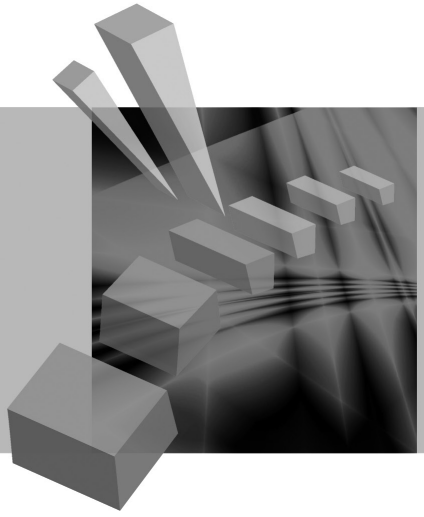
PRACE NAUKOWE
Uniwersytetu Ekonomicznego we Wrocławiu
RESEARCH PAPERS
of Wrocław University of Economics

278

Taksonomia 20

Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi
Krzysztof Jajuga
Marek Walesiak

Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2013

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:
www.ibuk.pl, www.ebscohost.com,
The Central and Eastern European Online Library www.ceeol.com,
a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon
http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php

Informacje o naborze artykułów i zasadach recenzowania znajdują się
na stronie internetowej Wydawnictwa
www.wydawnictwo.ue.wroc.pl

Tytuł dofinansowany ze środków Narodowego Banku Polskiego
oraz ze środków Sekcji Klasyfikacji i Analizy danych PTS

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

Spis treści

Summaries

PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS nr 278 • 2013

Taksonomia 20. Klasyfikacja i analiza danych – teoria i zastosowania ISSN 1899-3192

Joanna Trzęsiok
Uniwersytet Ekonomiczny w Katowicach

WYBRANE SYMULACYJNE TECHNIKI PORÓWNYWANIA NIEPARAMETRYCZNYCH METOD REGRESJI

Streszczenie: W artykule przedstawiono symulacyjną procedurę badawczą pozwalającą na porównywanie różnych nieparametrycznych modeli regresji. Procedura ta przebiega dwuetapowo. Na początku tworzonych jest wiele modeli regresji, spośród których wybrane i uszeregowane w postaci rankingu zostają te modele, które charakteryzują się najlepszą dokładnością predykcji, mierzoną za pomocą estymatora punkowego, jakim jest błąd średniokwadratowy obliczony metodą sprawdzania krzyżowego (MSE_{CV}). Drugi etap analizy ma na celu zbadanie istotności różnic pomiędzy uzyskanymi wartościami MSE_{CV}, a tym samym skorygowanie otrzymanych rankingów metod. Zaproponowaną procedurę badawczą zastosowano w badaniu empirycznym dla zbiorów danych standardowo wykorzystywanych do badania własności metod regresji.

Słowa kluczowe: regresja nieparametryczna, porównywanie modeli, symulacyjne procedury badawcze, testowanie hipotez.

1. Wstęp

Rozwój technologii informatycznych pozwolił na budowę nieparametrycznych, wielowymiarowych metod regresji, wykorzystujących do budowy modeli złożone algorytmy numeryczne. Metody te pozwalają na analizę zbiorów danych o dużej liczebności, opisywanych przez wiele zmiennych. Ze względu na sposób ich działania, polegający często na systematycznym przeszukiwaniu (eksplorowaniu) zbioru danych, zalicza się je do grona metod *Data Mining*. Stanowią one liczną grupę zróżnicowanych i dynamicznie rozwijających się metod. Tym samym pojawił się problem zarówno porównywania tych metod, jak i wyboru jednej z nich do rozwiązywania postawionego zadania regresji.

Wybór najlepszej metody do rozwiązania zadanego problemu jest dylematem, przed którym postawiony zostaje niejeden badacz. Analizy mające na celu porównywanie i testowanie różnych metod regresji pokazują, że niemożliwe jest wskazanie metody najlepszej, za pomocą której budowane są modele dające najmniejsze błędy średniokwadratowe, niezależnie od rozważanego zbioru danych (por. [Meyer, Le-

isch, Hornik 2003]). Charakter badanego zbioru danych czasem determinuje wybór odpowiedniej metody. Najczęściej jednak mamy do dyspozycji wiele modeli. Ponadto stosowane w praktyce coraz lepsze metody statystyczne są adekwatne do poziomu złożoności badanych zjawisk i niejednokrotnie pozwalają zbudować modele, które charakteryzują się równie wysoką dokładnością predykcji.

