \mathbf{w}_{i+}^T \mathbf{x} + b_{i+} = 0$ is closer to cluster \mathbf{A}_i and put class \mathbf{B} to one side; l hyperplanes such that: the j -th hyperplane, $f_{j-}(\mathbf{x}) = \mathbf{w}_{j-}^T \mathbf{x} + b_{j-} = 0$ is closer to cluster \mathbf{B}_j and put class \mathbf{A} to one side (see Figure 4.3 and Figure 4.4).

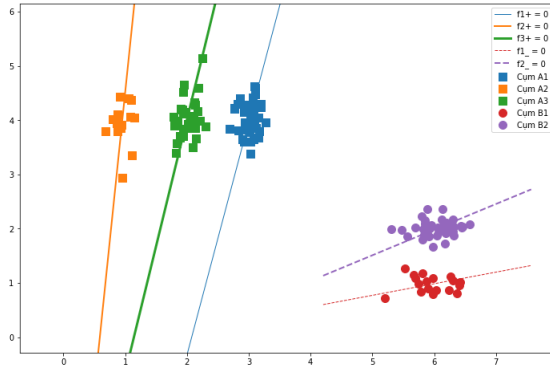


Figure 4.3: WLS-SVM in the case of simple structured data.

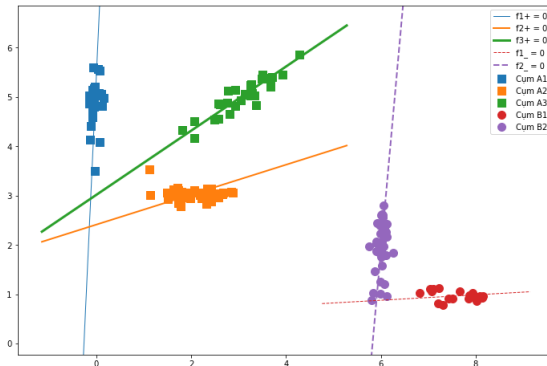


Figure 4.4: WLS-SVM in the case of complex structured data

The classifier is selected as:

$$f(\mathbf{x}) = \underset{+, -}{\operatorname{argmin}}(f_+(\mathbf{x}), f_-(\mathbf{x})),$$

$$f_+(\mathbf{x}) = \sum_{i=1}^k \frac{m_{A_i}}{m_A} f_{i+}(\mathbf{x}); \quad f_-(\mathbf{x}) = \sum_{j=1}^l \frac{m_{B_j}}{m_B} f_{j-}(\mathbf{x}).$$

A new data point \mathbf{x} is assigned to class $\{+\}$ or class $\{-\}$ depending on whether $f_+(\mathbf{x})$ is less than or greater than $f_-(\mathbf{x})$.

4.2.1. The linear case

We determine $(k + l)$ hyperplanes in WLS-SVM by solving $(k + l)$ convex QP problems as follows:

$$\begin{cases} \min_{\mathbf{w}_{i+}, b_{i+}, \boldsymbol{\xi}} & \frac{1}{2} \|\mathbf{A}_i \mathbf{w}_{i+} + \mathbf{e}_{A_i} b_{i+}\|^2 + \frac{1}{2} c_+ \boldsymbol{\xi}^T \boldsymbol{\xi} + \frac{1}{2} \mu_+ (\|\mathbf{w}_{i+}\|^2 + b_{i+}^2), \\ \text{s.t.} & (\mathbf{B} \mathbf{w}_{i+} + \mathbf{e}_B b_{i+}) + \boldsymbol{\xi} = \mathbf{e}_B, \end{cases}$$

$$\begin{cases} \min_{\mathbf{w}_{j-}, b_{j-}, \boldsymbol{\eta}} & \frac{1}{2} \|\mathbf{B}_j \mathbf{w}_{j-} + \mathbf{e}_{B_j} b_{j-}\|^2 + \frac{1}{2} c_- \boldsymbol{\eta}^T \boldsymbol{\eta} + \frac{1}{2} \mu_- (\|\mathbf{w}_{j-}\|^2 + b_{j-}^2), \\ \text{s.t.} & (\mathbf{A} \mathbf{w}_{j-} + \mathbf{e}_A b_{j-}) + \boldsymbol{\eta} = \mathbf{e}_A, \end{cases}$$

