PRACE NAUKOWE AKADEMII EKONOMICZNEJ WE WROCŁAWIU
Nr 1076 2005
TAKSONOMIA 12
Klasyfikacja i analiza danych – teoria i zastosowania

Michał Trzęsiok
Akademia Ekonomiczna w Katowicach

METODA WEKTORÓW NOŚNYCH W KONSTRUKCJI NIEPARAMETRYCZNYCH MODELI REGRESJI

1. Wstęp

Metoda wektorów nośnych (ang. *SVM – support vector machines*) została pierwotnie skonstruowana jako narzędzie służące do dyskryminacji. Jest naturalny sposób jej przeformułowania, prowadzący do zbudowania modelu regresji nieliniowej.

W zadaniu dyskryminacji metoda wektorów nośnych polega na przetransformowaniu obserwacji ze zbioru uczącego, za pomocą nieliniowego przekształcenia φ, w przestrzeń o dużo większym wymiarze, gdzie klasy rozdzielane są hiperpłaszczyzną optymalną (zob. [Trzęsiok 2003]), daną wzorem:

$$\boldsymbol{\beta} \cdot \varphi(\mathbf{x}) + \beta_0 = 0. \tag{1}$$

W zagadnieniu regresji celem jest znalezienie hiperpłaszczyzny leżącej jak najbliżej jak największej liczby punktów ze zbioru uczącego, po wcześniejszym przetransformowaniu tych punktów za pomocą nieliniowego przekształcenia w przestrzeń o większym wymiarze. Tak postawiony problem przekłada się na zadanie minimalizacji normy wektora kierunkowego hiperpłaszczyzny $\boldsymbol{\beta}$, przy jednoczesnej minimalizacji odległości punktów ze zbioru uczącego od tejże hiperpłaszczyzny. V. Vapnik – twórca metody wektorów nośnych – zaproponował Vapnik [1998] do pomiaru dopasowania funkcję straty, mający postać:

$$L^{\varepsilon}(y, f(\mathbf{x}, \boldsymbol{\beta})) = \begin{cases} 0, & \text{gdy} \quad |y - f(\mathbf{x}, \boldsymbol{\beta})| \le \varepsilon, \\ |y - f(\mathbf{x}, \boldsymbol{\beta})| - \varepsilon, & \text{gdy} \quad |y - f(\mathbf{x}, \boldsymbol{\beta})| > \varepsilon, \end{cases} \tag{2}$$

gdzie $f(\mathbf{x}, \boldsymbol{\beta}) = \boldsymbol{\beta} . \varphi(\mathbf{x}) + \beta_0$ jest funkcją regresji wyznaczoną przez hiperpłaszczyznę (1). Funkcja straty (2) zwraca wartość zero dla wszystkich punktów leżących bliżej hiperpłaszczyzny niż z góry zadana wartość $\varepsilon > 0$. Tak wyznaczonej hiper-

płaszczyźnie w nowej przestrzeni cech odpowiada nieliniowa funkcja regresji w przestrzeni pierwotnej.

Przekształcenie nieliniowe przestrzeni, które umożliwia konstruowanie za pomocą metody wektorów nośnych nieliniowych modeli regresji, realizuje się przez pewną funkcję jądrową definiującą iloczyn skalarny w nowej przestrzeni cech. Wybór postaci funkcji jądrowej, a także wybór wartości parametrów tej funkcji, wartości $\varepsilon > 0$, ma kluczowe znaczenie dla jakości otrzymywanego modelu, lecz stanowi najmniej rozpoznaną część teoretyczną metody wektorów nośnych. Domyślnie w dostępnym oprogramowaniu, jak w wielu publikacjach naukowych, autorzy wykorzystują funkcję jądrową Gaussa i przeszukują jedynie pewien zakres jej parametrów w celu uzyskania najlepszego modelu.

W następnych częściach pracy przedstawione będą algorytm metody wektorów nośnych w zagadnieniu regresji oraz próba odpowiedzi na pytanie o zasadność wyboru funkcji jądrowej Gaussa. Przeprowadzona zostanie również analiza błędu średniokwadratowego, który generuje metoda w zależności od postaci funkcji jądrowej oraz jej parametrów.

