

**Joanna Trzęsiok**

Akademia Ekonomiczna w Katowicach

## **OCENA ZASADNOŚCI ŁĄCZENIA WYBRANYCH NIEPARAMETRYCZNYCH MODELI REGRESJI**

### **1. Wstęp**

W ostatnich latach zarówno w regresji, jak i dyskryminacji coraz częściej wykorzystywane jest podejście wielomodelowe. Do prognozowania stosowane są wskazania wielu modeli skonstruowanych na losowo dobranych podpróbach, których wyniki predykcji łączone są według określonej formuły, w regresji najczęściej przez uśrednianie.

Celem podejścia wielomodelowego jest poprawa dokładności predykcji. Jak wykazują badania symulacyjne i wyniki teoretyczne, predykcja modelu zagregowanego jest na ogół dokładniejsza niż każdego z pojedynczych modeli składowych. Nie jest tak jednak zawsze. W 1996 r. Tumer i Ghosh udowodnili, że łączenie równoległe modeli (*bagging*) przynosi największą poprawę dokładności, gdy predykcja modeli indywidualnych, zbudowanych na losowo dobranych podpróbach uczących, charakteryzuje się relatywnie silnym zróżnicowaniem.

Celem artykułu było zweryfikowanie hipotezy o zasadności łączenia wybranych modeli regresji otrzymanych metodami nieparametrycznymi. Zachodzi pytanie, czy poprzez agregację tych modeli uzyskamy istotną poprawę wyników predykcji. W pierwszym etapie analizy mającej na celu zweryfikowanie powyższej hipotezy zbadano zróżnicowanie modeli regresji zbudowanych za pomocą metod: MARS, PPR, ACE i AVAS. W drugim etapie porównano dokładność predykcji modeli indywidualnych oraz zagregowanych.

### **2. Wybrane nieparametryczne modele regresji wykorzystane w analizie**

Jak pokazują badania, modele zbudowane za pomocą nieparametrycznej metody drzew regresyjnych charakteryzują się silnym zróżnicowaniem, a agregacja tych

modeli, np. poprzez metodę MART, przynosi dużą poprawę dokładności predykcji [Friedman 1999; Gatnar 2005]. W niniejszym artykule przedstawiona została ocena zasadności agregowania wyników predykcji modeli uzyskanych za pomocą innych nieparametrycznych metod regresji:

- wielowymiarowej metody krzywych sklepanych MARS,
- metody rzutowania PPR,
- metod opartych na jednoczesnej transformacji wszystkich zmiennych: ACE i AVAS.

### 2.1. Wielowymiarowa metoda krzywych sklepanych MARS

Wielowymiarowa metoda MARS (*multivariate adaptive regression splines*) została zaproponowana przez Friedmana w 1991 r. Natomiast w 1997 r. Kooperberg, Bose i Stone przedstawili jej modyfikację – metodę POLYMARS.

Metoda MARS, jako metoda nieparametryczna, nie wymaga znajomości rozkładów badanych zmiennych ani analitycznych postaci związków między nimi. Ważną jej zaletą jest to, że pozwala ona na wprowadzanie do modelu zarówno zmiennych metrycznych, jak i niemetrycznych.

Niech  $\mathbf{X} = (X_1, \dots, X_N)$  będzie  $N$ -wymiarowym wektorem zmiennych objaśniających,  $Y$  zaś zmienną objaśnianą. Przez  $x_{ij}$  oznaczmy  $i$ -tą realizację zmiennej  $X_j$  (dla  $j = 1, \dots, N$  oraz  $i = 1, \dots, n$ ).

Nieparametryczna metoda MARS oparta jest na funkcjach sklepanych pierwszego rzędu postaci:

$$(X_j - \xi_j)_+ = \begin{cases} X_j - \xi_j, & \text{dla } X_j \geq \xi_j \\ 0, & \text{dla } X_j < \xi_j \end{cases} \quad (1)$$

gdzie punkt  $\xi_j$  jest węzłem, czyli punktem podziału dziedziny zmiennej  $X_j$  (dla  $j = 1, \dots, N$ ), w którym funkcje typu (1) zostają „sklejone”.

