

mgr Wojciech Rejchel
 Uniwersytet Mikołaja Kopernika
 Wydział Matematyki i Informatyki

Estymatory regresji rangowej i ich asymptotyka

Regresja rangowa obejmuje zadania mające bardzo szerokie zastosowania w praktyce. Dla przykładu wymieńmy chociażby: wyszukiwanie informacji (IR), bankowość, medycynę czy kontrolę jakości.

W pracy rozważamy populację, której obiekty wyposażone są w relację (liniowego) porządku. Zakładamy ponadto, że dwa wybrane elementy z tej populacji opisane są przez parę niezależnych wektorów $Z = (X, Y)$ i $Z' = (X', Y')$ o tym samym rozkładzie i wartościach w $\mathbb{R}^d \times \mathbb{R}$. Porządek zdefiniowany jest poprzez etykiety Y i Y' w sposób następujący: $Z \preceq Z'$, gdy $Y \leq Y'$. Zadanie regresji rangowej polega na odgadnięciu porządku pomiędzy obiektami Z i Z' , znając tylko obserwacje X i X' , w możliwie najlepszy sposób, tzn. poszukujemy reguły rangującej ϕ , która minimalizuje

$$\mathbf{E} \mathbb{I}(Y > Y') \mathbb{I}(\phi(X) \leq \phi(X')). \quad (1)$$

Ograniczamy się do reguł liniowych $\phi(x) = \theta^T x$, $\theta \in \mathbf{R}^d$. Han [2] pokazał, że wektor losowy minimalizujący empiryczny odpowiednik (1) jest mocno zgodny i asymptotycznie normalny. Jednak funkcja $\mathbb{I}(y > y') \mathbb{I}(\theta^T x \leq \theta^T x')$ jest nieciągła względem θ , więc wyznaczenie tego estymatora jest obliczeniowo trudne i nieefektywne. Proponujemy, analogicznie do maszyn wektorów podpierających w statystycznej teorii klasyfikacji [3], poszukiwanie minimum następującej, wypukłej względem θ , funkcji

$$\frac{1}{n(n-1)} \sum_{i \neq j} \mathbb{I}(Y_i > Y_j) \psi(\theta^T X_i - \theta^T X_j), \quad (2)$$

gdzie $\psi(x) = (1 - x)_+$. Znane są algorytmy efektywnie wyliczające argument minimalizujący funkcję (2) [1]. W pracy pokazujemy, że nowy estymator jest mocno zgodny i asymptotycznie normalny. Ponadto przedstawimy wyniki praktycznych zastosowań estymatora.

Bibliografia

- [1] L. Bobrowski, W. Niemirowicz *A method of synthesis of linear discriminant function in the case of nonseparability*, Pattern Recognition 17 (1984), 205–210.
- [2] A. K. Han, *Non-parametric analysis of a generalized regression model*, Journal of Econometrics 35 (1987), 303–316.
- [3] V. N. Vapnik, *Statistical Learning Theory*, J. Wiley, New York 1998.

Słowa kluczowe: liniowa reguła rangowa, maszyny wektorów podpierających, minimalizacja ryzyka wypukłego, U-statystyka.