

**Akademia Górniczo-Hutnicza im. Stanisława Staszica
w Krakowie**

Wydział Elektrotechniki, Automatyki, Informatyki i Elektroniki

KATEDRA INFORMATYKI



AUTOREFERAT PRACY DOKTORSKIEJ

MARCIN ORCHEL

**Wykorzystanie wiedzy a priori w analizie danych
wielowymiarowych algorytmami SVM**

PROMOTOR:
prof. dr hab. Witold Dzwinel

Kraków 2012

Streszczenie

W tym raporcie przedstawiamy wyniki badań przeprowadzonych przez autora dotyczące metody regresji, zwanej δ regresja wektorów wspierających (δ -SVR), metody włączania wiedzy o marginesie per przykład, zwanej φ klasyfikacja wektorów wspierających (φ -SVC), implementacji maszyny wektorów wspierających (SVM) oraz zastosowania SVM do składania zleceń giełdowych. W tej pracy, proponujemy metodę, zwaną δ -SVR, która polega na zamianie problemu regresji w problemy binarnej klasyfikacji, które są rozwiązywane za pomocą SVM. Analizujemy statystyczną równoważność problemu regresji z problemem binarnej klasyfikacji. Pokazujemy potencjalną możliwość ulepszenia ograniczeń błędu generalizacji opartych na wymiarze Vapnik-Chervonenkis (VC), w porównaniu do SVM. Wykonaliśmy eksperymenty porównujące δ -SVR z ε regresja wektorów wspierających (ε -SVR) na syntetycznych i rzeczywistych zbiorach danych. Rezultaty wskazują, że δ -SVR osiąga porównywalny błąd generalizacji, mniej wektorów wspierających oraz mniejszy błąd generalizacji dla różnych wartości ε i δ . Metoda δ -SVR jest szybsza dla liniowych jąder używając metody minimalna optymalizacja sekwencyjna (SMO), dla nieliniowych jąder rezultaty szybkości zależą od zbioru danych. W tej pracy proponujemy metodę zwaną φ -SVC do włączania wiedzy marginesowej per przykład do problemów klasyfikacji i regresji. Proponujemy dwie aplikacje dla φ -SVC: zmniejszenie błędu generalizacji modeli zredukowanych przy zachowaniu podobnej liczby wektorów wspierających, oraz włączenie nieliniowego warunku specjalnego typu do problemu. Metoda była przetestowana dla klasyfikatora SVM, oraz dla metody regresji ε -SVR. Eksperymenty na danych rzeczywistych pokazują zmniejszony błąd generalizacji modeli zredukowanych dla liniowych i wielomianowych jąder. W tej pracy proponujemy dwa usprawnienia w implementacji SVM, pierwsze w szybkości trenowania SVM, drugie w uproszczeniu implementacji SVM. Pierwsze usprawnienie, zwane heurystyka alternatyw (HoA), dotyczy nowej heurystyki wyboru parametrów do zbioru roboczego. Bierze ona pod uwagę nie tylko spełnienie warunków Karush-Kuhn-Tucker (KKT), ale również zmianę wartości funkcji celu. Testy na rzeczywistych zbiorach danych pokazują, że HoA prowadzi do zmniejszenia czasu trenowania SVM, porównując do heurystyki standardowej. Drugie usprawnienie, zwane wielowymiarowa, sekwencyjna metod rozwiązywania podproblemów (SMS), dotyczy nowego sposobu rozwiązywania podproblemów o więcej niż dwóch parametrach, zamiast stosowania skomplikowanych metod rozwiązujących problemy z zakresu programowania kwadratowego, używamy do tego celu metodę SMO. Otrzymujemy prostszą implementację, z podobnymi wynikami szybkościowymi. W tej pracy proponujemy zastosowanie regresja wektorów wspierających (SVR) do wykonywania zleceń na rynkach akcyjnych. Używamy SVR do predykcji funkcji partycypacji w wolumenie. Proponujemy ulepszenie przewidywania funkcji partycypacji używając SVM z włączonym do problemu dodatkowym warunkiem nieliniowym. Pokazujemy, że jakość przewidywania wpływa na koszty egzekucji. Ponadto pokazujemy jak możemy włączyć wiedzę o cenach akcji. Porównaliśmy ε -SVR i δ -SVR z prostymi predyktorami takimi jak średnia cena wykonania z poprzednich dni. Testy zostały przeprowadzone na danych dla spółek z indeksu NASDAQ-100. Dla obu metod otrzymaliśmy mniejszą wariancję kosztów egzekucji zleceń. Ponadto, zmniejszyliśmy koszty wykonania zleceń wykorzystując dodatkowo predykcję cen giełdowych.

Przedmowa

Praca ta dotyczy maszyn wektorów wspierających, nowoczesnej metody maszynowego uczenia się, która wyszukuje efektywnie zależności w danych. Ostatnimi czasy SVMy stały się pełnoprawną dziedziną naukową, która obejmuje nie tylko teorię maszynowego uczenia się, ale także teorię optymalizacji, geometrię wielowymiarową oraz teorię funkcji liniowych. Podstawowymi problemami maszynowego uczenia się są klasyfikacja i regresja, dla obu problemów powstały efektywne warianty SVM. Powstało również wiele rozszerzeń SVM, a wśród nich rozszerzenia umożliwiające włączanie dodatkowej wiedzy.

Badania objęte tematyką pracy doprowadziły do opracowania nowoczesnej metody regresji, która rozwiązuje problem regresji za pomocą binarnej klasyfikacji. Metoda ta, oraz koncepcja stojąca za nią rzuca nowe światło na relację między dwoma najważniejszymi problemami maszynowego uczenia się. Drugim ważnym elementem badań była analiza dodatkowej wiedzy w postaci wielkości marginesu per przykład. Jest to naturalne rozszerzenie SVM, ponieważ podstawową koncepcją SVM jest margines. Trzecim elementem badań była analiza implementacji SVM oraz zastosowanie w tematyce finansowej.

Praca powstała dzięki wsparciu ze strony prof. Witolda Dzwiniela z Katedry Informatyki AGH. Była inspirowana współpracą z Marcinem Kurdzielem, Tomaszem Arodziem oraz Witoldem Dzwinielą.

Metody przedstawione w pracy były dyskutowane z prof. Vojislavem Kecmanem.

Chciałbym serdecznie podziękować Izie, rodzicom oraz siostrze. Podziękowania również dla Ryszarda Zięby za pokazanie piękna matematyki w liceum.

Jestem im wszystkim bardzo wdzięczny.

Badania były finansowane z

- grantu Ministerstwa Nauki i Szkolnictwa Wyższego, nr. projektu NN519579338,
- wewnętrznego grantu Katedry Informatyki na Akademii Górniczo-Hutniczej i projektu współfinansowanego przez EU i Ministerstwo Nauki i Szkolnictwa Wyższego, nr UDA-POKL.04.01.01-00-367/08-00 zatytułowanego "Doskonalenie i Rozwój Potencjału Dydaktycznego Kierunku Informatyka w AGH",
- grantu Ministerstwa Nauki i Szkolnictwa Wyższego, nr. projektu 3 T11F 010 30.

Marcin Orchel

Spis treści

I	Wstęp	1
I.1	Przegląd	1
I.2	Problemy optymalizacyjne SVC	5
I.2.1	Dualny problem optymalizacyjny SVC	5
I.2.2	ν -SVC	7
I.3	Problemy optymalizacyjne SVR	7
I.4	Włączanie wiedzy a priori do SVM	7
II	Regresja bazująca na klasyfikacji binarnej	10
II.1	Wstęp do metody δ -SVR	10
II.2	Podsumowanie	11
III	Wiedza marginesowa per przykład	12
III.1	Wprowadzenie do φ -SVC	12
III.2	Podsumowanie	13
IV	Techniki implementacyjne bazujące na warunkach KKT	14
IV.1	Podsumowanie	14
V	Aplikacje: strategie wykonywania zleceń	15
V.1	Podsumowanie	15
VI	Podsumowanie	16
	Bibliografia	17

Rozdział I

Wstęp

W ostatnich latach, SVM stały się popularne ze względu na znakomite teoretyczne i praktyczne wyniki, [29, 30]. SVM są używane z powodzeniem do rozwiązywania problemów klasyfikacji, regresji oraz wielu innych problemów maszynowego uczenia się. SVMy były już szeroko stosowane w rozmaitych dziedzinach, takich jak klasyfikacja tekstu, [8, 27], biotechnologia, [7, 13], ekonomia, [35], chemia, [3, 13], fizyka, [26], i wielu innych. Popularność SVM zawdzięczają w szczególności dobrej generalizacji oraz szybkim trenowaniu. A także solidnej podstawie w statystycznej teorii maszynowego uczenia się, zwracaniu rzadkich, nieliniowych rozwiązań, odporności na błędne dane, interpretacji geometrycznej, sformułowaniu jako problemów wypukłych z zakresu programowania kwadratowego.