Model regresyjny w tej metodzie można przedstawić w postaci addytywnej:

$$f(\mathbf{X}) = \alpha_0 + \sum_{m=1}^M \alpha_m h_m(\mathbf{X}), \quad (2)$$

gdzie funkcje  $h_m$  przedstawione we wzorze (2) są iloczynami tensorowymi funkcji sklepanych (1),

$$h_m(\mathbf{X}) = \prod_{p=1}^2 u_m(X_p - \xi_p)_+ \quad (3)$$

dla  $u_m \in \{-1, 1\}$ .

W omawianej metodzie algorytm budowy modelu regresyjnego składa się z dwóch głównych etapów: dołączania zmiennych do modelu oraz ich eliminacji. Procedury te zostały szczegółowo omówione w pracy [Trzęsiok 2004a].

## 2.2. Metoda rzutowania PPR

Metoda rzutowania (*projection pursuit regression*) została zastosowana po raz pierwszy w zagadnieniu regresji przez J. Friedmana i W. Stuetzle'a (zob. [Friedman, Stuetzle 1981]). Celem tej metody jest transformacja danych z przestrzeni wielowymiarowej w przestrzeń o niższym wymiarze, w której łatwiej jest badaczowi zaobserwować pewne własności analizowanego zbioru obserwacji. Transformacja ta odbywa się poprzez zrzutowanie wektora zmiennych objaśniających  $\mathbf{X}$  w kierunkach  $\mathbf{a}_k$ . Otrzymujemy w ten sposób nowe zmienne:

$$Z_k = \mathbf{a}_k^T \cdot \mathbf{X}, \text{ dla } k = 1, \dots, K, \quad (4)$$

gdzie  $\mathbf{a}_k \in \mathbf{R}^n$  są unormowanymi wektorami, nazywanymi kierunkami rzutowania.

Model zbudowany za pomocą metody rzutowania ma postać addytywną:

$$Y = f(\mathbf{X}) = \alpha_0 + \sum_{k=1}^K g_k(\mathbf{a}_k^T \cdot \mathbf{X}, \boldsymbol{\beta}_k). \quad (5)$$

Funkcje składowe modelu –  $g_k$  (dla  $k = 1, \dots, K$ ) są funkcjami jednej zmiennej o parametrach zapisanych w postaci wektorów  $\boldsymbol{\beta}_k$ . Estymatory współrzędnych wektorów parametrów  $\boldsymbol{\beta}_k$ , jak i kierunków rzutowania  $\mathbf{a}_k$ , otrzymujemy w kolejnych krokach algorytmu [Trzęsiok 2004b] poprzez minimalizację ryzyka empirycznego:

$$R_{emp}(\mathbf{a}, \boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n (y_i - f(\mathbf{x}_i))^2, \quad (6)$$

gdzie  $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_K)$  oraz  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_K)$ .

## 2.3. Metody transformujące zmienne: ACE i AVAS

W wielu przypadkach uzyskany model regresyjny charakteryzuje się słabym dopasowaniem do danych i dużymi błędami prognoz. Jednym ze sposobów uzyskania modelu bardziej dokładnego jest zastosowanie transformacji zmiennych. Metody ACE (*alternating conditional expectation*) [Breiman, Friedman 1985] oraz AVAS (*additivity and variance stabilization*) [Tibshirani 1988] oparte są na jednoczesnej transformacji wszystkich zmiennych, a model rozważany w tych metodach można zapisać w postaci:

$$\theta(Y) = \sum_{j=1}^N f_j(X_j) + \varepsilon. \quad (7)$$

Funkcje  $\theta$  oraz  $f_j$  (dla  $j=1, \dots, N$ ) są funkcjami mierzalnymi spełniającymi warunki:

$$E\theta^2(Y) < \infty, E f_j^2(X_j) < \infty, \text{ dla } j=1, \dots, N, \quad (8)$$

$$E\theta(Y) = 0, E f_j(X_j) = 0, \text{ dla } j=1, \dots, N, \quad (9)$$

przy czym nie zakładamy tutaj monotoniczności transformacji  $f_j$  ani tym bardziej znajomości analitycznych postaci tych przekształceń. W metodzie AVAS wymagane jest jedynie, aby funkcja  $\theta$  była rosnąca.

