Joanna Trzęsiok

OCENA WPŁYWU ZMIENNYCH OBJAŚNIAJĄCYCH NA ZMIENNĄ ZALEŻNĄ W METODZIE RZUTOWANIA PPR

Wprowadzenie

Metoda rzutowania PPR (projection pursuit regression), zaproponowana przez J. Friedmana i W. Stuetzle'a w 1981 roku, jest jedną z nieparametrycznych metod regresji. Przeprowadzone badania porównawcze pokazują, iż modele regresji otrzymane za jej pomocą charakteryzują się często wyższą dokładnością predykcji niż modele generowane przez inne metody, zarówno nieparametryczne, jak i klasyczne (zob. [4; 6; 7]). Jest to jednak jedna z metod, które często określa się mianem „czarnej skrzynki". Wyniki, które otrzymuje się przy jej użyciu, nie są zazwyczaj interpretowalne, dlatego tak ważne są wszelkie próby uzyskania dodatkowych informacji z otrzymanego modelu.

W artykule przedstawiono dwie procedury: eliminacji i dołączania zmiennych, które redukują złożoność modelu otrzymanego metodą PPR, pozwalają na wyodrębnienie zmiennych, które mają największy wpływ na zmienną zależną, jak również powiększają zasób informacji uzyskanych ze zbudowanego modelu.

Celem artykułu, jak również wspomnianych procedur eliminacji i dołączania zmiennych będzie zbudowanie rankingu zmiennych objaśniających pod względem ich siły wpływu na zmienną Y.

1. Metoda rzutowania PPR

Celem metody rzutowania 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 $\boldsymbol{\alpha}_k$. Uzyskuje się w ten sposób nowe zmienne:

$$Z_k = \boldsymbol{\alpha}_k^T \cdot \mathbf{X}, \text{ dla } k = 1,\ldots,K, \tag{1}$$

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

Model regresyjny, zbudowany za pomocą metody rzutowania, można przedstawić w postaci addytywnej:

$$Y = f(\mathbf{X}) = \alpha_0 + \sum_{k=1}^{K} \beta_k g_k\left(\boldsymbol{\alpha}_k^T \cdot \mathbf{X}\right). \tag{2}$$

Funkcje składowe modelu g_k (dla $k = 1,\ldots,K$) to funkcje jednej zmiennej, nazywane funkcjami grzbietowymi, o parametrach β_k. Estymatory tych parametrów, a także kierunków rzutowania $\boldsymbol{\alpha}_k$ otrzymuje się w kolejnych krokach algorytmu poprzez minimalizację błędu empirycznego (empirical risk):

$$R_{emp}(\boldsymbol{\alpha},\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - f(\mathbf{x}_i)\right)^2, \tag{3}$$

gdzie $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_K)$ oraz $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_K)$.

Błąd empiryczny przedstawiony we wzorze (3) można przekształcić do następującej postaci (zob. [5]):

$$R_{emp}(\boldsymbol{\alpha},\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\left(r_i - \beta_k g_k\left(\boldsymbol{\alpha}_k^T \cdot \mathbf{x}_i\right)\right)^2, \tag{4}$$

gdzie:

$$r_i = y_i - \alpha_0 - \sum_{l \neq k} \beta_l g_l\left(\boldsymbol{\alpha}_l^T \cdot \mathbf{x}_i\right) \text{ dla } i = 1,\ldots,n. \tag{5}$$

Otrzymano w ten sposób dekompozycję błędu empirycznego na dwa składniki: resztę częściową r_i opisującą zmienność, która nie została wyjaśniona przez funkcje składowe g_l (dla $l \neq k$), oraz funkcję g_k.