2. Metoda wektorów nośnych w regresji

2.1. Ryzyko rzeczywiste, empiryczne i funkcje straty

Zakładamy, że dany jest zbiór uczący o postaci $\left\{\left(\mathbf{x}^1, y^1\right),\ldots,\left(\mathbf{x}^N, y^N\right)\right\}$, gdzie $\mathbf{x}^i \in \mathbf{R}^d$ oraz $y^i \in \mathbf{R}$ dla $i = 1,\ldots,N$, który tworzą realizacje niezależnych zmiennych losowych o jednakowym, nieznanym rozkładzie:

$$p(\mathbf{x}, y) = p(\mathbf{x})p(y \mid \mathbf{x}). \tag{3}$$

Należy wyznaczyć funkcję regresji $f(\mathbf{x}, \boldsymbol{\beta})$ minimalizującą ryzyko rzeczywiste, tj. funkcjonał dany wzorem:

$$R(\boldsymbol{\beta}) = \int L(y, f(\mathbf{x}, \boldsymbol{\beta}))p(\mathbf{x}, y)d\mathbf{x}dy, \tag{4}$$

gdzie: $L(y, f(\mathbf{x}, \boldsymbol{\beta}))$ jest funkcją straty mierzącą jakość aproksymacji, a $\boldsymbol{\beta}$ – wektorem parametrów funkcji regresji. Wartości tego funkcjonału są nieznane, jako że nieznany jest rozkład $p(\mathbf{x}, y)$. Możemy zatem jedynie oszacować wartość $R(\boldsymbol{\alpha})$ na podstawie zbioru uczącego. Wobec tego w przeszukiwanej przestrzeni hipotez wybierana będzie ta funkcja f, która minimalizuje ryzyko empiryczne:

$$R_{emp}(\boldsymbol{\beta}) = \frac{1}{N}\sum_{i=1}^{N} L(y^i, f(\mathbf{x}^i, \boldsymbol{\beta})). \tag{5}$$

Jednakże mała wartość ryzyka empirycznego (dobre dopasowanie na zbiorze uczącym) nie gwarantuje niewielkiej wartości ryzyka rzeczywistego. Oznacza to, że otrzymany model może mieć słabą zdolność objaśniania (uogólniania). Metoda

wektorów nośnych została skonstruowana przez Vapnika [1998] na podstawie teoretycznego oszacowania ryzyka rzeczywistego. Vapnik udowodnił, że w liniowej funkcji regresji (definiującej hiperpłaszczyznę w przestrzeni cech) ryzyko rzeczywiste jest ograniczone z góry przez wartość ryzyka empirycznego oraz długość (normę) wektora kierunkowego hiperpłaszczyzny. Mała wartość ryzyka empirycznego jest związana z dobrym dopasowaniem modelu, a minimalizacja normy wektora kierunkowego hiperpłaszczyzny – z lepszą zdolnością modelu do uogólniania. Procedura jednoczesnej minimalizacji obu składników ograniczających ryzyko rzeczywiste nazywana jest minimalizacją ryzyka strukturalnego.

Istotnym zagadnieniem jest również wybór funkcji straty. Kwadratowa funkcja straty, mająca postać $L(y, f(\mathbf{x}, \boldsymbol{\beta})) = (y - f(\mathbf{x}, \boldsymbol{\beta}))^2$, wraz z minimalizacją ryzyka empirycznego stanowi podstawę klasycznej metody najmniejszych kwadratów. W metodzie wektorów nośnych jest wykorzystywana funkcja straty Vapnika o postaci (2).

2.2. Formalny zapis metody wektorów nośnych w regresji

Ponieważ aproksymowana zależność między zmienną objaśnianą y a zmienną objaśniającą $\mathbf{x}$ może mieć charakter nieliniowy, metoda wektorów nośnych wykorzystuje nieliniową transformację $\varphi: \mathbf{R}^d \rightarrow \mathbf{Z}$ danych ze zbioru uczącego do przestrzeni o większym wymiarze. Następnie w nowej przestrzeni cech jest wyznaczana hiperpłaszczyzna (odpowiadająca liniowej funkcji regresji) minimalizująca ryzyko strukturalne. Rozpatrujemy zatem zbiór $\left\{\left(\varphi(\mathbf{x}^1), y^1\right), \ldots, \left(\varphi(\mathbf{x}^N), y^N\right)\right\}$, gdzie $\varphi(\mathbf{x}^i) \in \mathbf{Z}$ oraz $y^i \in R$ dla $i = 1, \ldots, N$. Zakładamy, że wyznaczana funkcja regresji f ma postać:

