# Машинное обучение
# Логистическая регрессия

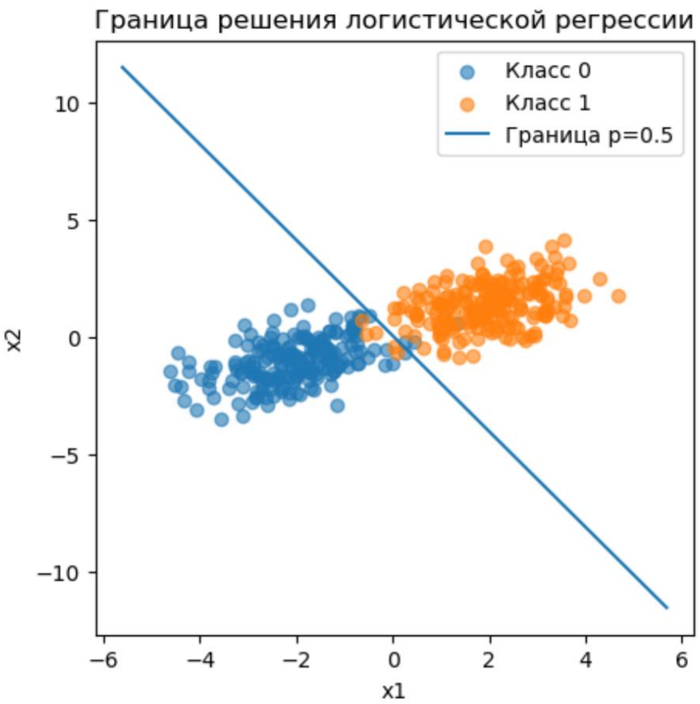
Житницкая Ирина гр. 20401
Черняховский Лев гр. 20202
Штейнберг Георгий гр. 20202

10.11.25

POLYTECH

# Содержание

- Что такое логистическая регрессия
  - Основная идея
  - Математическая модель
  - Обучение модели
  - Интерпретация результатов
- Пример применения
- Преимущества и недостатки
- Вариации и расширения






POLYTECH

# Определение

- Логистическая регрессия – это статистическая модель, используемая для прогнозирования вероятности возникновения некоторого события путём его сравнения с логистической кривой.