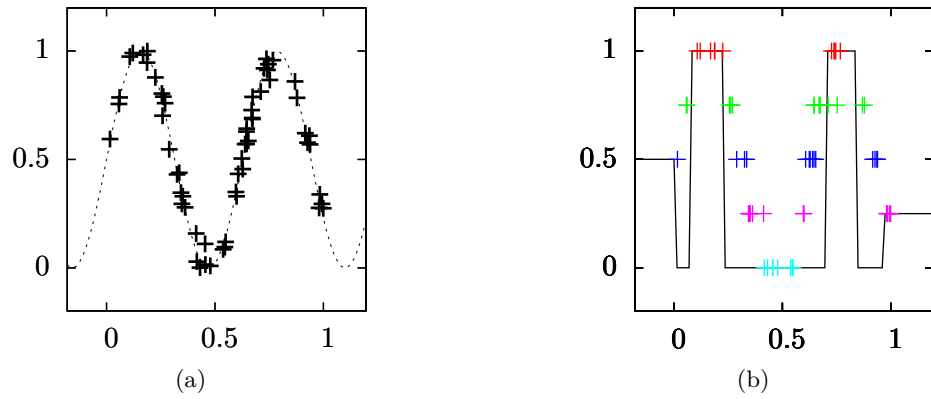
Dwoma najbardziej popularnymi problemami maszynowego uczenia się są klasyfikacja i regresja. Metody regresji mogą być z łatwością użyte do problemów klasyfikacji. Oznacza to możliwość wykorzystania potencjału metod regresji dla problemów klasyfikacji. Czy jest możliwe wykorzystanie metod klasyfikacji do problemów regresji? Na to pytanie będziemy starali się odpowiedzieć w tej pracy.

Drugim głównym zagadnieniem podjętym w niniejszej pracy jest zagadnienie włączania wiedzy a priori do SVM. Przegląd badań w tym temacie znajduje się w [11, 12]. Wiedza dodatkowa może przyczynić się do zmniejszenia błędów generalizacji. Była ona włączana do SVM dla wielu rzeczywistych problemów takich jak np. wyszukiwanie obrazów, [31], rozpoznawanie promotora DNA, [5], prognozowanie raka piersi, [5].

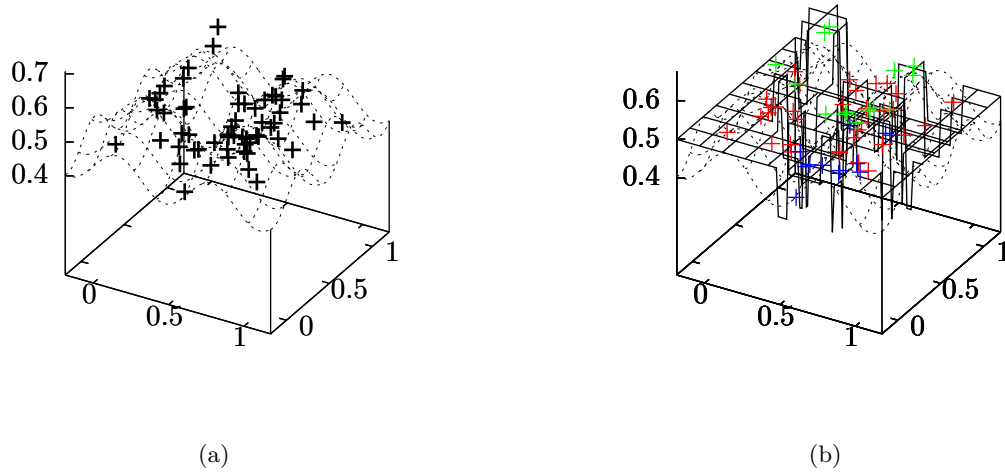
I.1 Przegląd

SVM są metodami maszynowego uczenia się używanymi głównie do rozwiązywania problemów klasyfikacji i regresji. Zostały one opracowane przez Vapnika, [29, 30] w latach 90-tych. Do dzisiaj stały się popularne ze względu na znakomite rezultaty teoretyczne i praktyczne. Wydajność metod maszynowego uczenia się jest głównie oceniana na podstawie następujących kryteriów: generalizacja, szybkość, rzadkość rozwiązania oraz możliwość włączania dodatkowej wiedzy.

Analiza związku między klasyfikacją a regresją jest istotna z punktu widzenia poprawy wydajności istniejących metod. Ponadto, związek między klasyfikacją, a regresją stanowi istotny element teorii uczenia maszynowego. Vapnik w [29] wyprowadził ograniczenia błędu generalizacji dla problemów regresji za pomocą koncepcji zamiany funkcji regresji na zbiór funkcji charakterystycznych. Na bazie tej koncepcji powstały metody, które rozwiązują problem regresji jako problem wieloklasowej klasyfikacji, np. w [9], Rys. I.1b, Rys. I.2b. Odmienna koncepcja zamiany problemów regresji na klasyfikację została zaproponowana przez autora w [19, 21] oraz jednocześnie w [15]. Polega ona na duplikacji punktów i przesunięciu oryginalnych punktów w górę, a zduplikowanych w dół, Rys. ??, Rys. ??. Bazując na tej idei stworzyliśmy nową metodę regresji, zwaną δ -SVR, która posiada wszystkie zalety metod SVM, a pod niektórymi względami okazuje się lepsza. Dodatkowo jedną z praktycznych zalet δ -SVR jest możliwość jej użycia z dowolnymi klasyfikatorami opartymi na funkcjach jądra. Daje to



Rysunek I.1: Transformacja regresji do klasyfikacji wieloklasowej, 2d. (a) Punkty - przykłady regresji sprzed transformacji, linia przerywana - oryginalna funkcja wg której zostały wygenerowane przykłady. (b) punkty - przykłady klasyfikacji po transformacji, linia ciągła - rozwiązanie

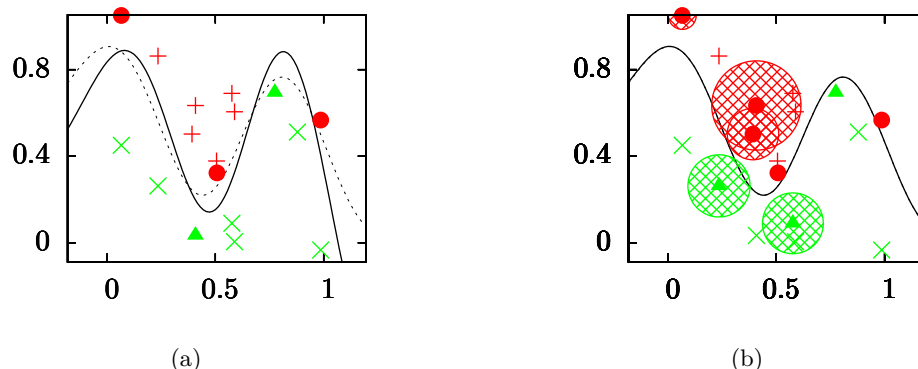


Rysunek I.2: Transformacja regresji do klasyfikacji wieloklasowej, 3d. (a) Punkty - przykłady regresji sprzed transformacji, linia przerywana - oryginalna funkcja wg której zostały wygenerowane przykłady. (b) punkty - przykłady klasyfikacji po transformacji, linia ciągła - rozwiązanie

szerokie możliwości do natychmiastowego zastosowania wszelkich usprawnień i modyfikacji wprowadzanych do metod klasyfikacji, bezpośrednio dla problemów regresji.