W metodzie ACE transformacje  $\theta$  oraz  $f_j$  wyznaczone są w ten sposób, by minimalizowany był współczynnik zbieżności przedstawiony wzorem (10):

$$e^2(\theta, f_1, \dots, f_N) = \frac{E \left( \left( \theta(Y) - \sum_{j=1}^N f_j(X_j) \right)^2 \right)}{E\theta^2(Y)}. \quad (10)$$

Natomiast celem metody AVAS jest uzyskanie funkcji  $\theta$  oraz  $f_j$  stabilizujących wariancję przekształconej zmiennej objaśnianej, czyli spełniających warunki:

$$E(\theta(Y) | X_1, \dots, X_N) = \sum_{j=1}^N f_j(X_j), \quad (11)$$

$$\text{var} \left( \theta(Y) \middle| \sum_{j=1}^N f_j(X_j) \right) = \text{const.} \quad (12)$$

Szczegółowe algorytmy, za pomocą których otrzymujemy model regresyjny (7) oraz szukane funkcje transformujące zmienne, przedstawione zostały w pracach [Breiman, Friedman 1985; Tibshirani 1988].

### 3. Procedura badawcza i wyniki analizy

Przeprowadzone badania porównawcze pokazują, że modele uzyskane za pomocą nieparametrycznych metod regresji, a szczególnie poprzez metody MARS i PPR, charakteryzują się relatywnie dużą dokładnością predykcji [Meyer, Leisch, Hornik 2002]. Mimo to można oczekiwać, iż przez agregację uzyskamy modele

dające jeszcze mniejsze błędy prognoz. Będą to jednak modele bardziej złożone, dla których oszacowanie wartości parametrów zajmie więcej czasu.

Jak już wspomniano, zostało udowodnione, że łączenie równoległe pojedynczych modeli przynosi największą poprawę dokładności w przypadku, gdy predykcja modeli indywidualnych wykazuje silne zróżnicowanie. Zatem weryfikacja hipotezy o zasadności stosowania podejścia wielomodelowego w przypadku rozważanych metod nieparametrycznych jest ściśle związana z badaniem zróżnicowania modeli indywidualnych. Analiza zróżnicowania wartości predykcji modeli pojedynczych, zbudowanych na próbach wylosowanych ze zbioru uczącego, stanowi pierwszy etap procedury badawczej. W drugim etapie porównano błędy średniokwadratowe, obliczone na zbiorach testowych dla modelu zagregowanego oraz indywidualnego, zbudowanego na całym zbiorze uczącym.

Analiza została przeprowadzona na zbiorach danych standardowo wykorzystywanych do badania własności różnych nieparametrycznych metod regresji. Były to dwa rzeczywiste zbiory danych: *Boston* oraz *Autompg* oraz trzy zbiory danych sztucznych: *Friedman 1*, *Friedman 2*, *Friedman 3* [Friedman 1991]. Wybrane charakterystyki tych zbiorów przedstawiono w tab. 1.

Tabela 1. Charakterystyki zbiorów danych wykorzystanych w analizie

Zbiór danych	Liczba zmiennych	Liczba zmiennych niemetrycznych	Liczba obserwacji
<i>Boston</i>	14	1	506
<i>Autompg</i>	9	3	392
<i>Friedman 1</i>	10	0	1000
<i>Friedman 2</i>	4	0	1000
<i>Friedman 3</i>	4	0	1000

Źródło: opracowanie własne.

Procedurę badawczą można przedstawić w następujących krokach:

1. W każdym zbiorze danych wyodrębniono część uczącą i testową:

a) rzeczywiste zbiory danych podzielono losowo na część uczącą  $U$  (67% obserwacji) oraz testową  $T$  (33% obserwacji);

b) część ucząca w sztucznych zbiorach danych zawierała 1000 obserwacji i dodany został do niej szum gaussowski  $e$  na poziomie 20% zmienności mierzonej wariancją. Zbiór testowy składał się z 200 elementów i nie zawierał szumu.

