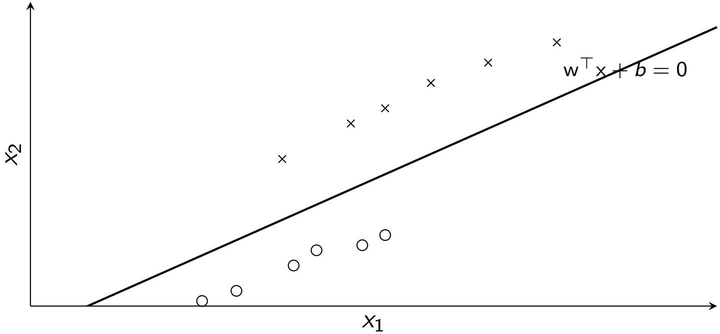
Логистическая регрессия

Лазар В. И., Козлова Е. Р.

24 сентября 2025 г.

Бинарная классификация: постановка

Дано: точки (x_i, y_i), где $y_i \in \{0,1\}$. Предсказываем вероятность $p(y=1 \mid \mathrm{x}) \in [0,1]$, решение по порогу (обычно 0,5).



Почему не подойдёт обычная линейная регрессия?

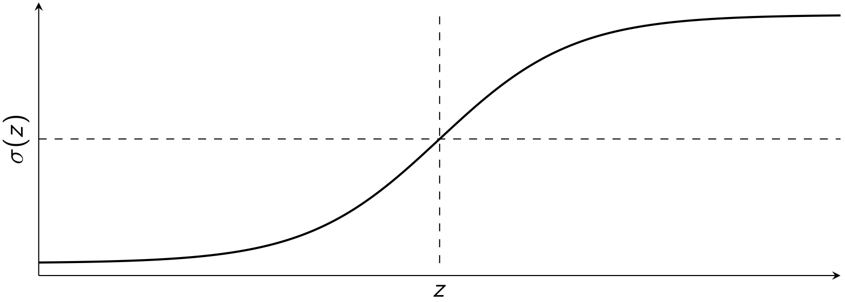
- Линейная регрессия выдаёт любые числа ($\mathbb{R}$), а нам нужна вероятность в $[0, 1]$.
- Квадратичная ошибка не «понимает» уверенность: за очень уверенную ошибку штраф почти как за небольшую.
- Хотим модель, у которой $\hat{p} = \Pr(y = 1 \mid x)$, и потери, сильно наказывающие «уверенно неправые» ответы.

Сигмоида: как из $\mathbb{R}$ получить $[0, 1]$

Определение логистической функции (сигмоиды):

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \qquad z = \mathrm{w}^\top \mathrm{x} + b.$$

Тогда $\hat{p} = \sigma(z) \in [0, 1]$ — это и есть оценка вероятности класса «1».



Вероятностная модель: Бернулли и правдоподобие

Для одного примера:

$$p(y \mid \mathrm{x}) = \sigma(z)^{y} \left(1 - \sigma(z)\right)^{(1-y)}, \quad y \in \{0, 1\}.$$

Для всей выборки (независимость примеров):

$$\mathcal{L}(\mathrm{w}, b) = \prod_{i=1}^{n} \sigma(z_i)^{y_i} \left(1 - \sigma(z_i)\right)^{(1-y_i)}, \quad z_i = \mathrm{w}^\top \mathrm{x}_i + b.$$

Учимся максимизируя правдоподобие $\mathcal{L}$ или (эквивалентно) лог-правдоподобие.

Функция потерь логистической регрессии: шаг-за-шагом

1) Логарифмируем произведение (сумма логов):

$$\log \mathcal{L} = \sum_{i=1}^{n} \Big[y_i \, \log \sigma(z_i) + (1 - y_i) \, \log\big(1 - \sigma(z_i)\big) \Big].$$

2) Меняем знак (минимизация вместо максимизации):

$$J(\mathrm{w}, b) = -\frac{1}{n} \sum_{i=1}^{n} \Big[y_i \, \log \sigma(z_i) + (1 - y_i) \, \log\big(1 - \sigma(z_i)\big) \Big].$$

Это **кросс-энтропийная** потеря (log loss). Она *сильно* штрафует уверенные ошибки.