Algorytm minimalizacji błędu R_{emp} jest pewnym uogólnieniem metody wykorzystującej sprzężenie zwrotne (backfitting algorithm) i składa się z następujących kroków [1, s. 255-259; 5]:

1. Ustal początkowe wartości współrzędnych wektorów $\boldsymbol{\alpha}_k$ oraz β_k (dla $k = 1,\ldots,K$) tak, aby:

$$\beta_k g_k\left(\boldsymbol{\alpha}_k^T \cdot \mathbf{x}_i\right) \equiv 0, \text{ dla } i = 1,\ldots,n. \tag{6}$$

Przyjmij:

$$\alpha_0 = \frac{1}{n}\sum_{i=1}^{n} y_i \,. \tag{7}$$

2. Dla każdego $k = 1,...,K$ wykonaj następujące kroki:
   a) Oblicz reszty częściowe:

$$r_i = y_i - \alpha_0 - \sum_{l \neq k} \beta_l g_l\left(\boldsymbol{\alpha}_l^T \cdot \mathbf{x}_i\right), \text{ dla } i = 1,...,n \,. \tag{8}$$

   b) Wykonaj rzutowanie, aż do osiągnięcia zbieżności:
      – ustal kierunek rzutowania $\boldsymbol{\alpha}_k$ i znajdź parametr β_k minimalizujący wyrażenie:

$$\mathrm{R}_{\mathrm{emp}}(\alpha,\beta) = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{r}_{\mathrm{i}} - \beta_{\mathrm{k}}\mathrm{g}_{\mathrm{k}}\left(\alpha_{\mathrm{k}}^{\mathrm{T}} \cdot \mathrm{x}_{\mathrm{i}}\right)\right)^2, \tag{9}$$

      – zmień współrzędne wektora $\boldsymbol{\alpha}_k$ w kierunku wyznaczonym przez wyrażenie:

$$\boldsymbol{\alpha}_k \leftarrow \boldsymbol{\alpha}_k - \gamma \cdot \nabla R_{emp}(\boldsymbol{\alpha}_k) \,, \tag{10}$$

      gdzie $\gamma > 0$.
3. Zakończ wykonywanie algorytmu, gdy jest spełnione ustalone wcześniej kryterium stopu lub gdy wartość funkcji (3) nie zmienia się znacząco.

Najmniej ugruntowanym elementem metody rzutowania jest wybór liczby funkcji składowych K w modelu regresyjnym (2). Wartość tego parametru jest zazwyczaj podawana przez użytkownika.

Poprawę jakości budowanego modelu można uzyskać wykorzystując algorytm SMART (smooth multiply additive regression), w którym badacz podaje dwie wartości parametru K:

– K_{pocz} – początkową (maksymalną) liczbę funkcji składowych,
– K_{konc} – liczbę funkcji użytych w końcowym modelu.

Stworzony zostaje model złożony z K_{pocz} funkcji g_k, który z wykorzystaniem przyjętego kryterium zostaje stopniowo przycinany, aż do uzyskania funkcji f, która jest sumą K_{konc} składowych g_k.

2. Procedura doboru zmiennych objaśniających do modelu zbudowanego metodą rzutowania PPR

Metoda rzutowania, jako nieparametryczna metoda regresji, nie zakłada znajomości rozkładu składnika losowego w modelu czy analitycznych postaci związków między zmiennymi. Jest narzędziem, które nie wymaga spełnienia wielu restrykcyjnych założeń, przez co znacząco został zwiększony jej obszar zastosowań. W praktyce często stosuję się tę metodę do analizy zbiorów danych charakteryzowanych przez dużą liczbę zmiennych. Otrzymuje się wtedy złożony model, którego współczynniki nie są interpretowalne.

Zastosowanie procedury doboru zmiennych objaśniających do modelu pozwala na istotną redukcję liczby zmiennych, a co za tym idzie – złożoności modelu. Okazuje się również, że procedura ta pozwala na poprawę dokładności predykcji, a także stworzenie rankingu zmiennych pod względem ich siły wpływu na zmienną zależną. W ten sposób badacz, czy decydent, otrzymuje prostszy model, dający mniejsze błędy prognoz, jak również dodatkową informację o tym, które zmienne są najbardziej istotne dla tego modelu.

W artykule przedstawiono dwa warianty procedury doboru zmiennych objaśniających do modelu regresyjnego: eliminację zmiennych oraz dołączanie zmiennych.