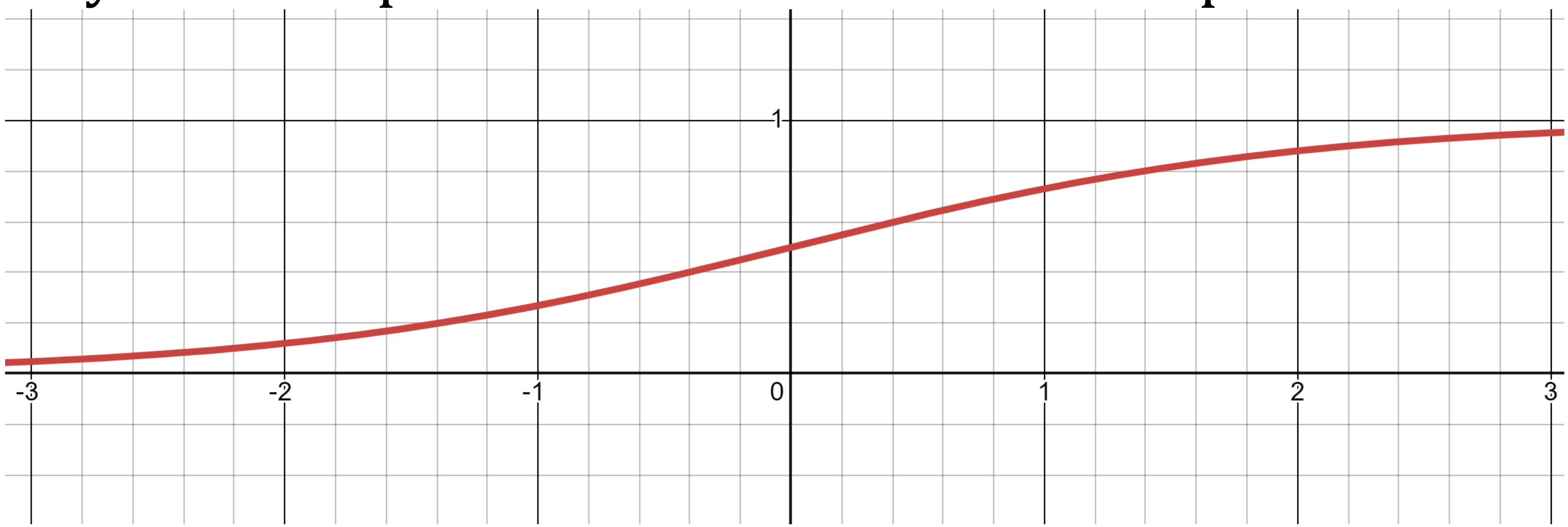



График логистической кривой: $y = \frac{1}{1+e^{-x}}$

# Математическая модель

- Вероятность того, что объект принадлежит классу y = 1:
  - $\mathcal{P}(y = 1|x) = \sigma(z) = \frac{1}{1+e^{-z}}$,
    где $z = a_0 + a_1x_1 + a_2x_2 + \cdots + a_nx_n$
    $x_1, x_2 \dots x_n$– независимые переменные
    $a_1, a_2 \dots a_n$– коэффициенты регрессии
    $a_0$– смещение
- $\mathcal{P}(y_i|x_i) = \sigma(z_i)^{y_i}(1 - \sigma(z_i))^{(1-y_i)}$
  - Если $y_i = 1$, то вероятность $\sigma(z_i)$
  - Если $y_i = 0$, то вероятность $1 - \sigma(z_i)$

# Математическая модель

- Логарифмическая функция потерь

Мы хотим подобрать веса $a_0, a_1 \dots a_n$, чтобы **правдоподобие данных** было максимальным

$$L(w) = \prod_{i=1}^{N} \mathcal{P}(y_i | x_i)$$

Для удобства используем логарифм:

$$\log L(w) = \sum_{i=1}^{N} (\, y_i \log(\sigma(z_i)) + (1 - y_i) \log(1 - \sigma(z_i))$$

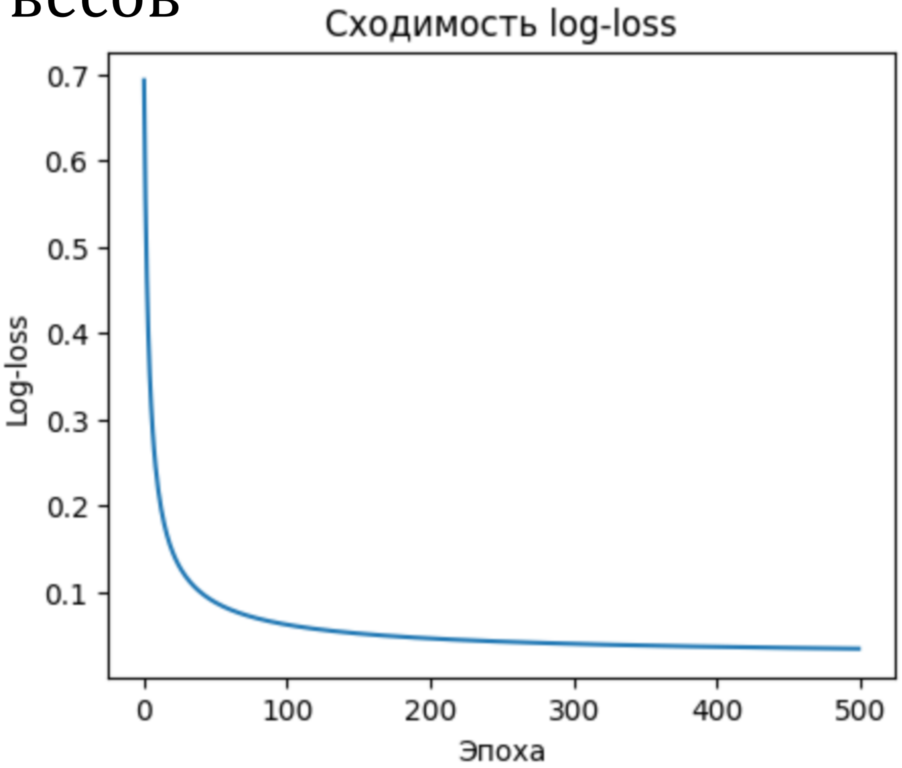
$\mathcal{L}(w) = -\log L(w)$ – это log loss