2. Z  $n$ -elementowego zbioru uczącego wylosowano 100 prób bootstrapowych  $D_1, \dots, D_{100}$ .

3. Dla każdej z próbek  $D_k$  zbudowano model regresyjny za pomocą wybranej nieparametrycznej metody regresji i dokonano predykcji na zbiorze testowym. W efekcie uzyskano macierz wartości teoretycznych o wymiarach  $\overline{T} \times 100$ , gdzie  $\overline{T}$  oznacza moc zbioru testowego.

4. Na podstawie wartości teoretycznych  $f_D(\mathbf{x})$ , dla każdego  $\mathbf{x} \in T$ , obliczono wariancję:

$$S^2(\mathbf{x}) = E_D \left[ \left( f_D(\mathbf{x}) - E_D [f_D(\mathbf{x})] \right)^2 \right] \quad (13)$$

oraz współczynnik zmienności  $V(\mathbf{x})$ , a następnie wyznaczono przeciętne wartości wariancji i współczynnika zmienności na całym zbiorze testowym:

$$S^2 = E_T [S^2(\mathbf{x})], \quad V = E_T [V(\mathbf{x})]. \quad (14)$$

5. Dokonano agregacji modeli poprzez uśrednienie wartości predykcji otrzymanych dla 100 prób bootstrapowych, a następnie dla uzyskanego modelu zagregowanego obliczono na zbiorze testowym błąd średniokwadratowy  $MSE_{agr}$ .

6. Na zbiorze testowym obliczono również błąd średniokwadratowy  $MSE$  dla pojedynczego modelu zbudowanego na całym zbiorze uczącym.

7. Porównano otrzymane miary dopasowania modelu zagregowanego z modelem pojedynczym zbudowanym na całym zbiorze uczącym.

Wyniki analizy przedstawiono w tab. 2-5.

Tabela 2. Wyniki analizy dla modeli uzyskanych metodą MARS

Zbiór danych	Pojedynczy model	Zagregowany model		
	$MSE$	$MSE_{agr}$	$S^2$	$V$
<i>Boston</i>	15,51	11,73	8,03	0,11
<i>Autompg</i>	8,71	7,90	1,38	0,05
<i>Friedman 1</i>	2,72	2,02	1,16	0,08
<i>Friedman 2</i>	161,38	144,22	720,15	0,14
<i>Friedman 3</i>	0,009	0,008	0,003	0,04

Źródło: opracowanie własne.

Tabela 3. Wyniki analizy dla modeli uzyskanych metodą PPR

Zbiór danych	Pojedynczy model	Zagregowany model		
	$MSE$	$MSE_{agr}$	$S^2$	$V$
<i>Boston</i>	21,38	13,26	11,79	0,16
<i>Autompg</i>	8,71	7,52	4,93	0,10
<i>Friedman 1</i>	6,14	5,35	1,80	0,10
<i>Friedman 2</i>	2580,43	3679,65	5499,29	0,28
<i>Friedman 3</i>	0,02	0,01	0,01	0,12

Źródło: opracowanie własne.

Tabela 4. Wyniki analizy dla modeli uzyskanych metodą ACE

Zbiór danych	Pojedynczy model	Zagregowany model		
	$MSE$	$MSE_{agr}$	$S^2$	$V$
<i>Boston</i>	12,08	11,29	4,00	0,10
<i>Autompg</i>	7,70	8,19	1,03	0,04
<i>Friedman 1</i>	1,97	1,96	0,26	0,06
<i>Friedman 2</i>	24 711,82	25 967,36	1239,56	0,04
<i>Friedman 3</i>	0,04	0,05	0,01	0,04

Źródło: opracowanie własne.

Tabela 5. Wyniki analizy dla modeli uzyskanych metodą AVAS

Zbiór danych	Pojedynczy model	Zagregowany model		
	$MSE$	$MSE_{agr}$	$S^2$	$V$
<i>Boston</i>	14,47	35,82	143,39	0,47
<i>Autompg</i>	7,87	13,07	17,05	0,11
<i>Friedman 1</i>	2,02	2,01	0,26	0,13
<i>Friedman 2</i>	20 210,88	23 606,92	2218,15	0,07
<i>Friedman 3</i>	0,02	0,02	0,003	0,03

Źródło: opracowanie własne.

