

Прикладні аспекти використання logit-регресії

Національний університет «Острозька академія»

Анотація:

Досліджено особливості застосування logit-регресії у банківській, медичній та екологічній сферах. Запропоновано використання реальних кейсах при вивченні дисциплін, пов'язаних з моделюванням та аналізом даних.

Ключові слова: logit-регресія, математичне моделювання, кластеризація.

Abstract

The peculiarities of logit-regression application in banking, medical and ecological spheres are investigated and learned. The use of real cases in the study of disciplines related to modeling and data analysis is proposed.

Keywords: logit-regression, mathematical modeling, clustering.

У процесі виробничої та невиробничої діяльності виникає необхідність кластеризувати об'єкти до однієї з відомих множин. Це відбувається за допомогою математичних методів, одним з яких є logit-регресія. Рівняння logit-регресії має вигляд [1]:

$$E(Y_i|X_i) = p_i = \frac{e^{(\beta_0 + \beta_1 x_i)}}{1 + e^{(\beta_0 + \beta_1 x_i)}}$$

Досліджувані параметри знаходяться у степені експоненти, що дозволяє отримувати результати у межах [0;1]. Таким чином можна отримати не лише розподіл на кластери, а і ймовірність потрапляння конкретного об'єкта з відомими параметрами до того чи іншого кластеру.

Застосування logit-регресії широко використовується у банківських системах. При вивченні даного методу студентам (зокрема економічних спеціальностей) доцільно наводити приклад такого використання. Для початку, можна розглянути бінарну класифікацію на основі даних одного з німецьких банків [2]. У даний вибірці міститься 21 атрибут (як кількісні, та і якісні), усі позичальники діляться на «хороших» і «поганих» (1 та 0 відповідно). Методом logit-регресії легко встановити ймовірність того, що позичальник буде «хорошим».

У дослідженні В.Я. Данилова, О.Л. Жирова, П.І. Бідюка [3] зазначається, що перевагою logit-регресії є те, що вона може поділяти клієнтів як на дві групи (0 – поганий, 1 – хороший), так і на кілька груп (1, 2, 3, 4 групи ризику). У цьому ж дослідженні підтверджується висока ефективність методу у кредитному скорингу.

У медичній сфері бінарна класифікація полягає у віднесенні пацієнта до кластеру «хворий» чи «здоровий» на основі показників його стану здоров'я. Для цього пропонується розглянути кейс на основі бази даних Клівленда щодо хвороб серця [4]. У дослідженні було зібрано 76 атрибутів проте для моделювання було достатньо 14, серед яких кількісні (*age, cigarettes per day, number of years as a smoker, тощо*), якісні (*chest pain location, chest pain type, тощо*) та бінарні *sex, fasting blood sugar > 120 mg/dl, amily history of coronary artery disease, та ін..*). За використання логістичної регресії високим є показник істинно позитивних значень, що підтверджує можливість застосування методу у цій сфері.

Прикладом використання logit-регресії у екології може бути кейс, досліджений Кеном Коглером щодо новонароджених черепашок. Він провів експеримент з черепашиними яйцями одного й того ж виду, які були транспортовані та розподілені на декілька груп, які вивчалися при різних температурах: від 27,2°C до 29,9°C [1]. За результатами дослідження було встановлено, що при підвищенні температури серед новонароджених переважають черепашки чоловічої статі, що у свою чергу впливає на довкілля.

Отже, застосування методу logit-регресії є актуальним та затребуваним у різних сферах. Особливістю використання методу є саме сфера використання, адже для різних потреб необхідно

коригувати методи дослідження. Проте для вивчення методу студентами рекомендовано використовувати кейси, які дають вищий рівень розуміння сутності проблеми та переваги методу.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Logistic Regression [Електронний ресурс] — Режим доступу: http://courses.ncssm.edu/math/Stat_Inst/PDFS/REG3_LOG.pdf
2. UCI Machine Learning Repository [Електронний ресурс] // UCI Machine Learning Repository. – Режим доступу: <https://archive-beta.ics.uci.edu/ml/datasets/statlog+german+credit+data>.
3. ДАНИЛОВ В. Я. Оцінювання кредитних ризиків методами інтелектуального аналізу даних / В. Я. Данилов, О. Л. Жиров, П. І. Бідюк // Системні дослідження та інформаційні технології. – 2017. – № 1. – С. 33–48.
4. UCI Machine Learning Repository [Електронний ресурс] // UCI Machine Learning Repository. – Режим доступу: <https://archive-beta.ics.uci.edu/ml/datasets/heart+disease>.

Артиух Ольга Миколаївна — викладач-стажист кафедри економіко-математичного моделювання та інформаційних технологій економічного факультету Національного університету «Острозька академія», Національний університет «Острозька академія», м. Острог, olha.artiukh@oa.edu.ua

Artiukh Olha M. — Lecturer-trainee of the Department of Economic-Mathematical Modeling and Information Technologies, National University of "Ostroh Academy", Ostroh