Celem artykułu było przedstawienie procedury badawczej pozwalającej na porównywanie metod nieparametrycznych, jak i wybór najlepszej z nich do rozwiązania postawionego zadania regresji. Procedura ta prowadzi do stworzenia rankingu nieparametrycznych modeli regresji pod względem generowanych błędów średniokwadratowych, uwzględniając istotność różnic pomiędzy otrzymanymi wartościami błędu *MSE*. Ze względu na charakter nieparametrycznych metod regresji – ich odmienne mechanizmy działania, niemożliwe jest analityczne porównanie otrzymywanych modeli. Badania porównawcze przeprowadzone zostały więc za pomocą procedur symulacyjnych, na zbiorach danych standardowo wykorzystywanych do badania własności różnych metod regresji.

2. Opis procedury badawczej

W zaproponowanej procedurze badawczej wybór najlepszego rozwiązania dla postawionego zadania regresji przebiega dwuetapowo.

W pierwszym etapie zbudowanych zostaje wiele modeli za pomocą różnych, zarówno nieparametrycznych, jak i klasycznych, metod regresji. Tworzone są one dla różnych zestawów parametrów, dla każdej z metod. Jednak w ostatecznym zestawieniu daną metodę reprezentuje zawsze tylko jeden model – ten, w którym wykorzystano optymalną kombinacją parametrów. Zwieńczeniem tego etapu procedury badawczej jest stworzenie rankingu modeli pod względem dokładności predykcji ocenianej za pomocą estymatora punktowego, jakim jest błąd średniokwadratowy obliczony metodą sprawdzania krzyżowego[1] (MSE_{CV}). Model będący najlepszym rozwiązaniem danego zadania regresji to ten o najmniejszej wartości błędu MSE_{CV}. Szczegółowo ten etap procedury badawczej przedstawiony został w tab. 1.

W drugim etapie w celu zapewnienia poprawności procedury badawczej należy zbadać istotność różnic pomiędzy otrzymanymi wartościami błędów średniokwadratowych (obliczonymi dla modeli zbudowanych różnymi metodami) (por. [Hothorn i in. 2005]). Jeżeli różnice te są nieistotne, to model najlepszy nie musi być tym o najmniejszej wartości błędu średniokwadratowego. W wyborze optymalnego rozwiązania można się wtedy kierować innymi własnościami modelu, jak choćby stopniem złożoności czy możliwościami interpretacji jego postaci.

[1] Metoda sprawdzania krzyżowego jest uniwersalną metodą estymacji, która polega na podziale zbioru danych na b rozłącznych i równolicznych (w przybliżeniu) części. W każdym z b kroków algorytmu tej metody jedną (ale za każdym razem inną) część z otrzymanego podziału wykorzystuje się do testowania modelu zbudowanego na pozostałych $b - 1$ częściach zbioru danych. Otrzymane wyniki zostają na końcu uśrednione. Statystyka MSE_{CV} jest nieobciążonym estymatorem błędu predykcji [Blum i in. 1999].

Tabela 1. Etap pierwszy procedury badawczej – porównywanie zdolności predykcyjnych modeli za pomocą estymatora punktowego MSE_{CV}

Krok 1.	Przygotowanie zbioru uczącego *D*, czyli podział *D* na 10 równolicznych (w przybliżeniu) oraz rozłącznych części[2]
Krok 2.	Wykonanie następujących czynności dla każdej z rozpatrywanych metod regresji: a) zbudowanie wielu modeli regresji (z wykorzystaniem jednej metody) dla różnych wartości parametrów tej metody; b) obliczanie błędu średniokwadratowego MSE_{CV} metodą sprawdzania krzyżowego dla modeli otrzymanych w punkcie a); c) wybór tego układu parametrów i odpowiadającego mu modelu, dla którego uzyskano najmniejszy błąd MSE_{CV}, czyli wybór modelu – reprezentanta danej metody do porównań
Krok 3.	Stworzenie rankingu analizowanych modeli regresji, pod względem otrzymanych wartości błędów MSE_{CV}

Źródło: opracowanie własne.