$$y = f(\mathbf{x}, \boldsymbol{\beta}) = \boldsymbol{\beta} \cdot \varphi(\mathbf{x}) + \beta_0, \tag{6}$$

czyli jest funkcją liniową w przestrzeni $\mathbf{Z}$ i definiuje w tej przestrzeni hiperpłaszczyznę (jednocześnie funkcja ta, rozważana w pierwotnej przestrzeni danych $\mathbf{R}^d$, jest funkcją nieliniową). Zgodnie z opisaną powyżej zasadą minimalizacji ryzyka strukturalnego, funkcję f znajdujemy jak rozwiązanie zadania minimalizacji następującego wyrażenia:

$$\tfrac{1}{2}\|\boldsymbol{\beta}\|^2 + C\sum_{i=1}^{N} L^{\varepsilon}(y^i, f(\mathbf{x}^i, \boldsymbol{\beta})), \tag{7}$$

gdzie C jest parametrem metody, określającym kompromis między dokładnością dopasowania a zdolnością uogólniania modelu. Wprowadzając zmienne $\xi_1, \ldots, \xi_N, \xi_1^*, \ldots, \xi_N^* \geq 0$ osłabiające założenie, że wszystkie obserwacje muszą znajdować się bliżej wyznaczanej hiperpłaszczyzny niż z góry zadana wartość ε, oraz wykorzystując funkcję straty (2), możemy zadanie minimalizacji wyrażenia (7) zapisać w postaci zadania programowania kwadratowego z liniowymi ograniczeniami:

$$\begin{cases} \min\limits_{\beta,\beta_0,\xi} \frac{1}{2}\|\beta\|^2 + C\sum\limits_{i=1}^{N}(\xi_i + \xi_i^*), \\ y^i - \left(\beta \cdot \varphi(\mathbf{x}^i) + \beta_0\right) \le \varepsilon + \xi_i, \quad i = 1,\ldots,N, \\ -y^i + \left(\beta \cdot \varphi(\mathbf{x}^i) + \beta_0\right) \le \varepsilon + \xi_i^*, \quad i = 1,\ldots,N, \\ \xi_i, \xi_i^* \ge 0. \end{cases} \tag{8}$$

Zadanie to rozwiązujemy metodą mnożników Lagrange'a. Funkcja Lagrange'a ma postać:

$$\begin{aligned} L_p(\beta,\beta_0,\alpha,\alpha^*) = \frac{1}{2}\|\beta\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) - \sum_{i=1}^{N}\alpha_i\left[\varepsilon + \xi_i - y^i + \beta \cdot \varphi(\mathbf{x}^i) + \beta_0\right] \\ - \sum_{i=1}^{N}\alpha_i^*\left[\varepsilon + \xi_i^* + y^i - \beta \cdot \varphi(\mathbf{x}^i) - \beta_0\right] - \sum_{i=1}^{N}\left(\eta_i\xi_i + \eta_i^*\xi_i^*\right), \end{aligned} \tag{9}$$

gdzie $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \ge 0$ są współczynnikami Lagrange'a. Przyrównując pochodne cząstkowe rzędu pierwszego funkcji Lagrange'a do zera i podstawiając otrzymane zależności do (9), otrzymujemy tzw. dualną postać funkcji Lagrange'a:

$$L_D(\alpha,\alpha^*) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right)\varphi(\mathbf{x}^i) \cdot \varphi(\mathbf{x}^j) - \varepsilon\sum_{i=1}^{N}\left(\alpha_i + \alpha_i^*\right) + \sum_{i=1}^{N}\left(\alpha_i - \alpha_i^*\right)y^i, \tag{10}$$

przy ograniczeniach $0 \le \alpha_i, \alpha_i^* \le C, i = 1,\ldots,N, \sum\limits_{i=1}^{N}\left(\alpha_i - \alpha_i^*\right) = 0.$

