

# Zgodność selekcji Uogólnionego Kryterium Informacyjnego dla wysokowymiarowej regresji logistycznej

Hubert Szymanowski

Instytut Podstaw Informatyki PAN

## Streszczenie

Rozważamy metodę selekcji zmiennych w wysokowymiarowych problemach regresyjnych, polegającą na wyborze podzbioru zmiennych minimalizującego funkcję celu Uogólnionego Kryterium Informacyjnego na rodzinie podzbiorów o z góry ograniczonej liczności. Przedstawiamy ogólne warunki przy których metoda ta z prawdopodobieństwem dążącym do 1 pozwala wybrać model prawdziwy, ograniczając rozważania do standardowej regresji logistycznej. Ponieważ wspomniana metoda przeważnie nie jest możliwa do zastosowania w praktyce ze względu za zbyt dużą liczbę porównywanych modeli, prezentujemy jej praktyczne modyfikacje. Przedstawiamy ich działanie na sztucznie generowanych zbiorach danych. Wstępne wyniki, przy znacznie bardziej restrykcyjnych założeniach niż omówione w referacie (spełnienie Rzadkiego Warunku Riesza) zostały opublikowane w Mielniczuk i Szymanowski (2015).

## Literatura

- Mielniczuk J., Szymanowski H., *Selection consistency of Generalized Information Criterion for sparse logistic regression model*, W Stochastic Models, Statistics and Their Applications, tom 122 Springer Proceedings in Mathematics & Statistics, strony 111–119, Wrocław, 2015.
- Chen J., Chen Z., *Extended BIC for small-n-large-p sparse GLM*, Statist Sinica, 22:555-574