W omawianej procedurze badawczej do badania istotności różnic pomiędzy wartościami błędu średniokwadratowego wykorzystano dwa nieparametryczne testy statystyczne:

- test Kruskala-Wallisa, w którym badamy hipotezę zerową o równości wartości MSE_{CV} obliczonych dla wszystkich wyznaczonych modeli regresji M_i (dla $i = 1, \ldots, K$):

$$H_0: \quad MSE_{CV}(M_1) = \ldots = MSE_{CV}(M_K), \tag{1}$$

wobec hipotezy alternatywnej:

$$H_1: \quad \bigvee_{i \neq j} MSE_{CV}(M_i) \neq MSE_{CV}(M_j); \tag{2}$$

- test Manna-Whitneya-Wilcoxona, sprawdzający istotność różnic pomiędzy parami liczb:

$$H_0: \quad MSE_{CV}(M_i) = MSE_{CV}(M_j) \quad \text{dla } i, j = 1, \ldots, K \tag{3}$$

wobec hipotezy alternatywnej:

$$H_1: \quad MSE_{CV}(M_i) \neq MSE_{CV}(M_j) \quad \text{dla } i, j = 1, \ldots, K. \tag{4}$$

Etap drugi procedury badawczej został szczegółowo przedstawiony w tab. 2.

[2] Możliwy jest podział zbioru danych na inną liczbę części, jednak Kohavi w pracy [1995] zaleca stosowanie metody sprawdzania krzyżowego z parametrem $b \leq 10$.

Tabela 2. Etap drugi procedury badawczej – testowanie istotności różnic pomiędzy wartościami błędu MSE_{CV}

Krok 1.	Przygotowanie zbioru uczącego *D*, czyli losowanie z niego *B* prób bootstrapowych: $\mathcal{L}_1, \ldots, \mathcal{L}_B$
Krok 2.	Wykonanie następujących czynności dla każdej próby $\mathcal{L}_b$ (dla $b = 1, \ldots, B$): podział $\mathcal{L}_b$ na 10 równolicznych (w przybliżeniu) oraz rozłącznych części; obliczenie, metodą sprawdzania krzyżowego, błędu średniokwadratowego $MSE_{CV}(M_i \mid \mathcal{L}_b)$ dla każdego z rozpatrywanych modeli regresji M_i (dla $i = 1, \ldots, K$) z optymalnym zestawem wartości parametrów (otrzymanym w pierwszym etapie procedury)
Krok 3.	Dla rozpatrywanych modeli regresji M_i (dla $i = 1, \ldots, K$): testowanie (parami lub wszystkich jednocześnie) na podstawie ciągów wartości $\left\{ MSE_{CV}(M_i \mid \mathcal{L}_b) \right\}_{b=1, \ldots, B}$ istotności różnic pomiędzy wartościami MSE_{CV} (otrzymanymi w etapie pierwszym); uwzględnienie wyników w rankingu metod regresji

Źródło: opracowanie własne.

Należy podkreślić, że w celu zapewnienia poprawności testowania istotności różnic pomiędzy MSE_{CV} konieczne jest zadbanie o jednolitą i przejrzystą procedurę badawczą, dającą jednakowe warunki do obliczeń i porównań. Oznacza to między innymi, że wszystkie rozpatrywane modele regresji budowane są na tych samych próbach bootstrapowych $\mathcal{L}_1, \ldots, \mathcal{L}_B$, wylosowanych z danego zbioru uczącego. Nie zmieniają się również wyznaczone w pierwszym etapie procedury optymalne kombinacje parametrów modeli.

3. Analiza z wykorzystaniem przedstawionej procedury badawczej

Analizę przeprowadzono na pięciu rzeczywistych zbiorach danych[3], standardowo wykorzystywanych do badania własności różnych metod regresji. Najważniejsze charakterystyki tych zbiorów zestawiono w tab. 3.

Tabela 3. Charakterystyki zbiorów danych wykorzystywanych w analizie

Nazwa zbioru	Liczba obserwacji	Liczba zmiennych
Autompg	398	8
Boston	506	14
Clothing	400	13
Ozone	366	13
Star	5748	6

Źródło: opracowanie własne.

[3] Zbiory danych wykorzystane w analizie pochodzą z bibliotek `Ecdat` oraz `mlbench` programu statystycznego **R**.

W badaniu porównywano nieparametryczne modele regresji zbudowane za pomocą:

1) metody rzutowania PPR [Friedman, Stuetzle 1981],
2) metody polegającej na równoległym łączeniu drzew regresyjnych [Breiman 1996] (oznaczonej jako BAGGING),
3) stochastycznej, addytywnej metody drzew regresyjnych MART [Friedman 1999a; Friedman 1999b],
4) metody zagregowanych drzew regresyjnych Breimana – RANDOM FORESTS [Breiman 2001],
5) wielowymiarowej metody krzywych sklejanych POLYMARS [Kooperberg i in. 1997],
6) metody wektorów nośnych SVM [Vapnik 1998],
7) metody wykorzystującej sieci neuronowe (oznaczonej jako NNET) (por. [Bishop 1995]).