Jedną z możliwości poprawy wydajności generalizacji jest włączenie dodatkowej wiedzy do problemu, czasami nazywanej *wiedzą a priori*. Różne typy wiedzy a priori były już włączane do metod SVM. W [11], autorzy dzielą wiedzę a priori na dwa główne typy, wiedzę o niezmiennikach klas, oraz o danych. Pierwsza grupa zawiera m.in. wiedzę na temat klasyfikacji w regionach przestrzeni wejściowej, [5, 4, 16], wiedzę o niezmienniczości klasyfikacji podczas transformacji wejścia. Druga grupa zawiera m.in. wiedzę na temat nieprzyporządkowanych przykładów, braku równowagi między klasami, o jakości danych. Istnieją różne sposoby włączania wiedzy a priori do metod SVM zależne od reprezentacji wiedzy. Możemy wyróżnić trzy główne sposoby:

1. modyfikacja danych wejściowych takich jak zbiór cech, wartości parametrów wejściowych
2. modyfikacja metody SVM



Rysunek I.3: Porównanie rozwiązań z wiedzą marginesową i bez, 2d

Porównanie rozwiązań z wiedzą marginesową i bez, 2d. Punkty - przykłady, trójkąty i okręgi - wektory wspierające, linie ciągłe - rozwiązania, linie kropkowane - oryginalne funkcje wg których zostały wygenerowane punkty. (b) Koła wypełnione siatką - wiedza marginesowa per przykład

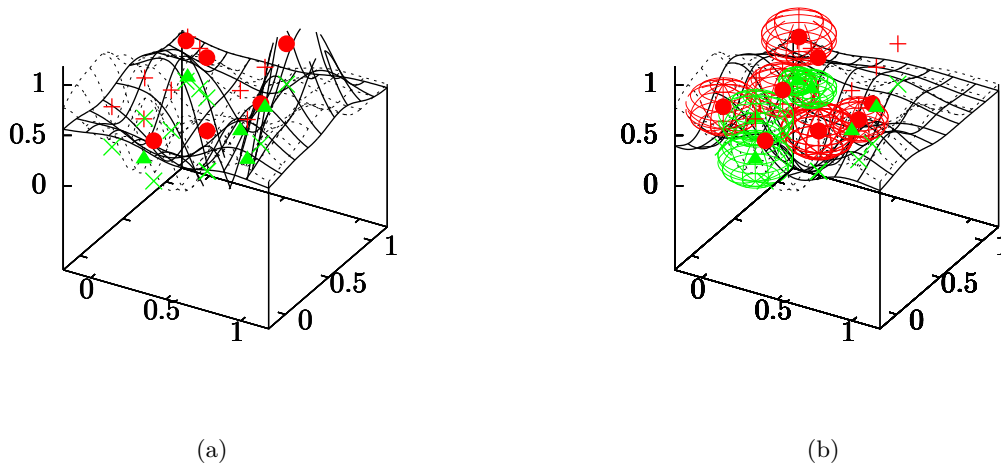
3. modyfikacja wyjścia.

Druga z metod sprowadza się do modyfikacji problemu optymalizacyjnego SVM, w szczególności do modyfikacji funkcji celu, zmiany ograniczeń, modyfikacji funkcji jądra.

W tej pracy analizujemy niedawno zaproponowaną przeze mnie wiedzę o marginesie przykładu, [18, 20, 22]. Jest ona ściśle związana ze sformułowaniem problemu optymalizacyjnego SVM. Wiedza ta per przykład może być interpretowana w uproszczeniu jako region w postaci hipersferycznego otoczenia przykładu o zmiennym promieniu zależnym od wielkości marginesu, z przyporządkowaną klasą, Rys. I.3, Rys. I.4. Regiony, które były już wcześniej włączane do SVM to regiony wielościanowe, [5, 4, 10, 33], regiony elipsoidalne, w tym hipersfery, [25] oraz regiony nieliniowe, [16]. Wiedza marginesowa włączona jest do SVM za pomocą uogólnienia problemu optymalizacyjnego SVM. Podobna idea była użyta do zdefiniowania γ -rozdzielenia (γ -shattering) w statystycznej teorii maszynowego uczenia. Pokażemy, że metoda ε -SVR może być traktowana jako problem klasyfikacji z wiedzą marginesową, [22]. Głównym zastosowaniem dla wiedzy marginesowej rozważanym w tej pracy jest zmniejszenie stopnia komplikacji rozwiązania, co oznacza mniejszą liczbę wektorów wspierających SVM. Prostsze rozwiązanie oznacza prostszą interpretację problemu, oraz zmniejszony czas testowania nowych przykładów. Drugim ważnym zastosowaniem wiedzy marginesowej zaproponowanym w tej pracy, jest możliwość wyrażenia innych szczególnych typów wiedzy a priori w tej formie. Pokażemy, że wiedza w postaci sumy wartości szukanej funkcji dla pewnych danych, może być efektywnie włączona do problemu SVM za pomocą dodatkowego warunku, i dalej jako wiedza marginesowa.

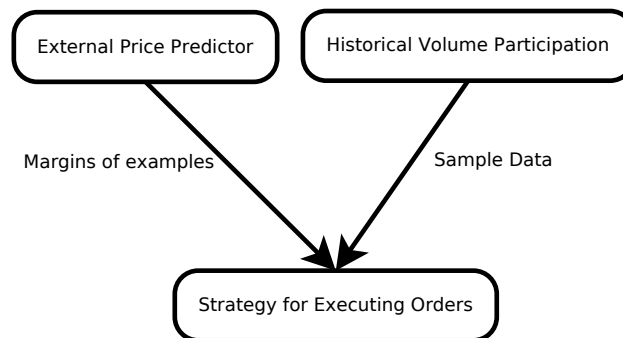
Istotnym zagadnieniem włączania wiedzy a priori jest możliwość włączania wiedzy w sposób płynny, co umożliwia operowanie *niepewną dodatkową wiedzą*, [17]. Włączanie wiedzy marginesowej jako dodatkowych parametrów do warunków nierównościowych problemu optymalizacyjnego SVM umożliwia otrzymanie rozwiązań będących kompromisem między wydajnością klasyfikacji danych oraz spełnienia warunków wiedzy marginesowej.

Niedawno, algorytmiczne składanie zleceń giełdowych stało się popularne ze względu na rozwój przemysłu komputerowego. Zlecenia są generowane automatycznie przez system inwestycyjny i wysyłane do system zarządzania zleceniami (OMS) który przekazuje zlecenia na giełdy. Jednym z zadań OMS jest efektywny podział zlecenia na mniejsze części i rozłożenie wysyłania zleceń w czasie. Jedną z metod oceny wydajności wykonania zlecenia jest porównanie ceny wykonania zlecenia z innymi uczestnikami rynku. Do tego celu służy m.in. miara zwana średnia cena ważona wolumenem (VWAP). Niedawno zaproponowano prosty model teoretyczny osiągający stosunek miar VWAP równy 1, [2]. W praktyce osiągnięcie tego wyniku zależy w dużej mierze od jakości przewidywania wskaźnika partycypacji w wolumenie. Predykcja ta sprowadza się do problemu regresji z dodatkowymi warunkami



Rysunek I.4: Porównanie rozwiązań z wiedzą marginesową i bez, 3d

Porównanie rozwiązań z wiedzą marginesową i bez, 3d. Punkty - przykłady, trójkąty i okręgi - wektory wspierające, linie ciągłe - rozwiązania, linie kropkowane - oryginalne funkcje wg których zostały wygenerowane punkty. (b) Koła wypełnione siatką - wiedza marginesowa per przykład



Rysunek I.5: Model danych dla strategii wykonywania zleceń giełdowych. Wiedza w postaci marginesu per przykład jest wykorzystywana do włączania informacji o predykcji cen. Wraz z historycznymi danymi o partycypacji w wolumenie są używane w strategii składania zleceń giełdowych

na rozwiązanie. W modelach osiągających lepszy stosunek miar dochodzi konieczność przewidywania dodatkowo cen instrumentów finansowych. Niedawno zaproponowaliśmy zastosowanie SVM wraz z wiedzą marginesową do systemu przewidującego partycypację w wolumenie oraz dodatkowo umożliwiającego dołączenie przewidywania cen instrumentów finansowych, [23].

Hipotezami pracy są:

1. Zaproponowana metoda regresji δ -SVR pozwala na zmniejszenie liczby wektorów wspierających oraz poprawienie elastyczności SVM.
2. Wiedza o marginesie przykładu prowadzi do zmniejszenia błędu generalizacji podczas tworzenia modeli zredukowanych dla problemów klasyfikacji i regresji.