$i = 1, \dots, k$ and $j = 1, \dots, l$. Substitute equality constraints into the objective function, solve derivatives equal to zero, we have:

$$\mathbf{z}_{i+} = [\mathbf{w}_{i+}^T, b_{i+}]^T = \left[\frac{1}{c_+} \mathbf{H}_i^T \mathbf{H}_i + \mathbf{G}^T \mathbf{G} + \frac{\mu_+}{c_+} \mathbf{I} \right]^{-1} \mathbf{G}^T \mathbf{e}_B,$$

$$\mathbf{z}_{j-} = [\mathbf{w}_{j-}^T, b_{j-}]^T = \left[\frac{1}{c_-} \mathbf{G}_j^T \mathbf{G}_j + \mathbf{H}^T \mathbf{H} + \frac{\mu_-}{c_-} \mathbf{I} \right]^{-1} \mathbf{H}^T \mathbf{e}_A.$$

Where $\mathbf{H}_i = [\mathbf{A}_i, \mathbf{e}_{A_i}]$, $\mathbf{G} = [\mathbf{B}, \mathbf{e}_B]$, $\mathbf{G}_j = [\mathbf{B}_j, \mathbf{e}_{B_j}]$, $\mathbf{H} = [\mathbf{A}, \mathbf{e}_A]$, \mathbf{I} being the unit matrix of order $(n + 1)$.

4.2.2. The nonlinear case

WLS-SVM determines $(k + l)$ hyperplanes by solving $(k + l)$ convex QP problems as follows:

$$\begin{cases} \min_{\mathbf{u}_{i+}, b_{i+}, \boldsymbol{\xi}} & \frac{1}{2} \|K(\mathbf{A}_i, \mathbf{C}^T) \mathbf{u}_{i+} + \mathbf{e}_{A_i} b_{i+}\|^2 + \frac{1}{2} c_+ \boldsymbol{\xi}^T \boldsymbol{\xi} + \frac{1}{2} \mu_+ (\|\mathbf{u}_{i+}\|^2 + b_{i+}^2), \\ \text{s.t.} & (K(\mathbf{B}, \mathbf{C}^T) \mathbf{u}_{i+} + \mathbf{e}_B b_{i+}) + \boldsymbol{\xi} = \mathbf{e}_B; \end{cases}$$

$$\begin{cases} \min_{\mathbf{u}_{j-}, b_{j-}, \boldsymbol{\eta}} \frac{1}{2} \|K(\mathbf{B}_j, \mathbf{C}^T) \mathbf{u}_{j-} + \mathbf{e}_{B_j} b_{j-}\|^2 + \frac{1}{2} c_- \boldsymbol{\eta}^T \boldsymbol{\eta} + \frac{1}{2} \mu_- (\|\mathbf{u}_{j-}\|^2 + b_{j-}^2), \\ \text{s.t. } (K(\mathbf{A}, \mathbf{C}^T) \mathbf{u}_{j-} + \mathbf{e}_A b_{j-}) + \boldsymbol{\eta} = \mathbf{e}_A; \end{cases}$$

In an exactly similar way to the linear data case, we have

$$\begin{aligned} \mathbf{z}_{i+} &= [\mathbf{u}_{i+}^T, b_{i+}]^T = \left[\frac{1}{c_+} \mathbf{H}_i^T \mathbf{H}_i + \mathbf{G}^T \mathbf{G} + \frac{\mu_+}{c_+} \mathbf{I} \right]^{-1} \mathbf{G}^T \mathbf{e}_B, \\ \mathbf{z}_{j-} &= [\mathbf{u}_{j-}^T, b_{j-}]^T = \left[\frac{1}{c_-} \mathbf{G}_j^T \mathbf{G}_j + \mathbf{H}^T \mathbf{H} + \frac{\mu_-}{c_-} \mathbf{I} \right]^{-1} \mathbf{H}^T \mathbf{e}_A. \end{aligned}$$

$\mathbf{H}_i = [K(\mathbf{A}_i, \mathbf{C}^T), \mathbf{e}_{A_i}]$, $\mathbf{G} = [K(\mathbf{B}, \mathbf{C}^T), \mathbf{e}_B]$, \mathbf{I} being the unit matrix of order $(m + 1)$, $\mathbf{G}_j = [K(\mathbf{B}_j, \mathbf{C}^T), \mathbf{e}_{B_j}]$, $\mathbf{H} = [K(\mathbf{A}, \mathbf{C}^T), \mathbf{e}_A]$.