Wyniki dla nieparametrycznych modeli regresji zestawiono również z wartościami błędu MSE_{CV}, obliczonego dla

8) klasycznego, liniowego modelu regresji wielorakiej (LM).

Do budowy modeli regresji wykorzystano program statystyczny **R** z dodatkowymi bibliotekami. Większość badanych metod wymaga ustalenia wartości pewnych parametrów budowanego modelu regresji. Przeszukiwane zakresy parametrów dla poszczególnych metod to:

- w metodzie rzutowania PPR wartość parametru opisującego początkową liczbę funkcji składowych modelu przyjmowano na poziomie: 10, 15, 20, 25, zaś końcowa liczba tychże funkcji w modelu zmieniała się od 1 do 10;
- w metodzie zagregowanych drzew regresyjnych Breimana liczbę zmiennych losowanych przy każdym podziale ustalano na poziomie: $\sqrt{m}$, $\frac{m}{3}$, $2\sqrt{m}$ (m – liczba zmiennych), liczbę drzew równą 100 oraz 200, zaś minimalną liczbę obserwacji w liściu: 1, 5, 10;
- w metodzie MART liczbę modeli składowych dobierano metodą sprawdzania krzyżowego, zakładając, że ich maksymalna możliwa liczba równa jest 10 000;
- w metodzie wektorów nośnych SVM wykorzystano wielomianową funkcję jądrową, przyjmując stopień wielomianu równy 2 lub 3, wartość parametru λ od 10^{-2} do 10, epsilon równe $0{,}1$ oraz $0{,}5$;
- w modelach sieci neuronowych z jedną ukrytą warstwą przyjmowano liczbę obserwacji w warstwie ukrytej zmieniającą się od 1 do $\ln(n)$ (gdzie n jest liczbą obserwacji);
- w pozostałych modelach przyjęto domyśle wartości parametrów zaproponowane przez funkcje realizujące daną metodę w programie statystycznym **R**.

Zgodnie z zaproponowaną procedurą badawczą analiza przebiegała dwuetapowo, a jej wyniki zestawiono w tabelach 4-8.

W pierwszej części badania dla każdego zbioru danych wyznaczono rankingi modeli regresji pod względem błędów średniokwadratowych obliczonych metodą sprawdzania krzyżowego (ten etap obrazują trzy pierwsze kolumny każdej z tab. 4-8).

W etapie drugim testowano różnice pomiędzy wartościami MSE_{CV}. W tym celu z każdego zbioru uczącego wylosowano po 100 prób bootstrapowych (B = 100), co oznacza, że w badaniu posłużono się ośmioma (dla każdego zbioru D), obliczonymi dla każdej z metod regresji, stuelementowymi ciągami wartości $\left\{ MSE_{CV}(M_i | \mathcal{L}_b) \right\}_{b=1,\ldots,100}$. Wyniki badania istotności różnic między błędami MSE_{CV} dały pewną korektę uzyskanych wcześniej rankingów (przedstawioną w kolumnach 4-6. w każdej z tab. 4-8).

Tabela 4. Wyniki analizy i rankingi modeli regresji dla zbioru *Autompg*

Etap 1.			Etap 2.		
Ranking	Metoda	MSE_{CV}	Ranking	Metoda	MSE_{CV}
1	R. FORESTS	4,04	1	R. FORESTS	4,04
2	MART	5,55	2	MART	5,55
3	BAGGING	6,45	**3**	**BAGGING**	**6,45**
4	SVM	6,53	**3**	**SVM**	**6,53**
5	POLYMARS	7,45	***5***	***POLYMARS***	***7,45***
6	PPR	7,62	***5***	***PPR***	***7,62***
7	NNET	8,75	7	NNET	8,75
8	LM	11,11	8	LM	11,11

Źródło: opracowanie własne.

Tabela 5. Wyniki analizy i rankingi modeli regresji dla zbioru *Boston*

Etap 1.			Etap 2.		
Ranking	Metoda	MSE_{CV}	Ranking	Metoda	MSE_{CV}
1	R. FORESTS	5,74	1	R. FORESTS	5,74
2	MART	8,21	2	MART	8,21
3	BAGGING	10,15	**3**	**BAGGING**	**10,15**
4	PPR	10,31	**3**	**PPR**	**10,31**
5	POLYMARS	11,85	5	POLYMARS	11,85
6	SVM	12,31	6	SVM	12,31
7	NNET	14,13	7	NNET	14,13
8	LM	22,70	8	LM	22,70

Źródło: opracowanie własne.