Powyższe zadanie ma jednoznacznie wyznaczone rozwiązanie:

$$\begin{aligned} \beta &= \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\varphi(\mathbf{x}^i), \\ \hat{\beta}_0 &= y^i - \beta \cdot \varphi(\mathbf{x}^i) - \varepsilon, \quad \text{dla } 0 < \alpha_i < C. \end{aligned} \tag{11}$$

Zatem funkcja regresji ma postać:

$$f(\mathbf{x}) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\varphi(\mathbf{x}^i) \cdot \varphi(\mathbf{x}) + \hat{\beta}_0. \tag{12}$$

Rozwiązanie optymalne (11) ponadto spełnia warunki Kuhna-Tuckera. Jeden z nich ma postać:

$$\alpha_i\alpha_i^* = 0, \quad i = 1,\ldots,N. \tag{13}$$

2.3. Wektory nośne

Wektorami nośnymi (ang. *support vectors*) nazywamy te obserwacje $\varphi(\mathbf{x}^i)$, dla których odpowiadające im współczynniki Lagrange'a α_i lub α_i^* są niezerowe (większe od zera). Widać zatem, że wyznaczona funkcja regresji (12) zależy nie od wszystkich obserwacji, a tylko od tych, będących wektorami nośnymi (pozostałe

obserwacje występują we wzorze (12) ze współczynnikami równymi zeru i są nieistotne dla rozważanego zagadnienia regresji). Oznaczmy przez I_{SV} zbiór tych indeksów ze zbioru $\{1,\ldots,N\}$, którym odpowiadają wektory nośne.

2.4. Funkcje jądrowe

Zarówno w sformułowaniu zadania optymalizacyjnego (10), jak i w jego rozwiązaniu (12) postać nieliniowej funkcji transformującej φ nie musi być znana. Wystarczy, jeśli zdefiniuje się iloczyn skalarny o postaci $K(\mathbf{u},\mathbf{v}) = \varphi(\mathbf{u}).\varphi(\mathbf{v})$ w nowej przestrzeni cech **Z**. Metoda wektorów nośnych wykorzystuje najczęściej funkcje z rodziny funkcji jądrowych:

1) Gaussa: $K(\mathbf{u},\mathbf{v}) = \exp(-\gamma\|\mathbf{u}-\mathbf{v}\|^2)$,
2) wielomianową: $K(\mathbf{u},\mathbf{v}) = \gamma(\mathbf{u}\cdot\mathbf{v}+\delta)^d$, $d = 1,2,\ldots$,
3) sigmoidalną: $K(\mathbf{u},\mathbf{v}) = \tanh(\gamma\mathbf{u}\cdot\mathbf{v}+\delta)$,
4) liniową: $K(\mathbf{u},\mathbf{v}) = \mathbf{u}\cdot\mathbf{v}$.

Przy zadanej funkcji jądrowej, definiującej iloczyn skalarny w **Z**, funkcję regresji (12) możemy zapisać następująco:

$$f(\mathbf{x}) = \sum_{i\in I_{SV}} (\alpha_i - \alpha_i^*)K(\mathbf{x}^i,\mathbf{x}) + \hat{\beta}_0. \tag{14}$$

Wektory nośne, definiujące funkcję regresji, leżą na brzegu epsylonowego otoczenia wyznaczonej hiperpłaszczyzny lub na zewnątrz tego otoczenia.

3. Analiza własności metody

Jak już wspomniano, jakość otrzymywanego modelu regresji zależy w sposób istotny od wyboru postaci funkcji jądrowej, wartości parametrów tej funkcji oraz od wartości dwóch dodatkowych parametrów metody wektorów nośnych: ε i C. Najczęściej do ustalenia optymalnych wartości parametrów jest proponowana metoda symulacyjnego przeszukiwania zadanego przez użytkownika zakresu wartości parametrów lub metoda sprawdzania krzyżowego. W opracowaniu do porównania jakości modeli regresji zbudowanych dla różnych funkcji jądrowych i różnych kombinacji wartości parametrów wykorzystano błąd średniokwadratowy, obliczony metodą sprawdzania krzyżowego z podziałem zbioru uczącego na dziesięć części.