#### 4. Wnioski

Wyznaczone wartości współczynnika zmienności  $V$  przedstawione w tab. 2-5 wskazują na niskie zróżnicowanie badanych nieparametrycznych metod regresji. Otrzymane modele charakteryzują się więc wysoką stabilnością i tym samym możemy się spodziewać niewielkiej poprawy dokładności predykcji poprzez zastosowanie podejścia wielomodelowego. Przypuszczenie to potwierdzają przedstawione w tab. 2-5 wartości błędu średniokwadratowego.

Porównując wartości  $MSE$  dla poszczególnych metod i zbiorów danych, dla modelu zagregowanego oraz indywidualnego zbudowanego na całym zbiorze uczącym, widzimy, że w większości przypadków następuje spadek tych wartości. Zazwyczaj jednak wartość analizowanego błędu średniokwadratowego nie zmniejsza się znacznie, a zdarza się nawet, że dla modelu pojedynczego ma on wartość niższą niż dla modelu zagregowanego. Zatem zastosowanie procedury agregacji nie poprawiło istotnie dokładności predykcji.

Przeprowadzona analiza pokazuje, że za pomocą rozpatrywanych nieparametrycznych metod regresji otrzymujemy modele stabilne, charakteryzujące się relatywnie wysoką dokładnością predykcji, których nie trzeba agregować. Stosowanie w tym przypadku podejścia wielomodelowego zwiększa jedynie złożoność otrzymywanego modelu regresyjnego.

## Literatura

- Breiman L., Friedman J. (1985), *Estimating Optimal Transformations for Multiple Regression and Correlation (with discussion)*, „Journal of the American Statistical Association” nr 80, s. 580-619.
- Friedman J., Stuetzle W. (1981), *Projection Pursuit Regression*, „Journal of the American Statistical Association” nr 76, s. 817-823.
- Friedman J.H. (1991), *Multivariate Adaptive Regression Splines*, „Annals of Statistics” nr 19, s.1-141.
- Friedman J.H. (1999), *Greedy Function Approximation: a Gradient Boosting Machine*, Technical report, Department of Statistics, Stanford University.
- Gatnar E. (2005), *Gradient Boosting in Regression*, „Acta Universitatis Lodzianis”, Folia Oeconomica 194, Łódź, s. 11-19.
- Kooperberg C., Bose S., Stone C.J. (1997), *Polychotomous Regression*, „Journal of the American Statistical Association” nr 92, s. 117-127.
- Meyer D., Leisch F., Hornik K. (2002), *Benchmarking Support Vector Machines*, Report no. 78, Vienna University of Economics and Business Administration, <http://www.wu.wien.ac.at/am/Download/report78.pdf>.
- Tibshirani R. (1988), *Estimating Transformations for Regression Via Additivity and Variance Stabilization*, „Journal of the American Statistical Association” nr 83, s. 394-405.
- Trzęsiok J. (2004a), *Wybrane nieparametryczne metody regresji i ich zastosowania*, [w:] Taksonomia 11, red. K. Jajuga, M. Walesiak, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1022, AE, Wrocław, s. 107-115.
- Trzęsiok J. (2004b), *Metoda rzutowania w budowie modelu regresyjnego*, [w:] *Postępy ekonometrii*, red. A.S. Barczak, AE, Katowice, s. 121-130.
- Tumer K., Ghosh J. (1996), *Analysis of Decision Boundaries in Linearly Combined Neural Classifiers*, „Pattern Recognition” nr 29, s. 341-348.

## ON AGGREGATING SOME NONPARAMETRIC REGRESSION MODELS

### Summary

The main goal of the paper is to analyze the feasibility of gaining more accurate predictions by aggregating nonparametric regression models such as MARS, PPR, ACE and AVAS. The aggregating approach is especially effective when single models are weak i.e. the variance of predictions of these models is high. The analysis was performed on benchmark data sets. The results showed that considered models had the built-in variance reduction mechanism, so the increase in the prediction accuracy of aggregating models compared to single models was not significant.