2.1. Procedura eliminacji zmiennych z modelu

Procedura eliminacji zmiennych opiera się na strategii wspinaczki. W pierwszym etapie tej procedury zostaje zbudowany model na oryginalnym zbiorze wszystkich zmiennych. W każdym kolejnym kroku zostaje usunięta jedna zmienna według ustalonego a priori kryterium i jest budowany model na pomniejszonym zbiorze zmiennych. Wykorzystywanym kryterium jest w tym przypadku minimalny błąd średniokwadratowy liczony metodą sprawdzania krzyżowego. W ten sposób kolejno są eliminowane zmienne, które mają najmniejszy wpływ na zmienną zależną. Procedura jest powtarzana tak długo, aż w zbiorze zostanie tylko jedna zmienna. Ta właśnie zmienna ma najsilniejszy wpływ na zmienną Y.

Procedurę eliminacji zmiennych z modelu można przedstawić w następujących krokach:

1. Za pomocą metody rzutowania PPR zbuduj model regresyjny f_0, wykorzystując kompletny zbiór zmiennych objaśniających:

$$V_0 = \{X_1, X_2, \ldots, X_m\}.$$

2. Dla $j = 1, \ldots, m-1$ wykonaj następujące kroki:

a) Ze zbioru zmiennych objaśniających V_{j-1} usuń tymczasowo jedną zmienną, wykonując tę czynność kolejno dla każdej ze zmiennych, i zbuduj $(m-j+1)$ modeli regresyjnych za pomocą metody PPR.
b) Dla wszystkich zbudowanych w poprzednim kroku modeli oblicz, metodą sprawdzania krzyżowego z podziałem zbioru danych na pięć części, błąd średniokwadratowy.
c) Ostatecznie w kroku j wyeliminuj zmienną, której usunięcie w najmniejszym stopniu zmieniło dokładność predykcji modelu, a więc tą, dla której obliczony błąd średniokwadratowy jest najmniejszy. Zredukowany zbiór zmiennych oznacz przez V_j, natomiast uzyskany najmniejszy błąd średniokwadratowy zapamiętaj jako MSE_j.
d) Przyjmij jako model f_j ten model regresyjny, który był zbudowany na zbiorze zmiennych oznaczonym przez V_j i któremu odpowiada błąd średniokwadratowy MSE_j.

3. Z otrzymanego ciągu modeli regresyjnych $\{f_j\}_{j=0,\ldots,m-1}$ (z malejącą liczbą zmiennych) wybierz ten model, dla którego błąd średniokwadratowy MSE_j jest najmniejszy. Jest to model końcowy zbudowany za pomocą metody rzutowania PPR z wykorzystaniem procedury eliminacji zmiennych.

W każdym kroku tej procedury zostaje wyeliminowana jedna zmienna, ta, która ma najmniejszy wpływ na zmienną zależną. Otrzymuje się zatem ranking zmiennych pod względem ich siły wpływu na zmienną Y, gdzie najbardziej istotna jest zmienna, która pozostaje na końcu w zbiorze zmiennych.

Obliczany na każdym etapie błąd średniokwadratowy pozwala na wybranie takiego modelu, któremu jest przyporządkowany najmniejszy MSE. Zmienne wykorzystane do budowy tego modelu to zmienne, które mają istotny wpływ na zmienną zależną. Pozostałe to zmienne redundantne.

W wyniku zastosowania procedury eliminacji otrzymuje się model, który jest rozwiązaniem optymalnym jedynie w sensie lokalnym. Zaletą tego podejścia jest jednak stosunkowo niska złożoność algorytmu.

2.2. Przykład ilustrujący procedurę eliminacji zmiennych z modelu

Przedstawiona procedura eliminacji zmiennych z modelu, zbudowanego metodą rzutowania, zostanie przedstawiona na przykładzie zbioru danych *Boston*. Obserwacje przedstawione w tym zbiorze zostały zebrane i opublikowane w 1978 roku przez Harrisona oraz Rubinfelda, badaczy, którzy zajmowali się wykrywaniem zależności pomiędzy cenami nieruchomości w Bostonie a jako-

ścią życia. Jest to zbiór szeroko znany i wykorzystywany do sprawdzania jakości modeli regresyjnych. Zgromadzone dane są charakteryzowane przez trzynaście zmiennych objaśniających:

crim – wskaźnik przestępstw,
zn – frakcja obszarów zaludnionych przekraczających 25 000 stóp kwadratowych,
indus – wskaźnik industrializacji,
chas – zmienna zero-jedynkowa wskazująca, czy teren znajduje się w pobliżu rzeki Charles,
nox – koncentracja tlenku azotu,
rm – średnia liczba pokoi,
age – procent budynków sprzed 1940 roku,
dis – ważona odległość do pięciu skupisk miejsc zatrudnienia w Bostonie,
rad – dostęp do autostrady,
tax – wysokość płaconych podatków,
ptratio – liczba uczniów na jednego nauczyciela,
black – procent ludności afroamerykańskiej,
lstat – procent ludności o niskim statusie społecznym.

