

prof. Leon Bobrowski

Wydział Informatyki Politechniki Białostockiej

Instytut Biocybernetyki i Inżynierii Biomedycznej PAN, Warszawa

Regresja przedziałowa oparta na funkcjach kryterialnych typu CPL

Określenie *regresja przedziałowa* (ang. *interval regression*) odnosi się do sytuacji, gdy model regresyjny budujemy na podstawie takiego zbioru danych, w którym wartości zmiennej zależnej nie są określone dokładnie lecz za pomocą przedziałów [1], [2]. Zbiór danych regresji przedziałowej ma poniższą strukturę:

$$C_m = \{\mathbf{x}_j[n], [l_j, u_j]\}, \quad \text{gdzie } j = 1, \dots, m, \quad (1)$$

gdzie $\mathbf{x}_j[n] = [x_{j1}, \dots, x_{jn}]^T$ jest n -wymiarowym wektorem cech x_i ($x_j[n] \in R^n$), natomiast l_j i u_j są liczbami rzeczywistymi spełniającymi relację $l_j \leq u_j$ (lub $l_j < u_j$).

Bierzemy pod uwagę odwzorowania liniowe $y = \mathbf{w}[n]^T \mathbf{x}[n]$ wektorów cech $\mathbf{x}_j[n]$ na liczby y_j :

$$(\forall j \in \{1, \dots, m\}) \quad y_j = \mathbf{w}[n]^T \mathbf{x}_j[n], \quad (2)$$

gdzie $\mathbf{w}[n] = [w_1, \dots, w_n]^T$ jest wektorem parametrów (*wag*) w_i ($\mathbf{w}[n] \in R^n$).

Odwzorowanie $y = \mathbf{w}[n]^T \mathbf{x}[n]$ tworzy model regresji przedziałowej wtedy, gdy możliwie dobrze spełnia układ nierówności

$$(\forall j \in \{1, \dots, m\}) \quad l_j \leq \mathbf{w}[n]^T \mathbf{x}_j[n] \leq u_j \quad (3)$$

Modele regresji przedziałowej pojawiają się między innymi w kontekście analizy przeżycia (ang. *survival analysis*).

Projektowanie modeli regresji przedziałowej można sprowadzić do zagadnienia liniowej separowalności dwu zbiorów oraz zastosować wypukłe i odcinkowo-liniowe (*CPL*) funkcje kryterialne do wyznaczenia optymalnego wektora parametrów $\mathbf{w}[n]$ [3].

Literatura

- [1] G. Li, C. Zhang, *Linear regression with interval censored data*, The Annals of Statistics 26:4 (1998), 1306–1327.
- [2] G. Gomez, A. Espinal, S. Lagakos, *Inference for a linear regression model with an interval-censored covariate*, Statistics in Medicine 22 (2003), 409–425.
- [3] L. Bobrowski, *Eksploracja danych oparta na wypukłych i odcinkowo-liniowych funkcjach kryterialnych*, Wydawnictwa Politechniki Białostockiej, Białystok 2005.