

РЕГРЕСІЙНИЙ ПРОГНОЗ ВІДСОТКУ ЖИРУ В ОРГАНІЗМІ ЛЮДИНИ МЕТОДАМИ МАШИННОГО НАВЧАННЯ

Яремчук О.М., Кочетов К.В.

Чорноморський національний університет ім. Петра Могили, м. Миколаїв

У роботі досліджується група з шести алгоритмів регресії, представлених у відомій платформі для машинного навчання WEKA [1] на прикладі набору даних щодо вмісту жиру в тканинах людського організму. Набір даних [2] містить 252 екземпляри з 15 атрибутами, включно з цільовим, для кожної з них. Атрибути мають антропометричний характер (вік, вага, зріст, обсяг грудної клітини тощо). Цільовою залежною змінною є відсоток жиру у тканинах тіла.

Засобами модуля Experimenter, включеного у платформу WEKA, розглядалися такі алгоритми як-от: Гаусова регресія (Gaussian Processes), Лінійна регресія (Linear Regression), k-найближчих сусідів (Ibk, k=3), Дерево рішень (REPTree), Метод опорних векторів (SMOreg), Багатошаровий персептрон (Multilayer Perceptron). Дизайн експерименту передбачав перехресну перевірку моделі з кількістю згорток, яка дорівнювала 10, тобто кожна комбінація алгоритм разом набором даних виконується 10 разів (ітераційний контроль).

Оцінка валідності регресійних моделей проводилася за відомим стандартними параметрами: 1) середня абсолютна помилка регресії (mae), 2) корінь квадратний із середньої квадратичної помилки (rmse), 3) відносна абсолютна помилка (%), 4) відносна rmse (%), 5) коефіцієнт кореляції (R^2).

Якщо пріоритетною оцінкою вважати коефіцієнт кореляції (R^2 , або коефіцієнт детермінації), який показує кореляцію між відомими та прогнозованими відсотками жиру в організмі, то найбільш ефективним є алгоритм Лінійної регресії (Linear Regression). Показник кореляції алгоритма-переможця 0.99 ± 0.02 .

Якщо ж орієнтуватися на відносну абсолютну похибку (Relative absolute error), то найменшу похибку (3.9 %) показує алгоритм «Метод опорних векторів» (SMOReg), у той час як відносна похибка Лінійної регресії (Linear Regression) дещо вища (7.6 %). Таким чином, оптимальність алгоритму для регресії суттєво залежить від тієї оцінки валідності, який дослідник надає пріоритет.

Література:

1. J. Brownlee. "How To Use Regression Machine Learning Algorithms in Weka." August 22, 2019, URL: <https://machinelearningmastery.com/use-regression-machine-learning-algorithms-weka/>
2. Weka/Weka datasets/bodyfat.arff., July 4, 2017, URL: <https://github.com/tertiarycourses/Weka/blob/master/Weka%20datasets/bodyfat.arff>