3. Zaproponowana heurystyka alternatyw HoA prowadzi do przyspieszenia heurystyki w metodzie SMO.
4. Zaproponowana sekwencyjna metoda rozwiązywania podproblemów wielowymiarowych SMS pozwala na ominięcie konieczności używania ogólnych bibliotek do programowania kwadratowego w algorytmie SVM.
5. SVM z dodatkową wiedzą prowadzą do zmniejszenia błędu generalizacji dla predykcji funkcji partycypacji w wolumenie w metodach optymalizujących koszt wykonania zleceń giełdowych.

Opis rozdziałów. Szczegółowy opis wspomnianych we wstępie metod znajduje się w odpowiednich rozdziałach. W pierwszym rozdziale prezentujemy krótko klasyfikacja wektorów wspierających (SVC), SVR, oraz włączanie dodatkowej wiedzy do SVM. W drugim rozdziale prezentujemy metodę δ -SVR. W trzecim rozdziale prezentujemy metodę φ -SVC. W czwartym rozdziale prezentujemy dwa usprawnienia implementacji SVM. W piątym rozdziale prezentujemy zastosowanie SVM do wykonywania zleceń giełdowych.

I.2 Problemy optymalizacyjne SVC

Dla problemu klasyfikacji, rozważamy zbiór n wektorów treningowych \vec{x}_i dla $i \in \{1, \dots, n\}$, gdzie $\vec{x}_i = (x_i^1, \dots, x_i^m)$, i -ty wektor treningowy jest zmapowany do $y_i^c \in \{-1, 1\}$, m to wymiar problemu. Problem optymalizacyjny SVC dla przypadku marginesu mocnego z normą to

OP 1.

$$\min_{\vec{w}_c, b_c} f(\vec{w}_c, b_c) = \|\vec{w}_c\|^2 \quad (\text{I.1})$$

z warunkami

$$y_i^c h(\vec{x}_i) \geq 1 \quad (\text{I.2})$$

dla $i \in \{1, \dots, n\}$, gdzie

$$h(\vec{x}_i) = \vec{w}_c \cdot \vec{x}_i + b_c \quad (\text{I.3})$$

Wszystkie punkty muszą być poprawnie zaklasyfikowane $\|\cdot\|_1$ Rys. I.6a Problem optymalizacyjny SVC dla przypadku marginesu słabego z normą $\|\cdot\|_1$ to

OP 2.

$$\min_{\vec{w}_c, b_c, \vec{\xi}_c} f(\vec{w}_c, b_c, \vec{\xi}_c) = \|\vec{w}_c\|^2 + C_c \sum_{i=1}^n \xi_c^i \quad (\text{I.4})$$

z warunkami

$$y_i^c h(\vec{x}_i) \geq 1 - \xi_c^i \quad (\text{I.5})$$

$$\vec{\xi}_c \geq 0 \quad (\text{I.6})$$

dla $i \in \{1, \dots, n\}$, gdzie

$$h(\vec{x}_i) = \vec{w}_c \cdot \vec{x}_i + b_c \quad (\text{I.7})$$

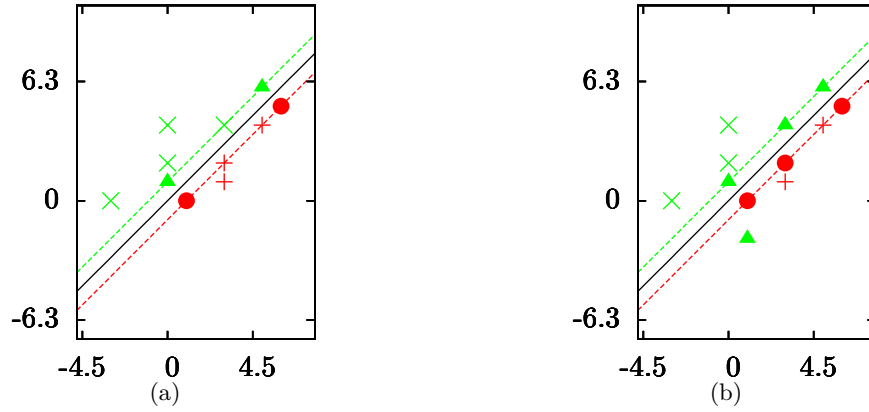
$h^*(\vec{x}) = \vec{w}_c^* \cdot \vec{x} + b_c^* = 0$ to krzywa decyzyjna problemu klasyfikacji. Niektóre punkty treningowe mogą być niepoprawnie zaklasyfikowane Rys. I.6b.

I.2.1 Dualny problem optymalizacyjny SVC

Problem PO 2 po transformacji do równoważnego problemu dualnego przyjmuje postać:

OP 3.

$$\max_{\vec{\alpha}} f(\vec{\alpha}) = 1 \cdot \vec{\alpha} - \frac{1}{2} \alpha^T \mathbf{Q} \vec{\alpha} \quad (\text{I.8})$$



Rysunek I.6: Dwa typy klasyfikatorów marginesowych: mocny i słaby. Punkty przykładowe, wektory wspierające (trójkąty i koła), rozwiązania (linie ciągłe), linie marginesowe (linie przerywane). (a) Mocny. (b) Słaby. Punkt niepoprawnie zaklasyfikowany jest w (1, -2)

z warunkami

$$\vec{\alpha} \cdot \vec{y} = 0 \quad (\text{I.9})$$

$$0 \leq \alpha_i \leq C \quad (\text{I.10})$$

gdzie

$$Q_{ij} = y_i y_j K(\vec{x}_i, \vec{x}_j) \quad (\text{I.11})$$

dla każdego $i, j \in \{1, \dots, n\}$.

Najbardziej popularne funkcje jądra to liniowa, wielomianowa, radialne funkcje bazowe (RBF) i sigmoidalna. Funkcja jądra będąca iloczynem skalarnym swoich argumentów nazywamy *prostą funkcją liniową*. Krzywa decyzyjna ma postać

$$h^*(\vec{x}) = \sum_{i=1}^n y_i^* \alpha_i^* K(\vec{x}_i, \vec{x}) + b_c^* = 0, \quad (\text{I.12})$$

gdzie α_i to mnożniki Lagrange'a problemu dualnego, $K(\cdot, \cdot)$ to funkcja jądra, która występuje tylko w problemie dualnym. *Granice marginesowe* są zdefiniowane jako dwie hiperpłaszczyzny $h(\vec{x}) = -1$ i $h(\vec{x}) = 1$. *Optymalne granice marginesowe* są zdefiniowane jako dwie hiperpłaszczyzny $h^*(\vec{x}) = -1$ i $h^*(\vec{x}) = 1$. i -ty przykład treningowy jest *wektorem wspierającym*, jeśli $\alpha_i^* \neq 0$. Można udowodnić, że zbiór wektorów wspierających zawiera wszystkie przykłady treningowe, które są położone poniżej optymalnych granic marginesowych ($y_c^i h^*(\vec{x}_i) < 1$), oraz większość przykładów położonych dokładnie na optymalnych granicach marginesowych ($y_c^i h^*(\vec{x}_i) = 1$).

Warunki dopełniające KKT dla PO 2 to

$$\alpha_i (y_c^i h(\vec{x}_i) - 1 + \xi_c^i) = 0 \quad (\text{I.13})$$

$$(C_c - \alpha_i) \xi_c^i = 0. \quad (\text{I.14})$$

Możemy znaleźć wartości parametrów ξ_i z rozwiązania problemu dualnego jak następuje. Kiedy

$$y_c^i h^*(\vec{x}_i) \geq 1, \quad (\text{I.15})$$

to $\xi_i = 0$, w przeciwnym razie

$$\xi_i = 1 - y_c^i h^*(\vec{x}_i). \quad (\text{I.16})$$

I.2.2 ν -SVC

Innym wariantem SVC jest ν klasyfikacja wektorów wspierających (ν -SVC) gdzie zastępujemy C , $\nu \in [0, 1]$. Zmodyfikowany problem optymalizacyjny to

OP 4.

$$\min_{\vec{w}, b, \vec{\xi}, p} f(\vec{w}, b, \vec{\xi}, p) = \frac{1}{2} \|\vec{w}\|^2 - \nu p + \frac{1}{n} \sum_{i=1}^n \xi_i \quad (\text{I.17})$$

z warunkami

$$y_i h(\vec{x}_i) \geq p - \xi_i \quad (\text{I.18})$$

$$\vec{\xi} \geq 0 \quad (\text{I.19})$$

$$p \geq 0 \quad (\text{I.20})$$

dla $i \in \{1, \dots, n\}$, gdzie

$$\vec{C} \gg 0 \quad (\text{I.21})$$

$$h(\vec{x}_i) = \vec{w} \cdot \vec{x}_i + b \quad (\text{I.22})$$

Możemy zauważyć inną funkcję kosztu, dodatkową zmienną p oraz dodatkowy warunek.

