

УДК 004.89

В. Яцишин, канд. техн. наук, доцент; А. Демиденко; В. Яцишин

(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

ЛОГІСТИЧНА РЕГРЕСІЯ В ЗАДАЧАХ ВИЯВЛЕННЯ АНОМАЛІЙ ТРАФІКУ КОМП'ЮТЕРНИХ МЕРЕЖ

UDC 004.89

V. Yatsyshyn, PhD., Assoc. Prof., A. Demydenko; V. Yatsyshyn

LOGISTIC REGRESSION IN COMPUTER NETWORK TRAFFIC ANOMALIES DETECTION PROBLEMS

Логістична регресія відноситься до класу керованих алгоритмів машинного навчання, які використовуються для задач класифікації. Основна ціль такого алгоритму полягає в тому, щоб спрогнозувати ймовірність приналежності екземпляра даних до одного чи іншого класу.

Логістична регресія доволі ефективно застосовується при бінарній класифікації з використанням сигмоїдної функції, яка приймає вхідні дані як незалежні змінні та генерує значення ймовірності в діапазоні від 0 до 1.

Наприклад, існує два класи (у даному випадку аномальний і нормальний трафік): клас 0 і клас 1. Якщо значення логістичної функції для вхідних даних перевищує 0,5 (порогове значення), тоді воно належить до класу 1, інакше воно належить до класу 0.

Логістична регресія прогнозує вихід категоріальної залежної змінної. Отже, результат повинен мати категоріальне або дискретне значення. Це може бути «Так» або «Ні», «0» або «1», «Істинно» або «Хибно» тощо, але замість того, щоб давати точні значення 0 або 1, алгоритм формує ймовірності, які знаходяться між 0 і 1.

У логістичній регресії замість «підгонки» лінії регресії підбирають логістичну функцію S-подібної форми, яка прогнозує два максимальні значення (0 або 1).

Логістична функція є сигмоїдною функцією, тобто математичною функцією, яка використовується для відображення прогнозованих значень у ймовірності. Сигмоїд перетворює будь-яке реальне значення на інше в діапазоні від 0 до 1. Значення логістичної регресії повинно бути між 0 і 1 і не може виходити за межі цього інтервалу, тому воно утворює криву, подібну до форми «S».

Крива S-подібної форми називається сигмоподібною функцією або логістичною функцією. У логістичній регресії використовується поняття порогового значення, яке визначає ймовірність 0 або 1. Наприклад, значення вище порогового значення прямує до 1, а значення нижче порогових значень прямують до 0.

На основі категорій логістичну регресію можна класифікувати на три типи:

- біноміальна або бінарна: у біноміальній логістичній регресії може бути лише два можливих типи залежних змінних, наприклад 0 або 1, «Пройшов» або «Не пройшов» тощо;
- мультиноміальна: у мультиноміальній логістичній регресії може бути 3 або більше можливих невпорядкованих типів залежної змінної, наприклад «кіт», «собаки» або «вівця»;
- порядкова: у порядковій логістичній регресії може бути 3 або більше можливих упорядкованих типів залежних змінних, таких як «низький», «середній» або «високий».

Модель логістичної регресії перетворює неперервні вихідні дані функції лінійної регресії у категоріальні вихідні значення за допомогою сигмоїдної функції, яка відображає будь-який дійсний набір вхідних незалежних змінних у значення від 0 до 1.