

O złej specyfikacji dla modelu logistycznego

Mariusz Kubkowski¹ i Jan Mielniczuk²

¹ Politechnika Warszawska

² Instytut Podstaw Informatyki PAN

Streszczenie

Rozpatrujemy model regresji z odpowiedzią binarną spełniający $P(y = 1|x) = q(\beta^T x)$, gdzie $q : R \rightarrow (0, 1)$, a x jest pewnym wektorem losowym lub deterministycznym. Niech X_n będzie macierzą eksperymentu, zaś Y_n wektorem odpowiedzi dla n obserwacji dla tego modelu, do których dopasowujemy model regresji logistycznej z funkcją odpowiedzi p . Wiadomo, że oczekiwana funkcja wiarygodności jest w przypadku losowym maksymalizowana przez wektor β^* taki, że $E(x(q(\beta^T x) - p(\beta^{*T} x))) = 0$, a w przypadku deterministycznym przez wektor $\beta_n^*(X_n)$ taki, że $X_n^T(q(X_n\beta) - p(X_n\beta_n^*(X_n))) = 0$, gdzie $q(X_n\beta) = (q(x_{1\cdot}^T\beta), \dots, q(x_{n\cdot}^T\beta))^T$ i x_i oznacza i -ty wiersz macierzy X_n . W referacie omówione zostaną wyniki dotyczące związków między wektorami β , β^* i $\beta_n^*(X_n)$ oraz ich nośnikami oraz pewne przykłady złej specyfikacji modelu binarnego.

Literatura

- Mielniczuk, J., Teisseyre, P. (2016). What Do We Choose When We Err? Model Selection and Testing for Misspecified Logistic Regression Revisited. *Challenges in Computational Statistics and Data Mining* (s. 271-296). Springer International Publishing.
- Lv, J., Liu, J. S. (2014). Model selection principles in misspecified models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1), 141-167.
- Gail, M. H., Wieand, S., Piantadosi, S. (1984). Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*, 71(3), 431-444.