I.3 Problemy optymalizacyjne SVR

W problemie regresji, rozważamy zbiór wektorów treningowych \vec{x}_i for $i \in \{1, \dots, n\}$, gdzie $\vec{x}_i = (x_i^1, \dots, x_i^m)$, i -ty wektor treningowy jest zmapowany do $y_i^i \in \mathbb{R}$, m jest wymiarem problemu. Problem optymalizacyjny ε -SVR dla przypadku słabego to

OP 5.

$$\min_{\vec{w}_r, b_r, \vec{\xi}_r, \vec{\xi}_r^*} f(\vec{w}_r, b_r, \vec{\xi}_r, \vec{\xi}_r^*) = \|\vec{w}_r\|^2 + C_r \sum_{i=1}^n (\xi_r^i + \xi_r^{*i}) \quad (\text{I.23})$$

z warunkami

$$y_r^i - g(\vec{x}_i) \leq \varepsilon + \xi_r^i \quad (\text{I.24})$$

$$g(\vec{x}_i) - y_r^i \leq \varepsilon + \xi_r^{*i} \quad (\text{I.25})$$

$$\vec{\xi}_r \geq 0 \quad (\text{I.26})$$

$$\vec{\xi}_r^* \geq 0 \quad (\text{I.27})$$

dla $i \in \{1, \dots, n\}$, gdzie

$$g(\vec{x}_i) = \vec{w}_r \cdot \vec{x}_i + b_r \quad (\text{I.28})$$

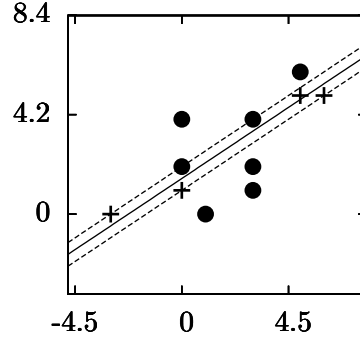
$g^*(\vec{x}) = \vec{w}_r^* \cdot \vec{x} + b_r^*$ to funkcja regresji. Problem optymalizacyjny 5 transformujemy do równoważnego problemu dualnego. Funkcja regresji przyjmuje postać

$$g^*(\vec{x}) = \sum_{i=1}^n (\alpha_i^* - \beta_i^*) K(\vec{x}_i, \vec{x}) + b_r^* \quad (\text{I.29})$$

gdzie α_i, β_i to mnożniki Lagrange'a, $K(\cdot, \cdot)$ to funkcja jądra. i -ty przykład treningowy jest *wektorem wspierającym*, jeśli $\alpha_i^* - \beta_i^* \neq 0$. Zbiór wektorów wspierających zawiera wszystkie wektory treningowe leżące poza granicami ε , oraz większość przykładów położonych dokładnie na granicach ε Rys. I.7. Liczba wektorów wspierających może być kontrolowana za pomocą parametru ε .

I.4 Włączanie wiedzy a priori do SVM

Różne schematy włączania wiedzy a priori do problemów optymalizacyjnych SVM były już zaproponowane. Należą one do następujących kategorii:



Rysunek I.7: Idea metody ε -SVR. Przykłady, wektory wspierające (koła), rozwiązanie (linia ciągła), oraz granice ε (linie przerywane)

1. zmodyfikowana funkcja kosztu (albo pierwotna albo dualna),
2. zmodyfikowane warunki (albo pierwotne albo dualne),
3. nowe warunki (albo pierwotne albo dualne),

Poza wymienionymi kategoriami, mogą być zmodyfikowane dane wejściowe, rozwiązanie, a także funkcja jądra. Niektóre włączenia wymagają wielu zmian problemu optymalizacyjnego, nowych zmiennych lub parametrów problemu. Przegląd metod włączania wiedzy a priori do SVM znajduje się w [11, 12]. W tym raporcie, koncentrujemy się na wiedzy per przykład, to znaczy wiedzy związanej z poszczególnymi przykładami. Analizujemy w szczególności wiedzę per przykład w formie dodatkowych parametrów do problemu optymalizacyjnego dla każdego przykładu.

Jednym z typów wag dla przykładów są wagi oznaczające różne koszty błędnej klasyfikacji C_i per przykład. Szczególnym przypadkiem jest różny koszt błędnej klasyfikacji dla ujemnych i dodatnich przykładów treningowych, C_+ and C_- używany do włączania informacji o niezbalansowanych danych dla C maszyny wektorów wspierających (C-SVM), [8], dla ν maszyny wektorów wspierających (ν -SVM), [32]. Wagi C_i były również używane do rozmytych maszyn wektorów wspierających, [14], dla oznaczania pewności pseudo etykiet, [31] oraz oznaczania pewności zależnej od jakości danych, [34]. Problem optymalizacyjny SVC we wersji słabej z normą $\|\cdot\|_1$ dla przykładów treningowych \vec{x}_i z wagami C_i to

OP 6.

$$\min_{\vec{w}, b, \vec{\xi}} f(\vec{w}, b, \vec{\xi}) = \frac{1}{2} \|\vec{w}\|^2 + \vec{C}_c \cdot \vec{\xi} \quad (\text{I.30})$$

z warunkami

$$y_i h(\vec{x}_i) \geq 1 - \xi_i \quad (\text{I.31})$$

$$\vec{\xi} \geq 0 \quad (\text{I.32})$$

dla $i \in \{1, \dots, n\}$, gdzie

$$\vec{C}_c \gg 0 \quad (\text{I.33})$$

$$h(\vec{x}_i) = \vec{w} \cdot \vec{x}_i + b \quad (\text{I.34})$$

Wagi C_i były również używane z ε -SVR do predykcji szeregów czasowych, [28]. Innym typem wag, które są używane z ε -SVR są wagi ε_i per przykład zastępujące parameter ε . Były używane do estymacji gęstości, [30]. Problem optymalizacyjny z dodatkowymi wagami ε_i to

OP 7.

$$\min_{\vec{w}_r, b_r, \vec{\xi}_r, \vec{\xi}_r^*} f(\vec{w}_r, b_r, \vec{\xi}_r, \vec{\xi}_r^*) = \|\vec{w}_r\|^2 + C_r \sum_{i=1}^n (\xi_r^i + \xi_r^{*i}) \quad (\text{I.35})$$

z warunkami

$$y_r^i - g(\vec{x}_i) \leq \varepsilon_i + \xi_r^i \quad (\text{I.36})$$

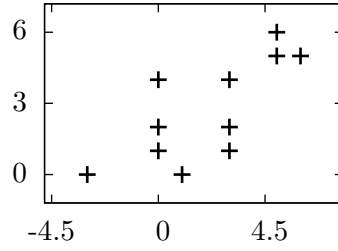
$$g(\vec{x}_i) - y_r^i \leq \varepsilon_i + \xi_r^{i*} \quad (\text{I.37})$$

$$\vec{\xi}_r \geq 0 \quad (\text{I.38})$$

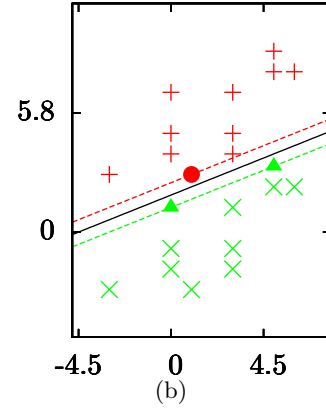
$$\vec{\xi}_r^* \geq 0 \quad (\text{I.39})$$

dla $i \in \{1, \dots, n\}$, gdzie

$$g(\vec{x}_i) = \vec{w}_r \cdot \vec{x}_i + b_r \quad (\text{I.40})$$



(a)



