«Машинное обучение»

Линейные методы (часть 2): логистическая регрессии

Александр Дьяконов

05 октября 2021 года

Линейные скоринговые модели в задаче бинарной классификации

Пусть $X = \mathbb{R}^n, Y = \{0, 1\}$

Как решать задачи классификации с помощью линейной модели: будем получать вероятность принадлежности к классу 1

$$a(x) \in [0, 1]$$

Любая линейная функция на $\mathbb{R}^n$ **будет получать значения в** $\mathbb{R}$**, поэтому нужна деформация (transfer function):**

$$\sigma : \mathbb{R} \to [0, 1]$$

Функции деформации

В логистической регрессии

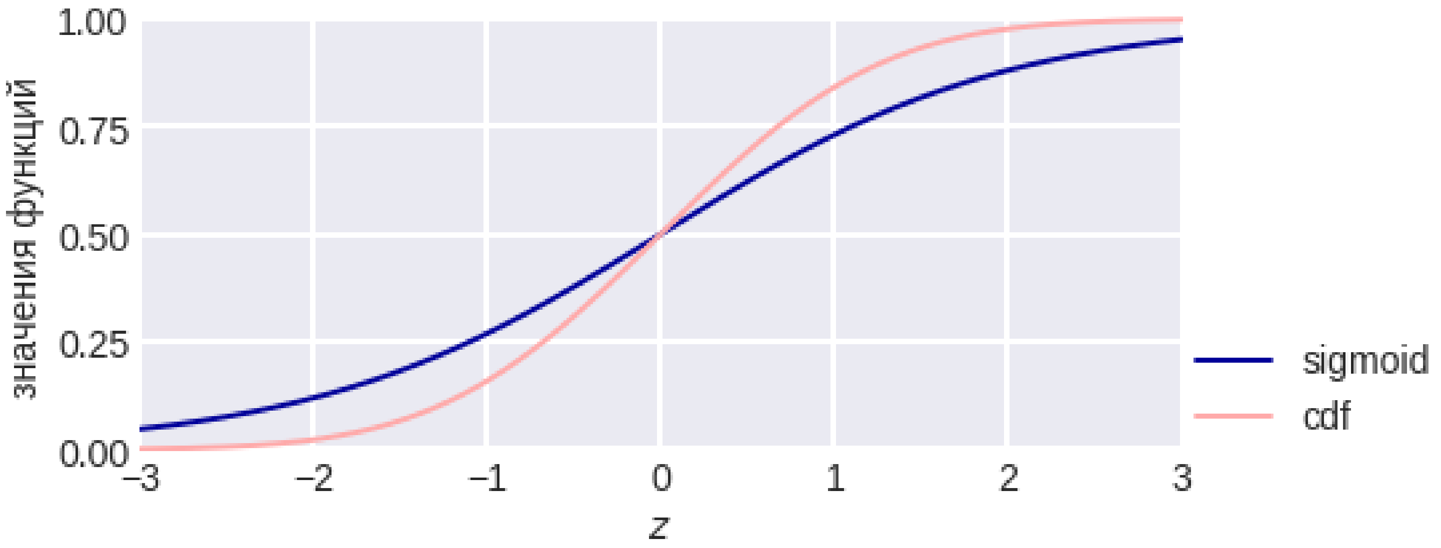
$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Логистическая функция (сигмоида)

В Probit-регрессии

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-t^2 / 2)\partial t$$

Normal Cumulative distribution function



Логистическая регрессия

$$p(x) \equiv P(Y=1\,|\,x) = \sigma(z) = \frac{1}{1+e^{-z}} \in (0,1),$$

$$z = w_0 + w_1 X_1 + \ldots + w_n X_n,$$

$$\log\left(\frac{p(x)}{1-p(x)}\right) = z$$

– монотонное преобразование, которое называют logit-transformation

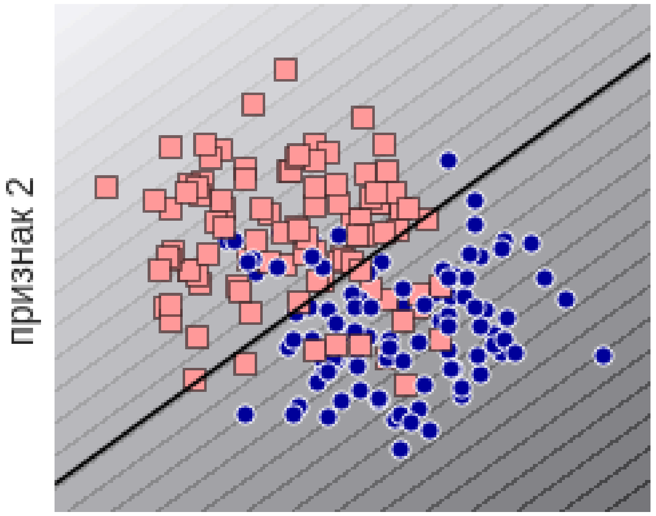
Решаем задачу классификации, но метод называется логистическая регрессия