Zmienną zależną jest Y = medv, czyli mediana wartości domu w tys. dolarów. Zbiór *Boston* składa się z 506 obserwacji.

Wyniki uzyskane poprzez zastosowanie procedury eliminacji przedstawiono w tabeli 1.

Tabela 1

Wyniki działania procedury eliminacji zmiennych

Etap	Wyeliminowana zmienna	Numery zmiennych usuniętych z modelu	*MSE*
0	–	–	14,089
1	crim	1	11,964
2	rad	1 9	14,350
3	chas	1 9 4	**11,331**
4	age	1 9 4 7	12,435
5	zn	1 9 4 7 2	12,338
6	indus	1 9 4 7 2 3	12,042
7	dis	1 9 4 7 2 3 8	14,055
8	black	1 9 4 7 2 3 8 12	14,992
9	tax	1 9 4 7 2 3 8 12 10	14,995
10	ptratio	1 9 4 7 2 3 8 12 10 11	17,814
11	nox	1 9 4 7 2 3 8 12 10 11 5	19,962
12	rm	1 9 4 7 2 3 8 12 10 11 5 6	27,242
13	lstat		

Błąd średniokwadratowy osiąga najmniejszą wartość, równą 11,331, dla modelu, z którego wyeliminowano zmienne: crim, rad, chas. Są to zmienne redundantne. Wprowadzenie ich do modelu powoduje zwiększenie wartości MSE oraz złożoności modelu. Pozostałe dziesięć zmiennych ma istotny wpływ na zmienną zależną i postać modelu.

Największy wpływ na medv ma zmienna, którą otrzymano w ostatnim, 13. kroku, natomiast najmniejsze znaczenie ma zmienna wyeliminowana w pierwszym etapie. Otrzymane wyniki pozwalają na stworzenie rankingu zmiennych objaśniających pod względem siły wpływu na zmienną zależną (zob. tabela 2).

Tabela 2

Ranking zmiennych objaśniających pod względem siły wpływu na zmienną zależną uzyskany za pomocą procedury eliminacji zmiennych

Nr w rankingu	Zmienne	
1	lstat	zmienne istotne
2	rm	
3	nox	
4	ptratio	
5	tax	
6	black	
7	dis	
8	indus	
9	zn	
10	age	
11	chas	zmienne redundantne
12	rad	
13	crim	

Największe znaczenie dla zmiennej medv ma tutaj zmienna lstat, tak więc największy wpływ na medianę wartości domu ma procent ludności o niskim statusie społecznym. Kolejną ważną zmienną jest rm – średnia liczba pokoi.

2.3. Procedura dołączania zmiennych do modelu

Alternatywnym podejściem do eliminacji zmiennych z modelu regresyjnego jest procedura dołączania zmiennych do modelu. Zaczyna się w tym przypadku od modelu zbudowanego dla jednej zmiennej, by sukcesywnie dołączać do niego kolejne zmienne i na końcu otrzymać model zbudowany na kompletnym zbiorze zmiennych.

W pierwszym etapie tej procedury buduje się m modeli dla pojedynczych zmiennych (gdzie m jest liczbą zmiennych objaśniających). Wybiera się z nich

najlepszy i w każdym kolejnym etapie dołącza się do niego zmienną według ustalonego a priori kryterium, którym ponownie jest minimalny błąd średniokwadratowy.