(b)

Rysunek II.1: Idea transformacji problemu w δ -SVR dla 2d. (a) Punkty - przykłady regresji. (b) Punkty - przykłady klasyfikacji po transformacji, trójkąty i koła - wektory wspierające, linia ciągła - rozwiązanie, linie przerywane - linie marginesowe

Rozdział II

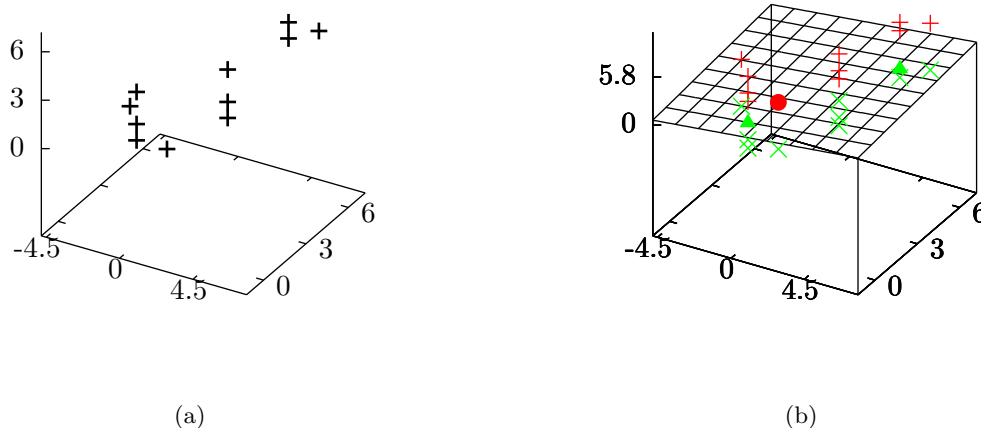
Regresja bazująca na klasyfikacji binarnej

II.1 Wstęp do metody δ -SVR

Rozważmy zbiór wektorów treningowych \vec{x}_i dla $i \in \{1, \dots, n\}$, gdzie $\vec{x}_i = (x_i^1, \dots, x_i^m)$, i -ty wektor treningowy jest zmapowany do $y_i^i \in \mathbb{R}$. Metoda δ -SVR bazuje na następującym schemacie wyszukiwania funkcji regresji:

1. Każdy przykład treningowy \vec{x}_i jest zduplikowany, wartość wyjściowa y_i^i jest zwiększana o wartość parametru $\delta \geq 0$ dla oryginalnych przykładów treningowych, oraz zmniejszana o δ dla zduplikowanych przykładów treningowych.
2. Każdy przykład treningowy \vec{x}_i jest konwertowany do przykładu klasyfikacyjnego poprzez włączenie wyjścia do wektora wejściowego jako dodatkowej cechy i ustawienie klasy 1 dla oryginalnych przykładów treningowych, i klasy -1 dla zduplikowanych przykładów treningowych.
3. Metoda SVC jest uruchamiana dla problemu klasyfikacji.
4. Rozwiązanie metody SVC jest konwertowane z powrotem do formy funkcyjnej.

Idea transformacji jest przedstawiona na rysunkach Rys. II.1, Rys. II.2.



Rysunek II.2: Idea transformacji problemu w δ -SVR dla 3d. (a) Punkty - przykłady regresji. (b) Punkty - przykłady klasyfikacji po transformacji, trójkąty i koła - wektory wspierające, linia ciągła - rozwiązanie

II.2 Podsumowanie

W tym raporcie przedstawiliśmy nowoczesną metodę regresji, zwaną δ -SVR. Wykonaliśmy eksperymenty porównujące δ -SVR z ε -SVR na syntetycznych i rzeczywistych zbiorach danych. Rezultaty wskazują, że δ -SVR osiąga podobny błąd generalizacji. Pierwszą zaletą δ -SVR jest mniejsza liczba wektorów wspierających. Tym samym rezultatem są prostsze modele predykcyjne. Przekłada się to na zmniejszenie czasu testowania nowych przykładów. Kolejną zaletą jest mniejszy błąd generalizacji dla różnych wartości ε i δ . Tym samym istnieje możliwość przyspieszenia czasu trenowania, ale z akceptacją rozwiązań suboptymalnych. Kolejną zaletą jest szybszy czas trenowania dla liniowych jąder używając metody SMO. Ostatnią zaletą δ -SVR jest możliwość zastąpienia standardowej klasyfikacji dowolną metodą klasyfikacji bazującą na funkcjach jądra. W szczególności, dowolne udoskonalenia metod klasyfikacji pod względem błędu generalizacji, szybkości trenowania i testowania, możliwości włączania wiedzy a priori mogą być bezpośrednio użyte dla problemów regresji.

Wadą zaproponowanej metody są niejednoznaczne wyniki porównujące czas trenowania dla nieliniowych jąder. Dla niektórych zbiorów danych δ -SVR jest wolniejsza od ε -SVR.

W przyszłych badaniach planujemy przetestować δ -SVR z różnymi parametrami przesunięcia w górę i w dół. Planujemy również wykonać testy dla zbiorów danych zawierających błędy nie tylko na wyjściu, ale również w wektorach wejściowych.

Rozdział III

Wiedza marginesowa per przykład

III.1 Wprowadzenie do φ -SVC

Definiujemy margines w następujący sposób:

Definicja III.1.1. Mając pewną krzywą $h(\vec{x}) = 0$, *margines dla pewnego przykładu* \vec{x}_p jest zdefiniowany jako wartość $|h(\vec{x}_p)|$.

Dla SVC z mocnym marginesem, PO 1, margines dla najbliższych przykładów jest równy 1. *Wiedza marginesowa per przykład* jest zdefiniowana jako informacja a priori o marginesie dla pewnych przykładów. Metoda φ -SVC jest niedawno zaproponowaną metodą włączania wiedzy marginesowej per przykład do SVC, [18, 20, 22]. Problem optymalizacyjny φ -SVC jest zdefiniowany z dodatkowym parametrem per przykład dodanym do prawej strony nierówności (I.5). Inna modyfikacja warunków nierównościowych została zaproponowana w [34]. Autorzy modyfikują nierówności za pomocą pomnożenia lewej strony nierówności przez pewną funkcję malejącą opartą na dodatkowych wagach. Metoda φ -SVC jest koncepcją ogólniejszą wag per przykład o dowolnych możliwych wartościach i z odmienną interpretacją.

Teraz, przyjrzymy się bliżej problemowi optymalizacyjnemu φ -SVC. Bazuje on na problemie SVC z wagami kosztowymi per przykład, PO 6. Problem optymalizacyjny φ -SVC jest następujący

OP 8.

$$\min_{\vec{w}_c, b_c, \vec{\xi}_c} f(\vec{w}_c, b_c, \vec{\xi}_c) = \frac{1}{2} \|\vec{w}_c\|^2 + \vec{C}_c \cdot \vec{\xi}_c \quad (\text{III.1})$$

z warunkami

$$y_c^i h(\vec{x}_i) \geq 1 + \varphi_i - \xi_c^i \quad (\text{III.2})$$

$$\vec{\xi} \geq 0 \quad (\text{III.3})$$

dla $i \in \{1, \dots, n\}$, gdzie

$$\vec{C}_c \gg 0 \quad (\text{III.4})$$

$$\varphi_i \in \mathbb{R} \quad (\text{III.5})$$

$$h(\vec{x}_i) = \vec{w}_c \cdot \vec{x}_i + b_c \quad (\text{III.6})$$

Nowe wagi φ_i są obecne tylko w (III.2). Kiedy $\vec{\varphi} = 0$, PO 8 jest równoważne z PO 6. Kiedy wszystkie φ_i są równe pewnej stałej $\varphi > -1$, otrzymujemy tę samą krzywą decyzyjną co dla $\varphi = 0$ kiedy zmienimy \vec{C}_c to $\vec{C}_c / (1 + \varphi)$.