Tabela 6. Wyniki analizy i rankingi modeli regresji dla zbioru *Clothing*

Etap 1.			Etap 2.		
Ranking	Metoda	MSE_{CV}	Ranking	Metoda	MSE_{CV}
1	PPR	$10525 \cdot 10^6$	1	PPR	$10525 \cdot 10^6$
2	SVM	$22417 \cdot 10^6$	2	SVM	$22417 \cdot 10^6$
3	MART	$38486 \cdot 10^6$	3	MART	$38486 \cdot 10^6$
4	R. FORESTS	$47579 \cdot 10^6$	4	R. FORESTS	$47579 \cdot 10^6$
5	BAGGING	$62471 \cdot 10^6$	5	BAGGING	$62471 \cdot 10^6$
6	NNET	$68114 \cdot 10^6$	6	NNET	$68114 \cdot 10^6$
7	LM	$82610 \cdot 10^6$	7	LM	$82610 \cdot 10^6$
8	POLYMARS	$94507 \cdot 10^9$	8	POLYMARS	$94507 \cdot 10^9$

Źródło: opracowanie własne.

Tabela 7. Wyniki analizy i rankingi modeli regresji dla zbioru *Ozone*

Etap 1.			Etap 2.		
Ranking	Metoda	MSE_{CV}	Ranking	Metoda	MSE_{CV}
1	R. FORESTS	8,93	1	R. FORESTS	8,93
2	MART	9,45	2	MART	9,45
3	BAGGING	11,27	**3**	**BAGGING**	**11,27**
4	SVM	11,67	**3**	**SVM**	**11,67**
5	NNET	13,08	5	NNET	13,08
6	POLYMARS	14,59	6	POLYMARS	14,59
7	PPR	17,06	7	PPR	17,06
8	LM	19,17	8	LM	19,17

Źródło: opracowanie własne.

Tabela 8. Wyniki analizy i rankingi modeli regresji dla zbioru *Star*

Etap 1.			Etap 2.		
Ranking	Metoda	MSE_{CV}	Ranking	Metoda	MSE_{CV}
1	R. FORESTS	1 812,1	1	R. FORESTS	1 812,1
2	MART	1 963,7	2	MART	1 963,7
3	PPR	1 988,3	3	PPR	1 988,3
4	NNET	2 037,8	**4**	**NNET**	**2 037,8**
5	BAGGING	2 041,7	**4**	**BAGGING**	**2 041,7**
6	SVM	2 052,2	***5***	***SVM***	***2 052,2***
7	POLYMARS	2 082,2	*7*	*POLYMARS*	*2 082,2*
8	LM	2 088,7	*7*	*LM*	*2 088,7*

Źródło: opracowanie własne.

W tych przypadkach, w których nie było podstaw do odrzucenia hipotez zerowych (zapisanych wzorami (1), (3)), wyniki analiz wyróżniono w tabelach 4-8 pogrubioną lub pochyłą czcionką. Dla zbioru *Autompg* nieistotnie różne okazały się wartości błędów średniokwadratowych obliczonych dla dwóch par modeli: zbudowanych za pomocą metod BAGGING i SVM oraz POLYMARS i PPR. Modele zbudowane na zbiorze *Boston* tylko w przypadku jednej pary metod – BAGGING i PPR, generowały błędy MSE_{CV}, których różnica była nieistotna. Analogiczny przypadek, tyle że dla metod BAGGING i SVM, uzyskano dla zbioru *Ozone*. Wartości MSE_{CV}, obliczone dla różnych modeli regresji zbudowanych na zbiorze *Clothing*, w każdym z przypadków różniły się istotnie pomiędzy sobą. Najciekawsze wyniki testowania uzyskano dla zbioru *Star*. Nieistotnie różniące się wartości błędu średniokwadratowego uzyskano dla modeli zbudowanych metodami NNET i BAGGING oraz BAGGING i SVM. Jednak różnica wartości MSE_{CV} dla modeli NNET i SVM okazała się istotna.