3.1. Zbiory danych

Badanie przeprowadzono na pięciu zbiorach danych, standardowo wykorzystywanych do porównywania własności metod wielowymiarowej analizy statystycznej. Pierwsze dwa zbiory zawierają dane rzeczywiste i mają swoje źródło w ogól-

nodostępnej bazie „UCI Repository Of Machine Learning Databases", zlokalizowanej w Uniwersytecie Kalifornijskim[1], a wykorzystane do obliczeń zbiory zaczerpnięto z pakietu Leisch i Dimitriadou [2004], którego autorzy przygotowali kilkanaście zbiorów danych do bezpośredniego przetwarzania w pakiecie statystycznym *R*. Kolejne trzy zbiory danych to zbiory sztuczne, wygenerowane komputerowo za pomocą funkcji dostępnych w pakiecie `mlbench`, zgodnie z koncepcją Friedmana, opisaną w dokumentacji pakietu `mlbench`, oraz w artykule Friedmana [1991]. Nazwy oraz krótką charakterystykę zbiorów przedstawiono w tab. 1.

Tabela 1. Zbiory danych wykorzystane do analizy.

Nazwa	Liczebność	Liczba zmiennych
BostonHousing	506	14
Servo	167	5
Friedman1	200	11
Friedman2	200	5
Friedman3	200	5

Źródło: opracowanie własne.

Na przykład zbiór „Friedman1" został utworzony jako realizacje dziesięciu niezależnych zmiennych objaśniających, o rozkładzie jednostajnym na przedziale [0, 1], przy czym zmienna objaśniana y jest zależna tylko od pięciu i wyznaczona jest według formuły:

$$y = 10\sin(\pi x_1 x_2) + 20(x_3 - 0{,}5)^2 + 10x_4 + 5x_5 + e,$$

gdzie e ma rozkład $N(0, 1)$.

Do obliczeń wykorzystano program statystyczny *R* z dodatkowymi pakietami: `e1071` (zawierającym implementację metody wektorów nośnych) oraz `mlbench` (zawierającym analizowane zbiory danych, wraz z ich opisem, oraz funkcje, za pomocą których wygenerowano zbiory sztuczne).

3.2. Procedura badawcza

Dla każdego zbioru zbudowano modele regresji metodą wektorów nośnych, używając:

a) funkcji jądrowej Gaussa z parametrami (γ, C, ε), gdzie γ przebiega zakres od 2^{-2} do 2^3, C – od 10^{-2} do 10^3, $\varepsilon \in \{0{,}1; 0{,}5; 0{,}8\}$,

b) wielomianowej funkcji jądrowej z parametrami $(d, \gamma, \delta, C, \varepsilon)$, gdzie d przebiega zakres[2] od 2 do 5, γ – od 2^{-2} do 2^0, $\delta \in \{2{,}10\}$, oraz C – od 10^{-2} do 10^0, $\varepsilon \in \{0{,}1; 0{,}5; 0{,}8\}$,

[1] Dostępne przez: ftp://ftp.ics.uci.edu/pub/machine-learning-databases, http://www.ics.uci.edu/~mlearn/MLRepository.html.

[2] W trzech zbiorach dla wielomianowej funkcji jądrowej analizę przeprowadzono, zawężając zakres niektórych parametrów ze względu na bardzo długi czas oczekiwania na wyniki (program przerwano po 12 godz. obliczeń).

c) sigmoidalnej funkcji jądrowej z parametrami $(\gamma, \delta, C, \varepsilon)$, gdzie γ przebiega zakres od 2^{-2} do 2^{3}, $\delta \in \{2,10\}$, zaś C – od 10^{-2} do 10^{3}, $\varepsilon \in \{0,1; 0,5; 0,8\}$,

d) liniowej funkcji jądrowej z parametrami (C, ε), gdzie C przebiega zakres od 10^{-2} do 10^{3}, $\varepsilon \in \{0,1; 0,5; 0,8\}$.

Dla każdego tak otrzymanego modelu obliczono błąd średniokwadratowy metodą sprawdzania krzyżowego z podziałem zbioru uczącego na dziesięć części.

3.3. Wyniki analizy

Przeprowadzono analizę błędu średniokwadratowego generowanego przez model najpierw ze względu na wybór postaci funkcji jądrowej, a następnie ze względu na wartości poszczególnych parametrów.

Tabela 2 przedstawia minimalne wartości błędu śreniokwadratowego, w zależności od postaci funkcji jądrowej. Pogrubioną czcionką zaznaczono najlepsze rezultaty, a kursywą – najgorsze.