4.3. Experiments

4.3.1. The 2-dimensional simulated data

Implement algorithms TSVM, LSTSVM, S-TSVM and WLS-SVM on large scale simulated data.

4.3.2. The UCI datasets

Compare the training time, testing and 10-fold CV accuracy between algorithms WLS-SVM, S-TSVM, LSTSVM and TSVM.

4.4. Chapter sub-conclusion

Thus, with the cluster-vs-class strategy, WSL-SVM (works 5) exploited information about the number of data points of each cluster, and structure information of each cluster in model learning. The algorithm proved to be effective in simulating the distribution trend of clusters for both simple and complex structured data.

CONCLUSION

In the field of Machine Learning, supervised pattern classification has been and will continue to evolve continuously. In which, the SVM and its variants are still effective compare to other machine learning. Because, SVMs bring the pattern classification problem to an optimization problem, namely the convex or strictly convex QP problems. The dissertation has clarified that: the mathematical idea of the support vector machine is essential to find a way to separate the data classes by a hyperplane with the largest distance to the datasets, by a consistent method is to use the Lagrange multiplier.

Chapter 1 provides basic math concepts and results. Specifically, it is the quadratic function, the QP problem, the optimal condition of the QP problem, the dual problem of the convex QP problem. Next is the mathematical basis of support vector machine in data classification technique, for different cases, from simple to complex. The simplest case is the linear classification function, followed by the soft margin hyperplane technique for the linear non-separable problem, the nonlinear classifier case, and finally the weighted classifier.

Chapter 2 briefly presents several variants of SVM. The first is the approach using two parallel hyperplanes to classify data (PSVM), then the methods of data classification using two hyperplanes that are not necessarily parallel (GEPSVM, LSTSVM, S-TSVM). We have shown the advantages and disadvantages of the above methods when the two data classes has a complex structure.

Chapters 3 and 4 are the new results that we have published. Specifically, Chapter 3 covers two classification algorithms using a class-vs-cluster strategy: Weighted structured support vector machine (called WS-SVM, work 2) and Improvement least squares support vector machine (called ILS-SVM, work 3). The classifiers of these algorithms are both based on a weighted average of the distances from a data point to the hyperplanes close to each class. WS-SVM uses cluster structure information and is solved by a dual problem, while ILS-SVM uses least squares to find the solution of QP problems.

Chapter 4 is a transformation of the S-TSVM algorithm (work 4) and the weighted least squares support vector machine algorithm (called WLS-SVM, work 5) for classification problems with the cluster-vs-class strategy. The classifier is based on the weighted average of the distances from a point to the hyperplanes close to the clusters. WLS-SVM is solved by using least squares method. The algorithms have two steps: The first step is to cluster in each class using Ward's linkage method; The second step is to train the model.

For problems with large scale data and each class containing many clusters with different distribution trends, the cluster-vs-class method proves to be more effective in simulating the distribution trend of clusters and thus achieves higher accuracy in classification. This method may not be suitable for multi-class classification problems, but it can be effective for binary classification problems with unbalanced data. Combining class-vs-cluster and cluster-vs-class methods can solve the multi-class classification problem. This is also one of interesting research directions for academic community. In addition, can apply one of the above methods with audio signal processing techniques to build an emotion speech recognition application.

LIST OF SCIENTIFIC WORKS OF THE AUTHOR RELATED TO THE DISSERTATION

1. Nguyen The Cuong, Twin support vector machine and application, *Journal of science and technology, University of Sciences, Hue University*, E. 17, pp. 1-13, 2020.
2. Nguyen The Cuong and Huynh The Phung, Weighted structural support vector machine, *Journal of Computer Science and Cybernetics*, Vietnam, vol. 37, no. 1, pp. 43–56, 2021.
3. Nguyen The Cuong and Nguyen Thanh Vi, Improvement of least square - twin support vector machine, *Journal of Research and Development on Information and Communication Technology*, vol. 2021, no. 1, pp. 8-13, 2021.
4. Nguyen The Cuong, Hierarchical multi twin support vector machine, *Hue University Journal of Science: Techniques and Technology*, vol. 130, no. 2B, 2021.
5. Nguyen The Cuong and Huynh The Phung, Weighted least square - support vector machine, in *2021 RIVF International Conference on Computing and Communication Technologies (RIVF)*, Hanoi, ser. 15. IEEE, 2021.