

mgr Magdalena Malina

Instytut Matematyczny Uniwersytetu Wrocławskiego

dr Małgorzata Bogdan

Instytut Matematyki i Informatyki Politechniki Wrocławskiej

## Zastosowanie logicznej regresji do lokalizacji genów wpływających na cechy ilościowe

Regresja logiczna wprowadzona w [2] przez autorów Ruczinski, Kooperberg, LeBlanc jest uogólnioną metodą regresji użyteczną w sytuacji, gdy większość zmiennych ma charakter binarny. Często w takim przypadku interakcje pomiędzy wieloma predyktorami mają znaczący wpływ na zmienną wynikową. W kontekście genetycznym zagadnienie tego typu pojawia się w przypadku danych dotyczących pojedynczych polimorfizmów (*Single Nucleotide Polimorphisms*) i wówczas celem staje się identyfikacja takich polimorfizmów lub ich interakcji odpowiedzialnych za rozwój konkretnej choroby.

W regresji logicznej, mając zbiór predyktorów binarnych, staramy się stworzyć nowe lepsze predyktory, będące kombinacjami wyjściowych zmiennych binarnych. Dopasowujemy model regresji postaci

$$g(E(Y)) = b_0 + b_1L_1 + b_2L_2 + \dots + b_nL_n,$$

gdzie  $L_j$ ,  $j = 1, 2, \dots, n$ , są wyrażeniami logicznymi zależnymi od stanu objaśniających zmiennych binarnych. Wyboru wyrażeń  $L_j$  i estymacji parametrów  $b_j$  dokonujemy za pomocą algorytmu *simulated annealing*.

Przedstawimy przykłady i wyniki zastosowania logicznej regresji do objaśniania zmiennej o charakterze ilościowym.

### Literatura

- [1] H. Schwender, *Statistical analysis of genotype and gene expression data*, Ph.D. thesis, URL: [hdl.handle.net/2003/23306](http://hdl.handle.net/2003/23306).
- [2] I. Ruczinski, C. Kooperberg, M. LeBlanc, *Logic regression*, *J. Comput. Graphical Statist.* 12:3 (2003), 474–511, URL: [biostat.jhsph.edu/iruczins/publications/publications.html](http://biostat.jhsph.edu/iruczins/publications/publications.html).
- [3] C. Kooperberg, I. Ruczinski, *Identifying Interacting SNPs Using Monte Carlo Logic Regression*, *Genetic Epidemiology* 28 (2005), 157–170.
- [4] A. Fritsch, K. Ickstadt, *Comparing Logic Regression Based Methods for Identifying SNP Interactions*, Springer, Berlin/Heidelberg, Lecture Notes in Computer Science 4414, 2007, 90–103.