Tabela 2. Minimalny błąd średniokwadratowy w zależności od postaci funkcji jądrowej

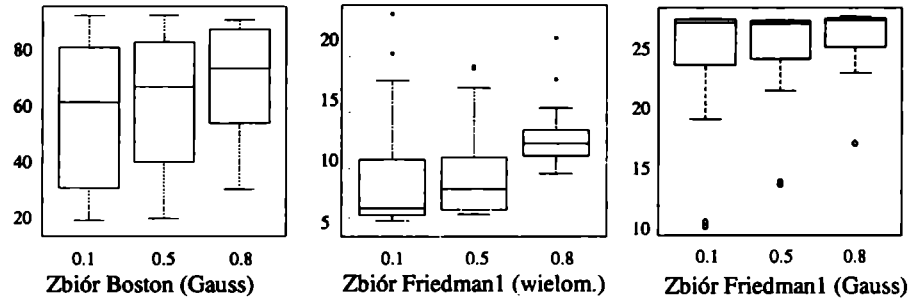
Nazwa	Gaussa	Wielomianowa	Sigmoidalna	Liniowa
BostonHousing	14,55	**11,15**	*54,4*	25,6
Servo	29,65	**23,88**	*91,94*	30,67
Friedman1	9,52	**4,7**	*15,5*	8,98
Friedman2	24661,0	**20670,0**	*95248,0*	38565,0
Friedman3	0,029	**0,024**	*0,105*	0,064

Źródło: opracowanie własne.

Widać, że we wszystkich pięciu przypadkach najmniejsze błędy generuje metoda wektorów nośnych z wielomianową funkcją jądrową. Ponadto zdecydowanie najgorsze rezultaty otrzymujemy po wykorzystaniu sigmoidalnych funkcji jądrowych. Funkcje Gaussa, ustawione jako domyślny wybór w programach komputerowych, dają rezultaty dobre, ale modele regresji na nich oparte ustępowały jakością modelom z wielomianowymi funkcjami jądrowymi.

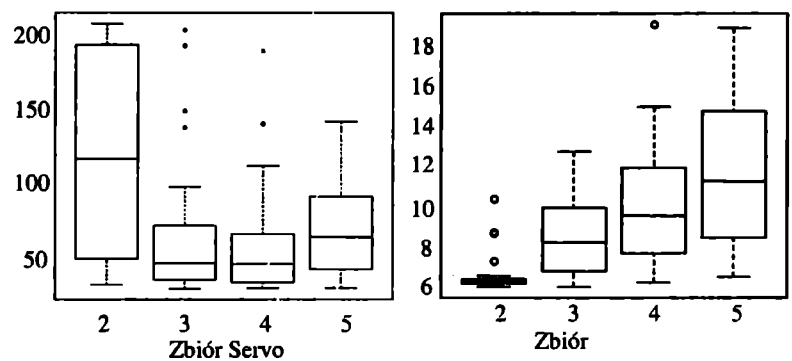
Przekrojowe porównania błędów średniokwadratowych dla różnych wartości badanych parametrów pozwalają sformułować pewne wnioski jedynie w przypadku parametru ε oraz parametru d – stopnia wielomianu.

Ponieważ, oprócz minimalnych błędów, interesująca wydaje się również informacja o innych charakterystykach opisowych (szczególnie informacja o zróżnicowaniu wartości błędów), dla ustalonych wartości jednego z parametrów wyniki analizy przedstawiono na wykresach pudełkowych (rys. 1-2).



Rys. 1. Wykresy pudełkowe błędów średniokwadratowych dla wartości $\varepsilon = 0{,}1$; 0,5; 0,8, dla trzech wybranych zbiorów: zbioru Boston z funkcją jądrową Gaussa, zbioru Friedman1 z wielomianową funkcją jądrową oraz funkcją Gaussa
Źródło: opracowanie własne.

Na przedstawionych wykresach (rys.1) widać, że błędy średniokwadratowe charakteryzują się stosunkowo silnym zróżnicowaniem. Potwierdza to również wcześniejsze stwierdzenie, że wybór wartości parametrów jest istotny dla jakości otrzymywanego modelu. Ponadto można wnioskować, że małe wartości parametru ε dają lepsze rezultaty.