Эквивалентные формы log loss

Через $y \in \{0, 1\}$:

$$\ell(y, z) = -\,y\ \log \sigma(z) - (1 - y)\ \log\bigl(1 - \sigma(z)\bigr).$$

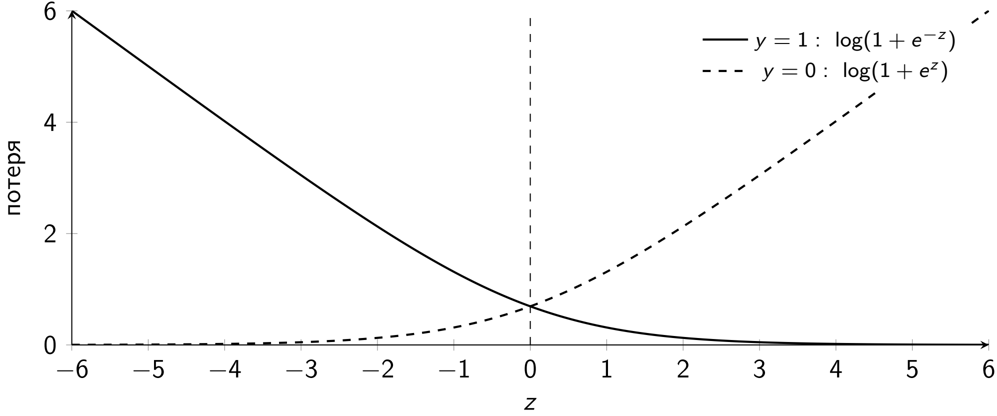
Через $y \in \{-1, +1\}$ и $z = \mathrm{w}^\top \mathrm{x} + b$:

$$\ell(y, z) = \log\bigl(1 + e^{-y\,z}\bigr).$$

Эти записи эквивалентны: выберите удобную для анализа/реализации.

Интуиция потерь: как растёт штраф

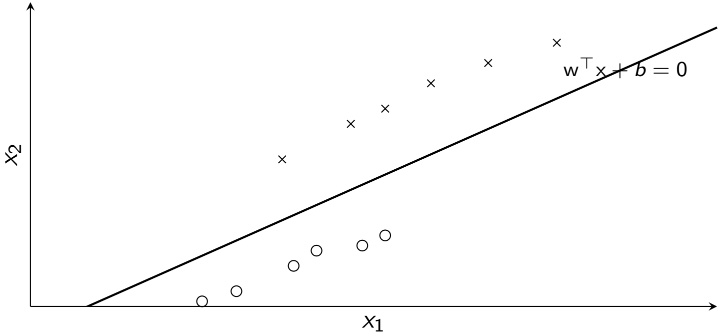
Для $y = 1$: $\ell = -\log \sigma(z) = \log(1 + e^{-z})$. Для $y = 0$: $\ell = -\log(1 - \sigma(z)) = \log(1 + e^{z})$.



Чем больше модель уверена и неправа (большой по модулю «не тот» знак z), тем больше штраф.

Граница решения и порог

Решаем по порогу $\hat{p} = 0{,}5$ (то есть $z = 0$) — это **линейная** граница $\mathrm{w}^\top \mathrm{x} + b = 0$.



Регуляризация (кратко)

Чтобы избежать переобучения и сделать веса стабильнее, добавляют штрафы:

$$J_{\text{ridge}} = J + \lambda \left\|\mathrm{w}\right\|_2^2, \qquad J_{\text{lasso}} = J + \lambda \left\|\mathrm{w}\right\|_1.$$

На практике **стандартизируйте** признаки и подбирайте λ по валидации.