Możemy zauważyć, że pomijając ξ_i , dostajemy różne ograniczenia dla marginesu: kiedy $y_c^i = 1$ i $g(\vec{x}_i) \geq 0$, wtedy dostajemy dolną granicę $1 + \varphi_i$, kiedy $y_c^i = -1$ i $g(\vec{x}_i) \geq 0$, wtedy dostajemy górną granicę $-(1 + \varphi_i)$, kiedy $y_c^i = 1$ i $g(\vec{x}_i) < 0$, wtedy dostajemy górną granicę $-(1 + \varphi_i)$, kiedy $y_c^i = -1$ i $g(\vec{x}_i) < 0$, wtedy dostajemy dolną granicę $1 + \varphi_i$. Możemy rozróżnić trzy przypadki: $1 + \varphi_i > 0$, $1 + \varphi_i < 0$ i $1 + \varphi_i = 0$. Dla pierwszego z nich, dostajemy dolną granicę na margines równą $1 + \varphi_i$. Dla

drugiego, dostajemy górną granicę na margines równą $-(1 + \varphi_i)$, dla trzeciego, nie mamy żadnych ograniczeń na margines. Dlatego też, używając wag φ_i , możemy włączyć wiedzę o marginesie.

Następną właściwością wag φ_i jest to, że mają one wpływ na odległość pomiędzy krzywą $h(\vec{x}) = 0$ i i -tym przykładem. Możemy przeprowadzić tą samą analizę co wyżej dla odległości za pomocą podzielenia obu stron (III.2) przez $\|\vec{w}_c\|$, i dostajemy $y_c^i h(\vec{x}_i) / \|\vec{w}_c\| \geq 1 / \|\vec{w}_c\| + \varphi_i / \|\vec{w}_c\|$. Wniosek jest podobny: dla przypadku gdy $1 + \varphi_i > 0$, dostajemy dolną granicę na odległość równą $(1 + \varphi_i) / \|\vec{w}_c\|$. Dla przykładu gdy $1 + \varphi_i < 0$, dostajemy górną granicę na odległość równą $-(1 + \varphi_i) / \|\vec{w}_c\|$. Dla przykładu gdy $1 + \varphi_i = 0$, nie mamy żadnych ograniczeń na odległość.

Zauważmy, że jeśli weźmiemy pod uwagę ξ_i , możemy zauważyć, że wiedza marginesowa jest włączona jako wiedza niepewna. Naruszenie wiedzy marginesowej jest kontrolowane za pomocą parametru C_i . Porównując nieformalnie wagi φ_i ze zmiennymi luźnymi (ξ_i): wagi φ_i są stałe, są nieobecne we funkcji celu, zaś suma luźnych zmiennych jest minimalizowana jako część funkcji celu.

Możemy także wyprowadzić równoważny problem optymalizacyjny do PO 8, gdzie wagi φ_i są obecne w ograniczeniach na luźne zmienne

OP 9.

$$\min_{\vec{w}_c, b_c, \vec{\xi}_c} f(\vec{w}_c, b_c, \vec{\xi}_c) = \frac{1}{2} \|\vec{w}_c\|^2 + \vec{C}_c \cdot \vec{\xi}_c \quad (\text{III.7})$$

z warunkami

$$y_c^i h(\vec{x}_i) \geq 1 - \xi_i^i \quad (\text{III.8})$$

$$\xi_i \geq \varphi_i \quad (\text{III.9})$$

dla $i \in \{1, \dots, n\}$.

III.2 Podsumowanie

W tym raporcie przeanalizowaliśmy zastosowanie wiedzy marginesowej per przykład włączonej do SVM, do zmniejszenia błędu generalizacji modeli zredukowanych. Wykonaliśmy testy dla klasyfikatora SVM, oraz dla metod regresji ε -SVR i δ -SVR. Eksperymenty na danych rzeczywistych pokazują mniejszy błąd generalizacji dla modeli zredukowanych z wiedzą marginesową. Możliwości wykorzystania wiedzy marginesowej są dużo szersze. W przyszłych badaniach planujemy zbadać możliwość tworzenia wiedzy marginesowej przez ekspertów, a także wykorzystanie wiedzy marginesowej do szeregów czasowych. Ponadto planujemy wykorzystać wiedzę marginesową do łączenia ze sobą klasyfikatorów SVM w celu zmniejszenia błędu generalizacji.

Rozdział IV

Techniki implementacyjne bazujące na warunkach KKT

Jedną z kategorii metod rozwiązywania problemu optymalizacyjnego PO 3 są metody wyboru podzbioru. W każdej iteracji jedynie kilka mnożników Lagrange’a jest optymalizowanych. Szczególnym przypadkiem jest metoda SMO zaproponowana w [24], która rozwiązuje analitycznie podproblemy 2 parametrowe w każdej iteracji. Dla podproblemów więcej niż 2 parametrowych używane są zaawansowane metody ogólne do rozwiązywania problemów kwadratowych. W pierwszej części tego raportu proponujemy użycie SMO do rozwiązywania podproblemów więcej niż 2 parametrowych. Zaletą takiego rozwiązania jest prostsza metoda bez zewnętrznych ogólnych metod programowania kwadratowego.

Istotną częścią metod wyboru podzbioru jest strategia wyboru parametrów w każdej iteracji. Najbardziej popularna strategia bazuje na kryterium KKT. Proponujemy strategię, która oblicza wartość funkcji celu w każdej iteracji dla kilku alternatywnych par parametrów. Wybierana jest para, dla której uzyskaliśmy największy spadek wartości funkcji celu. Zaletą tej strategii jest zmniejszona liczba iteracji oraz możliwość efektywnego zrównoleglenia algorytmu.

Wszystkie zaproponowane metody mogą być stosowane dla SVC, dlatego też również dla δ -SVR. Również mogą być stosowane z φ -SVC, dlatego też również z ε -SVR.

IV.1 Podsumowanie

W tym raporcie przeanalizowaliśmy dwa usprawnienia implementacyjne SVM, pierwsze w szybkości trenowania SVM, drugie w uproszczeniu implementacji SVM. Testy na rzeczywistych zbiorach danych pokazują, że metoda HoA może prowadzić do zmniejszenia czasu trenowania SVM, porównując do standardowej heurystyki. Stosując metodę SMS, otrzymujemy prostszą implementację SVM, z podobnymi wynikami szybkościowymi. Obie metody mogą być stosowane z δ -SVR i ε -SVR do rozwiązywania problemów regresji.

Rozdział V

Aplikacje: strategie wykonywania zleceń

Duże zlecenia giełdowe nie mogą być wykonane na giełdzie na raz ponieważ liczba ofert po przeciwnej stronie jest ograniczona. Muszą być podzielone na mniejsze zlecenia, których wykonanie zostanie rozłożone w czasie. Istnieje wiele miar jakości wykonania zleceń giełdowych. Najbardziej popularne to VWAP rynkowy, tzn. dla wszystkich zleceń, cena instrumentu tuż przed wykonaniem zlecenia, cena instrumentu po wykonaniu zlecenia, wszystkie wartości w porównaniu do VWAP zlecenia. W tym raporcie, analizujemy pierwszą miarę. Model strategii osiągającej VWAP rynkowy został zaproponowany niedawno w [1, 2]. W [1], autorzy stwierdzili, że poprawa jakości przewidywania wolumenu prowadzi do lepszego wykonania zleceń, jakkolwiek znaleźli oni przeciwstawne wyniki w [6].

Celem przeprowadzonych badań było rozszerzenie wyników teoretycznych dla strategii wykonania zleceń osiągających VWAP oraz pokazanie od jakich czynników zależy ostateczny błąd realizacji zleceń. Ponadto chcieliśmy zaimplementować część strategii za pomocą ogólnej metody maszynowego uczenia, SVR.

V.1 Podsumowanie

W tym raporcie przeanalizowaliśmy zastosowanie SVR do wykonywania zleceń na rynkach akcyjnych. Porównaliśmy ϵ -SVR i δ -SVR z prostymi predyktorami takimi jak średnia cena egzekucji z poprzednich dni. Testy zostały wykonane na danych dla spółek z indeksu NASDAQ-100. Dla obu metod otrzymaliśmy mniejszą wariancję kosztów wykonania. Ponadto, zmniejszyliśmy koszty wykonania zleceń wykorzystując dodatkowo predykcję cen giełdowych.

W przyszłych badaniach planujemy wykonać testy na szerszym zakresie spółek i giełd.