Rys. 2. Wykresy pudełkowe błędów średniokwadratowych dla wielomianowych funkcji jądrowych dla wielomianów stopnia 2, 3, 4, 5 dla zbioru Servo i zbioru Friedman1
Źródło: opracowanie własne.

Bardzo istotnym parametrem w wielomianowej funkcji jądrowej jest stopień wielomianu d. Dla dwóch zbiorów dodatkowo obliczono błędy średniokwadratowe dla wielomianów stopnia szóstego i siódmego. Na podstawie przeprowadzonych badań można wnioskować, że najlepsze rezultaty dają wielomiany stopnia 3 oraz 4. Zwiększanie stopnia wielomianu powoduje na ogół znaczne pogorszenie jakości otrzymywanego modelu. Wyniki analizy na przykładowych dwóch zbiorach przedstawia rys. 2.

Pozostałe parametry mają również istotny wpływ na jakość dyskryminacji, ale nie można wskazać takich ich wartości, które dawałyby małe błędy średniokwadratowe, niezależnie od analizowanego zbioru danych.

4. Podsumowanie

Metoda wektorów nośnych umożliwia budowę nieliniowych modeli regresji z zachowaniem dużego uogólnienia otrzymywanego modelu. Nieliniowość metody realizuje się przez wybór funkcji jądrowej. Różnorodność możliwości wyboru rodzaju funkcji jądrowej oraz jej parametrów sprawiają, że metoda pozwala przeszukiwać duży zbiór hipotez (potencjalnych funkcji regresji opisujących badaną zależność). Najlepsze rezultaty dają wielomianowe funkcje jądrowe dla wielomianów stopnia 3 lub 4 oraz małe wartości parametru ε (np. 0,1). Wskazanie wyboru wielomianowej funkcji jądrowej stopnia 3 lub 4 jest zgodne z wynikami analiz, przedstawionymi w pracy Trzęsioka [2004], które dotyczyły analogicznego badania, lecz przypadku zastosowania metody wektorów nośnych w dyskryminacji.

Słabością wielomianowych funkcji jądrowych jest, po pierwsze, to, że użytkownik musi określić wartości czterech parametrów, które nie mają jasnej interpretacji (w przeciwieństwie do dwóch interpretowalnych parametrów dla funkcji jądrowej Gaussa). Po drugie, symulacyjne przeszukiwanie zakresu parametrów dla określenia układu dającego najmniejszy błąd klasyfikacji może być, bardzo czasochłonne szczególnie przy dużych zbiorach danych.

Literatura

Cristianini N., Shawe-Taylor J., *An Introduction To Support Vector Machines (and other Kernel-Based Learning Methods),* Cambridge University Press, Cambridge 2000.

Friedman J., *Multivariate Adaptive Regression Splines*, „The Annals of Statistics" 1991, 19 (1).

Gunn S.R., *Support Vector Machines for Classification and Regression,* Technical Report, Image Speech and Intelligent Systems Research Group, University of Southampton, 1997.

Hastie T., Tibshirani R., Friedman J., *The Elements of Statistical Learning*, Springer Verlag, New York 2001.

Leisch F., Dimitriadou E., *The* `mlbench` *Package – a Collection for Artificial and Real-World Machine Learning Benchmarking Problems, R* package, Version 1.0-0, 2004, http://cran.R-project.org.

Trzęsiok M., *Analiza wybranych własności metody dyskryminacji wykorzystującej wektory nośne*, [w:] *Postępy ekonometrii*, red. A.S. Barczak, AE, Katowice 2004.

Trzęsiok M., *Zastosowanie metody SVM w klasyfikacji danych,* Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1022, AE, Wrocław 2003.

Vapnik V., *Statistical Learning Theory*, John Wiley & Sons, New York 1998.

SUPPORT VECTOR MACHINES FOR REGRESSION

Summary

For nonlinear regression problem, support vector machines (SVM) map the input space into a high-dimensional feature space first, and then perform linear regression in the high-dimensional feature space. The nonlinearity of SVM is realized by choosing the kernel function. Performance of SVM is very sensitive to the choice of the kernel and model parameters. In the paper the method is presented and the dependency of its performance on the kernel and the model parameters selection is analyzed.