

Логистическая регрессия (Logistic Regression)

Синонимы: Логит-регрессия, Logit Regression

Разделы: [Алгоритмы](#)

В математической статистике логистическая регрессия (или логит-регрессия) является широко используемой статистической моделью, которая использует логистическую функцию для моделирования зависимости выходной переменной от набора входных в случае, когда первая является бинарной.

Это разновидность множественной регрессии, общее назначение которой состоит в анализе связи между несколькими независимыми переменными (называемыми также **регрессорами** или **предикторами**) и зависимой переменной. Регрессия в общем виде применяется, когда входные и выходная переменные непрерывные. А логистическая регрессия лучшим образом подходит, когда выходная переменная принимает только два значения.

Важность логистической регрессии обусловлена тем, что многие задачи анализа данных могут быть решены с помощью бинарной классификации или сведены к ней.

Например, с помощью логистической регрессии можно оценивать вероятность наступления (или не наступления) некоторого события: пациент болен (здоров), заемщик вернул кредит (допустил просрочку) и т.д. Благодаря этому логистическую регрессию можно рассматривать как мощный инструмент поддержки принятия решений.

Как известно, все регрессионные модели могут быть записаны в виде формулы:

$$y = F(x_1, x_2, \dots, x_n).$$

Например, если рассматривается исход по займу, задается переменная y со значениями 1 и 0, где 1 означает, что соответствующий заемщик расплатился по кредиту, а 0 — что имел место дефолт.

Однако здесь возникает проблема: множественная регрессия не «знает», что переменная отклика бинарная по своей природе. Это неизбежно приведет к модели с предсказываемыми значениями большими 1 и меньшими 0. Но такие значения вообще не допустимы для первоначальной задачи. Таким образом, множественная регрессия просто игнорирует ограничения на диапазон значений для y .

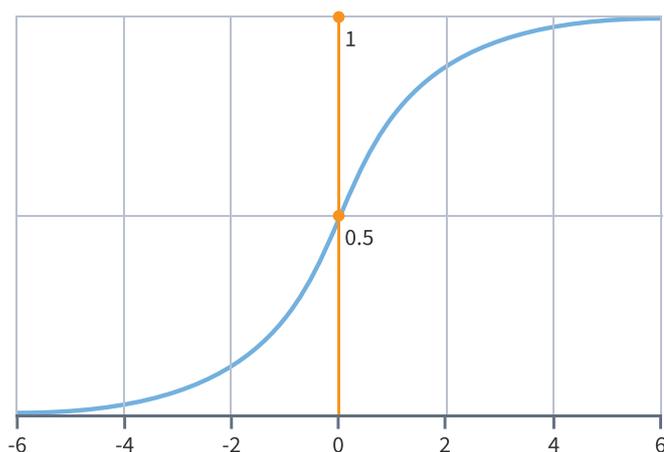
Для решения проблемы задача регрессии может быть сформулирована иначе: вместо предсказания бинарной переменной мы предсказываем непрерывную переменную со значениями на отрезке $[0,1]$ при любых значениях независимых переменных. Это

достигается применением следующего регрессионного уравнения (логит-преобразование):

$$p = \frac{1}{1+e^{-y}},$$

где p — вероятность того, что произойдет интересующее событие; e — основание натуральных логарифмов 2,71...; y — стандартное уравнение регрессии.

Зависимость, связывающая вероятность события и величину y , показана на следующем графике:



Преобразование вида:

$$P' = \log_e \left(\frac{P}{1-P} \right)$$

называют **логистическим**, или **логит-преобразованием**.

Существует несколько способов нахождения коэффициентов логистической регрессии. На практике часто используют метод максимального правдоподобия. Он применяется в статистике для получения оценок параметров генеральной совокупности по выборочным данным.

В Logipom существует специализированный обработчик логистическая регрессия, с помощью которого можно оценивать вероятность того, что событие наступит для конкретного объекта испытания (больной/здоровый, возврат кредита/дефолт и т.д.).

Логистическая регрессия является традиционным статистическим инструментом для расчета коэффициентов (баллов) скоринговой карты на основе накопленной кредитной истории. Подробнее в статье «Логистическая регрессия и ROC-анализ — математический аппарат».

О прикладном применении логистической регрессии в двух областях — диагностика заболеваний и оценка кредитоспособности физических лиц узнайте в статье «Применение логистической регрессии в медицине и скоринге».