Procedurę dołączania zmiennych do modelu można przedstawić w następujący sposób:

1. Za pomocą metody rzutowania PPR zbuduj m modeli regresyjnych dla pojedynczych zmiennych objaśniających. Dla każdego modelu oblicz błąd średniokwadratowy metodą sprawdzania krzyżowego. Model, który odpowiada najmniejszej wartości MSE, przyjmij jako model początkowy f_1, zaś ze zmiennej wykorzystanej do budowy modelu f_1 stwórz początkowy, jednoelementowy zbiór zmiennych V_1. Pozostałe zmienne niech tworzą zbiór W_{m-1}.
2. Dla $j = 2,\ldots,m$ wykonaj kroki:
   a) Do zbioru zmiennych objaśniających V_{j-1} dodaj tymczasowo jedną zmienną ze zbioru W_{m-j+1}, wykonując tę czynność kolejno dla każdej zmiennej, i zbuduj $(m-j+1)$ modeli regresyjnych za pomocą metody PPR.
   b) Dla wszystkich zbudowanych w poprzednim kroku modeli oblicz, metodą sprawdzania krzyżowego z podziałem zbioru danych na pięć części, błąd średniokwadratowy.
   c) Ostatecznie w kroku j dołącz do modelu tę zmienną, dla której obliczony błąd średniokwadratowy jest najmniejszy. Powiększony zbiór zmiennych tworzących model oznacz przez V_j, pozostałe zmienne przez W_{m-j}. Uzyskany najmniejszy błąd średniokwadratowy zapamiętaj jako MSE_j.
   d) Przyjmij jako model f_j ten model regresyjny, który był zbudowany na zbiorze zmiennych oznaczonym przez V_j i któremu odpowiada błąd średniokwadratowy MSE_j.
3. Z otrzymanego ciągu modeli regresyjnych $\{f_j\}_{j=1,\ldots,m}$ (z rosnącą liczbą zmiennych) wybierz ten model, dla którego błąd średniokwadratowy MSE_j jest najmniejszy. Jest to model końcowy zbudowany za pomocą metody rzutowania PPR z wykorzystaniem procedury dołączania zmiennych.

Podobnie jak dla procedury eliminacji, można uzyskać ranking zmiennych objaśniających pod względem ich siły wpływu na zmienną zależną. Przy czym najistotniejsza tym razem jest zmienna otrzymana w pierwszym kroku procedury, natomiast najmniejsze znaczenie ma zmienna, którą dołącza się do modelu w ostatnim etapie. Zmienne, których nie wykorzystano do budowy modelu końcowego, to zmienne redundantne.

Model końcowy, otrzymany za pomocą procedury dołączania zmiennych, jest rozwiązaniem optymalnym jedynie w sensie lokalnym. Ponadto procedura dołączania zmiennych do modelu, ze względu na pierwszy etap – budowy modelu dla pojedynczej zmiennej, jest uważana za mniej stabilną niż metoda eliminacji zmiennych. Z tego też powodu jest ona rzadziej wykorzystywana w praktyce.

2.4. Przykład ilustrujący procedurę dołączania zmiennych do modelu

Ponownie, w celu ilustracji procedury dołączania zmiennych, wykorzystano zbiór danych *Boston*. Uzyskane wyniki przedstawiono w tabeli 3.

Tabela 3

Wyniki działania procedury dołączania zmiennych

Etap	Dołączona zmienna	Numery zmiennych wykorzystanych do budowy modelu	*MSE*
1	lstat	13	27,242
2	rm	13 6	19,962
3	tax	13 6 10	15,785
4	nox	13 6 10 5	14,316
5	black	13 6 10 5 12	14,134
6	age	13 6 10 5 12 7	14,135
7	dis	13 6 10 5 12 7 8	14,134
8	rad	13 6 10 5 12 7 8 9	14,134
9	ptratio	13 6 10 5 12 7 8 9 11	**14,133**
10	indus	13 6 10 5 12 7 8 9 11 3	15,643
11	chas	13 6 10 5 12 7 8 9 11 3 4	15,643
12	crim	13 6 10 5 12 7 8 9 11 3 4 1	16,298
13	zn	13 6 10 5 12 7 8 9 11 3 4 1 2	16,298

Największy wpływ na medianę wartości domu, tak samo jak poprzednio, mają zmienne: lstat oraz rm, które zostały dołączone do modelu w pierwszym i drugim kroku algorytmu. Kolejne zmienne w coraz mniejszym stopniu wpływają na zmienną zależną. Ranking wszystkich zmiennych przedstawiono w tabeli 4.