4. Podsumowanie

Nieparametryczne metody regresji nie wymagają znajomości analitycznych postaci związków między zmiennymi ani testowania normalności składnika losowego. Pozwalają na budowę modeli nieliniowych, również dla bardzo dużych zbiorów danych, charakteryzowanych przez wiele zmiennych objaśniających (dla których nie wprowadza się założeń o postaciach ich rozkładów). Ponadto metody wykorzystujące drzewa regresyjne, metoda krzywych sklejanych POLYMARS oraz metoda wektorów nośnych dopuszczają wprowadzanie do modelu zmiennych mierzonych na różnych skalach pomiaru. W związku z tym modele nieparametryczne charakteryzują się dużo większą elastycznością, a dodatkowo zakres ich potencjalnych zastosowań jest znacznie szerszy.

Do wad metod nieparametrycznych zaliczamy to, że ich odmienne mechanizmy działania powodują, iż niemożliwe staje się analityczne porównywanie tych metod. Stąd też ważne są próby badań porównawczych omawianych metod za pomocą procedur symulacyjnych.

W przeprowadzonym badaniu empirycznym modele charakteryzujące się najlepszymi wynikami dokładności predykcji to zazwyczaj modele zbudowane za pomocą drzew regresyjnych – najczęściej metodą RANDOM FORESTS, lecz dobre wyniki uzyskujemy również dla modeli MART i BAGGING. Należy jednak pamiętać, że badanie przeprowadzono jedynie na kilku zbiorach danych (standardowo wykorzystywanych do badania własności różnych metod regresji) i nie można wskazać żadnych wyników analitycznych porównań, które by udowodniły przewagę modeli zbudowanych za pomocą drzew regresyjnych nad pozostałymi modelami. Tym samym otrzymywane najniższe wartości błędów predykcji dla modeli wykorzystujących zagregowane drzewa regresyjne nie są regułą, co pokazuje przykład zbioru *Clothing*.

W każdym z analizowanych przypadków wartości MSE_{CV} dla najlepszego modelu są istotnie różne od wartości MSE_{CV} obliczonych dla modeli znajdujących się na niższych miejscach w rankingach. Oznacza to, że w przypadku badanych zbiorów danych, wybierając model najlepszy ze względu na własności predykcyjne, powinno się brać pod uwagę tylko ten, który znajduje się na szczycie rankingu. Wybór innego modelu, choćby takiego, który dawałby większe możliwości interpretacyjne, oznacza zgodę na istotnie większy błąd predykcji.

Literatura

Bishop C. (1995), *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford.

Blum A., Kalai A., Langford J. (1999), *Beating the hold-out: bounds for K-fold and progressive cross-validation*, „COLT", s. 203-208.

Breiman L. (1996), *Bagging predictors*, „Machine Learning", 24, s. 123-140.

Breiman L. (2001), *Random forests*, „Machine Learning, 45, s. 5-32.

Friedman J. (1999a), *Greedy Function Approximation: a Gradient Boosting Machine*, Technical Report, Stanford University, Dept. of Statistics.

Friedman J. (1999b), *Stochastic Gradient Boosting*, Technical Report, Stanford University, Dept. of Statistics.

Friedman J., Stuetzle W. (1981), *Projection pursuit regression*, „Journal of the American Statistical Association", 76, s. 817-823.

Hothorn T., Leisch F., Zeileis A., Hornik K. (2005), *The design and analysis of benchmark experiments*, „Journal of Computational and Graphical Statistics", 14(3), s. 675-699.

Kohavi R. (1995), *A study of cross-validation and bootstrap for accuracy estimation and model selection*, „IJCAI", s. 1137-1145.

Kooperberg C., Bose S., Stone C. (1997), *Polychotomous regression*, „Journal of the American Statistical Association", 92, s. 117-127.

Meyer D., Leisch F., Hornik K. (2003), *The support vector machine under test*, „Neurocomputing", 55(1-2), s. 169-186.

Vapnik V. (1998), *Statistical learning theory*, „Adaptive and Learning Systems for Signal Processing, Communications, and Control", John Wiley & Sons, Nowy Jork.

ON SOME SIMULATIVE PROCEDURES FOR COMPARING NONPARAMETRIC METHODS OF REGRESSION

Summary: The paper presents the simulative procedure for comparing the performance of several competing algorithms of nonparametric regression. This procedure has two stages. In the first one, the ranking of nonparametric models of regression is created. In the second stage, statistical test procedures can be used to test the significance of differences in the performances of models presented in the ranking. The procedure is applied to regression benchmark studies based on real world data.

Keywords: nonparametric regression, model comparison, benchmarking experiments, hypothesis testing.