Rozdział VI

Podsumowanie

Głównym wkładem w tą pracę jest

1. zaproponowanie nowej metody regresji, która działa wydajnie również dla rozwiązań nieliniowych, analiza teoretyczna i praktyczna
2. zaproponowanie wiedzy marginesowej per przykład dla klasyfikacji i regresji, analiza teoretyczna i praktyczna
3. użycie wiedzy marginesowej per przykład do zmniejszenia liczby wektorów wspierających
4. zaproponowanie dwóch usprawnień do algorytmu SVM bazującego na warunkach KKT
5. zastosowanie SVM do wykonywania zleceń giełdowych

W przyszłych badaniach planujemy rozszerzyć analizę teoretyczną zaproponowanych rozwiązań.

Bibliografia

- [1] Jędrzej Białkowski, Serge Darolles, and Gaelle Le Fol. Improving vwap strategies: A dynamic volume approach. *Journal of Banking & Finance*, 32(9):1709–1722, September 2008. [15](#)
- [2] Christian T. Brownlees, Fabrizio Cipollini, and Giampiero M. Gallo. Intra-daily volume modeling and prediction for algorithmic trading. Econometrics working papers archive, Università degli Studi di Firenze, Dipartimento di Statistica "G. Parenti", February 2009. [3](#), [15](#)
- [3] Nianyi Chen, Wencong Lu, Jie Yang, and Guozheng Li. *Support Vector Machine in Chemistry*. World Scientific Publishing Co., 2004. [1](#)
- [4] Glenn M. Fung, Olvi L. Mangasarian, and Jude W. Shavlik. Knowledge-based nonlinear kernel classifiers. In *Learning Theory and Kernel Machines*, Lecture Notes in Computer Science, pages 102–113. Springer Berlin / Heidelberg, 2003. [2](#), [3](#)
- [5] Glenn M. Fung, Olvi L. Mangasarian, and Jude W. Shavlik. Knowledge-based support vector machine classifiers. In S. Thrun S. Becker and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15*, pages 521–528. MIT Press, Cambridge, MA, 2003. [1](#), [2](#), [3](#)
- [6] D. Hobson. Vwap and volume profiles. *Journal of Trading*, 1(2):38–42, 2006. [15](#)
- [7] Te-Ming Huang, Vojislav Kecman, and Ivica Kopriva. *Kernel Based Algorithms for Mining Huge Data Sets*. Springer, 2006. [1](#)
- [8] Thorsten Joachims. *Learning to Classify Text Using Support Vector Machines*. Kluwer Academic Publishers, 2002. [1](#), [8](#)
- [9] Vojislav Kecman and Tao Yang. Adaptive local hyperplane for regression tasks. In *Proceedings of the 2009 international joint conference on Neural Networks, IJCNN'09*, pages 2371–2375, Piscataway, NJ, USA, 2009. IEEE Press. [1](#)
- [10] Gautam Kunapuli, Kristin P. Bennett, Amina Shabbeer, Richard Maclin, and Jude W. Shavlik. Online knowledge-based support vector machines. In *ECML/PKDD (2)*, pages 145–161, 2010. [3](#)
- [11] Fabien Lauer and Gérard Bloch. Incorporating prior knowledge in support vector machines for classification: A review. *Neurocomputing*, 71(7-9):1578–1594, 2008. [1](#), [2](#), [8](#)
- [12] Fabien Lauer and Gérard Bloch. Incorporating prior knowledge in support vector regression. *Mach. Learn.*, 70:89–118, January 2008. [1](#), [8](#)
- [13] Yizeng Liang, Qing-Song Xu, Hong-Dong Li, and Dong-Sheng Cao. *Support Vector Machines and Their Application in Chemistry and Biotechnology*. Taylor and Francis Group, LLC, 2011. [1](#)
- [14] Chun-Fu Lin and Sheng-De Wang. Fuzzy support vector machines. *IEEE Transaction on Neural Networks*, 13(2):464–471, 2002. [8](#)

- [15] Fuming Lin and Jun Guo. A novel support vector machine algorithm for solving nonlinear regression problems based on symmetrical points. In *Proceedings of the 2010 2nd International Conference on Computer Engineering and Technology (ICCET)*, pages 176–180, 2010. [1](#)
- [16] Olvi L. Mangasarian and Edward W. Wild. Nonlinear knowledge-based classification. *IEEE Transactions on Neural Networks*, 19(10):1826–1832, 2008. [2](#), [3](#)
- [17] Tom M. Mitchell. *Machine Learning*. McGraw-Hill, New York, 1997. [3](#)
- [18] Marcin Orchel. Incorporating detractors into svm classification. In Krzysztof Cyran, Stanislaw Kozielski, James Peters, Urszula Stańczyk, and Alicja Wakulicz-Deja, editors, *Man-Machine Interactions*, volume 59 of *Advances in Intelligent and Soft Computing*, pages 361–369. Springer Berlin / Heidelberg, 2009. [3](#), [12](#)
- [19] Marcin Orchel. Paper id 55. In *Submitted to European Conference of Machine Learning, ECML 2010*, April 2010. [1](#)
- [20] Marcin Orchel. Incorporating a priori knowledge from detractor points into support vector classification. In Andrej Dobnikar, Uroš Lotric, and Branko Šter, editors, *Adaptive and Natural Computing Algorithms*, volume 6594 of *Lecture Notes in Computer Science*, pages 332–341. Springer Berlin / Heidelberg, 2011. [3](#), [12](#)
- [21] Marcin Orchel. Regression based on support vector classification. In Andrej Dobnikar, Uroš Lotric, and Branko Šter, editors, *Adaptive and Natural Computing Algorithms*, volume 6594 of *Lecture Notes in Computer Science*, pages 353–362. Springer Berlin / Heidelberg, 2011. [1](#)
- [22] Marcin Orchel. Support vector regression as a classification problem with a priori knowledge in the form of detractors. In Tadeusz Czachorski, Stanislaw Kozielski, and Urszula Stańczyk, editors, *Man-Machine Interactions 2*, volume 103 of *Advances in Intelligent and Soft Computing*, pages 353–362. Springer Berlin / Heidelberg, 2011. [3](#), [12](#)
- [23] Marcin Orchel. Support vector regression with a priori knowledge used in order execution strategies based on vwap. In Jie Tang, Irwin King, Ling Chen, and Jianyong Wang, editors, *Advanced Data Mining and Applications*, volume 7121 of *Lecture Notes in Computer Science*, pages 318–331. Springer Berlin / Heidelberg, 2011. [4](#)
- [24] John C. Platt. *Fast training of support vector machines using sequential minimal optimization*, pages 185–208. MIT Press, Cambridge, MA, USA, 1999. [14](#)
- [25] Jean-Baptiste Pothin and Cedric Richard. Incorporating prior information into support vector machines in the form of ellipsoidal knowledge sets, 2006. [3](#)
- [26] Manel Martinez Ramon and Christos Christodoulou. *Support Vector Machines for Antenna Array Processing and Electromagnetics*. Margan & Claypool Publishers, 2006. [1](#)
- [27] Catarina Silva and Bernardete Ribeiro. *Inductive Inference for Large Scale Text Classification*. Springer, 2010. [1](#)
- [28] Francis Eng Hock Tay and Lijuan Cao. Modified support vector machines in financial time series forecasting. *Neurocomputing*, 48(1-4):847–861, 2002. [8](#)
- [29] Vladimir N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995. [1](#)
- [30] Vladimir N. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, September 1998. [1](#), [8](#)

- [31] Lei Wang, Ping Xue, and Kap Luk Chan. Incorporating prior knowledge into svm for image retrieval. In *ICPR '04: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 2*, pages 981–984, Washington, DC, USA, 2004. IEEE Computer Society. 1, 8
- [32] Meng Wang, Jie Yang, Guo-Ping Liu, Zhi-Jie Xu, and Kuo-Chen Chou. Weighted-support vector machines for predicting membrane protein types based on pseudo amino acid composition. *Protein engineering, design & selection*, 17(6):509–516, 2004. 8
- [33] Edward W. Wild. *Optimization-based Machine Learning and Data Mining*. PhD thesis, University of Wisconsin-Madison, 2008. 3
- [34] Xiaoyun Wu and Rohini Srihari. Incorporating prior knowledge with weighted margin support vector machines. In *KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 326–333, New York, NY, USA, 2004. ACM. 8, 12
- [35] L. Yu, S. Wang, K.K. Lai, and L. Zhou. *Bio-inspired credit risk analysis: computational intelligence with support vector machines*. Springer Verlag, 2008. 1