Model końcowy, w tym przypadku, to model, dla którego błąd średniokwadratowy jest równy 14,133. Do budowy tego modelu wykorzystano dziewięć zmiennych mających istotny wpływ na zmienną medv. Pozostałe zmienne: indus, chas, crim i zn są, w tym przykładzie, zmiennymi redundantnymi.

Tabela 4

Ranking zmiennych objaśniających pod względem siły wpływu na zmienną zależną uzyskany za pomocą procedury dołączania zmiennych

Nr w rankingu	Zmienne	
1	lstat	zmienne istotne
2	rm	
3	tax	
4	nox	
5	black	
6	age	
7	dis	
8	rad	
9	ptratio	
10	indus	zmienne redundantne
11	chas	
12	crim	
13	zn	

Wartość współczynnika Spearmana – zgodności uzyskanych rankingów wynosi:

$$r_S = 0{,}833.$$

Podsumowanie

W artykule przedstawiono dwie metody doboru zmiennych objaśniających do modelu regresyjnego: eliminację oraz dołączanie zmiennych. Pomimo wbudowanego w algorytmie metody PPR mechanizmu selekcji zmiennych opartego na rzutowaniu zastosowanie omawianych metod doboru zmiennych doprowadziło do poprawy dokładności predykcji modelu. Wykorzystanie mniejszej liczby zmiennych dało w konsekwencji mniej skomplikowany model końcowy.

Systematyczna eliminacja lub dołączanie zmiennych pozwoliły na zbudowanie rankingu zmiennych objaśniających pod względem: ich siły wpływu na zmienną zależną oraz zdolności poprawiania jakości modelu PPR. W tym przypadku można także oddzielić zmienne istotne od zmiennych redundantnych. Otrzymany ranking jest również dodatkową, ważną informacją dla badacza czy decydenta posługującego się w analizie regresji metodą rzutowania PPR.

Literatura

1. Cherkassky V., Mulier F.: *Learning from Data – Concepts, Theory, and Methods*. Wiley, New York 1998.
2. Friedman J.H., Stuetzle W.: *Projection Pursuit Regression*. „Journal of the American Statistical Association" 1981, No. 76, s. 817-823.
3. Harrison D., Rubinfeld D.L.: *Hedonic Prices and the Demand for Clean Air*. „Journal of Environmental Economics and Management" 1978, No. 5, s. 81-102.
4. Meyer D., Leisch F., Hornik K.: *Benchmarking Support Vector Machines*. Report No. 78, Vienna University of Economics and Business Administration, 2002, http://www.wuwien.ac.at/am/Download/ report78.pdf.
5. Trzęsiok J.: *Metoda rzutowania w budowie modelu regresyjnego*. W: *Postępy ekonometrii*. Red. A.S. Barczak. Wydawnictwo Akademii Ekonomicznej, Katowice 2004, s. 121-130.
6. Trzęsiok J.: *Analiza wybranych własności metody MART*. W: *Taksonomia 13. Klasyfikacja i analiza danych*. Red. K. Jajuga, M. Walesiak. Prace Naukowe Akademii Ekonomicznej, Wrocław 2006, No. 1126, s. 510-518.
7. Trzęsiok J.: *Ocena zasadności łączenia wybranych nieparametrycznych modeli regresji*. W: *Taksonomia 15. Klasyfikacja i analiza danych*. Red. K. Jajuga, M. Walesiak. Prace Naukowe Uniwersytetu Ekonomicznego, Wrocław 2008, No. 1207, s. 346-353.

DETERMINING THE INFLUENCE OF PREDICTOR VARIABLES ON THE RESPONSE VARIABLE IN THE PPR MODELS

Summary

Projection Pursuit Regression (PPR) was introduced by J. Friedman and W. Stuetzle in 1981. It is one of the nonparametric regression methods. The benchmark studies show very often the superiority of PPR models over other nonparametric or classical regression models in terms of the test error. PPR produces a "black-box" prediction machine and it suffers from the lack of interpretation. Thus it seems to be an important issue to find the method for evaluating the influence of the predictor variables on the response.

We present the procedure that might be used to examine the strength of the influence of every predictor variable on the response variable in Projection Pursuit Regression models.