POLYTECH

# Процесс обучения модели

- Начинаем с произвольных весов $a_0, a_1 \dots a_n$
- Используем градиентный спуск для минимизации log loss и корректировки весов
- Повторяем, пока log loss не станет минимальной



# Интерпретация результатов

- Чтобы принять окончательное решение о принадлежности объекта к классу вводится пороговое значение (обычно 0.5)
$$y = \begin{cases} 1, & \text{если } \mathcal{P} \geq 0.5 \\ 0, & \text{если } \mathcal{P} \leq 0.5 \end{cases}$$
- Каждый вес $a_0, a_1 \dots a_n$ показывает влияние признака $x_1, x_2 \dots x_n$ на вероятность

# Пример

- Дано: N=3 наблюдения. X=[1, 2, 3] Y=[0, 0, 1]
- Модель: $\mathcal{P}(y=1|x) = \frac{1}{1+e^{-(a_0+a_1 x)}}$,

  log loss:

  $$\mathcal{L}(a_0, a_1) = -\sum_{i=1}^{N} (y_i \log(\mathcal{P}_i) + (1-y_i)\log(1-\mathcal{P}_i))$$

  Градиенты:

  $$\frac{\partial \mathcal{L}}{\partial a_0} = \sum_{i=1}^{N} (\mathcal{P}_i - y_i) \qquad \frac{\partial \mathcal{L}}{\partial a_1} = \sum_{i=1}^{N} (\mathcal{P}_i - y_i) x_i$$

  Инициализация:

  $\alpha = 0.1, a_0 = 0, a_1 = 0$

# Пример

- 1 итерация:
  - $\mathcal{P} = \frac{1}{1+e^0} = 0.5$
  - $\mathcal{L} = -\,(0 \cdot \log(0.5) + 0 \cdot \log(0.5) + 1 \cdot \log(0.5) = 0.693147$
  - $\frac{\partial \mathcal{L}}{\partial a_0} = 0.1667\,;\ \frac{\partial \mathcal{L}}{\partial a_1} = 0$
  - Новые веса: $a_0 = a_0 - \alpha \frac{\partial \mathcal{L}}{\partial a_0} = -0.01667\,;\ a_1 = a_1 - \alpha \frac{\partial \mathcal{L}}{\partial a_1} = 0$
- 2 итерация:
  - $\mathcal{P} = \frac{1}{1+e^{-0.1667}} = 0.4958$
  - $\mathcal{L} = -\,(2 \cdot 0 \cdot \log(0.4958) + 1 \cdot \log(0.5042)) = 0.690404$
  - Новые веса: $a_0 = a_0 - \alpha \frac{\partial \mathcal{L}}{\partial a_0} = -0.03291\,;$

    $a_1 = a_1 - \alpha \frac{\partial \mathcal{L}}{\partial a_1} = 0.0008$

# Пример

- После 500 итераций:
  - $\mathcal{P}_1 = 0.086$ ; $\mathcal{P}_2 = 0.334$ ; $\mathcal{P}_3 = 0.728$
  - $\mathcal{L} = 0.271315$
  - $a_0 = -4.0430$ ; $a_1 = 1.6764$
  - $\mathcal{P}_1 < 0.5 \ \rightarrow x_1$ принадлежит классу 0;
    $\mathcal{P}_2 < 0.5 \ \rightarrow x_2$ принадлежит классу 0;
    $\mathcal{P}_3 > 0.5 \ \rightarrow x_3$ принадлежит классу 1.
  - Модель правильно классифицирует все три наблюдения

# Преимущества и недостатки

- **Преимущества**:
- Простая и понятная интерпретация
- Быстро обучается и не требует больших вычислений
- Предсказывает вероятность принадлежности
- Работает даже при небольших наборах данных
- **Недостатки:**
- Плохо справляется с нелинейными зависимостями
- Неустойчива к зависимым признакам

# Вариации и расширения

- **Мультиномиальная** — применяется, когда целевая переменная имеет три или более категорий.
- **Ординарная** — используется, когда целевая переменная имеет три или более порядковых категорий.
- Регуляризация (L1, L2)
- Часто служит финальным классификатором

# Источники

- Logistic Regression in Machine Learning
- Логистическая регрессия: что это простыми словами
- Логистическая, порядковая и мультиномиальная регрессия
- Исследование ординальных данных: Примеры и применение