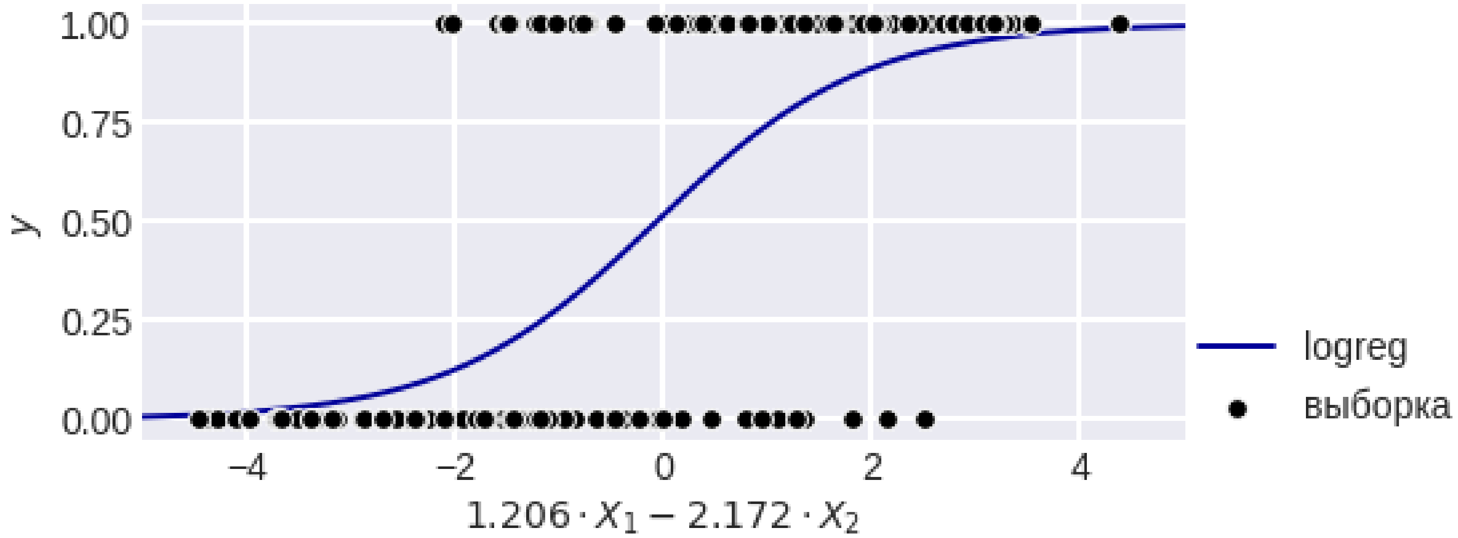
Кстати,

$$\underbrace{\sigma(z)}_{P(Y=1|x)} = \frac{1}{1+e^{-z}} = \frac{e^{+z}}{1+e^{+z}} = \frac{1+e^{+z}-1}{1+e^{+z}} = 1 - \frac{1}{1+e^{+z}} = 1 - \underbrace{\sigma(-z)}_{P(Y=0|x)}$$

Геометрический смысл логистической регрессии

```
from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
model.fit(X, y)
a = model.predict_proba(X_test)[:,1]
```

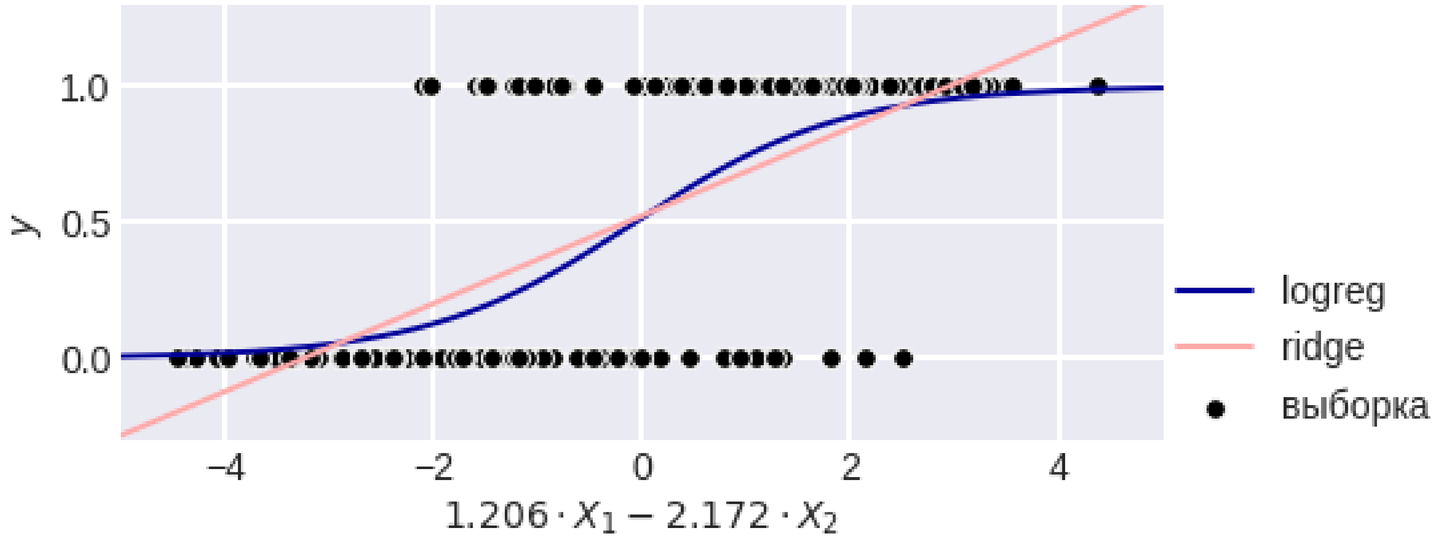




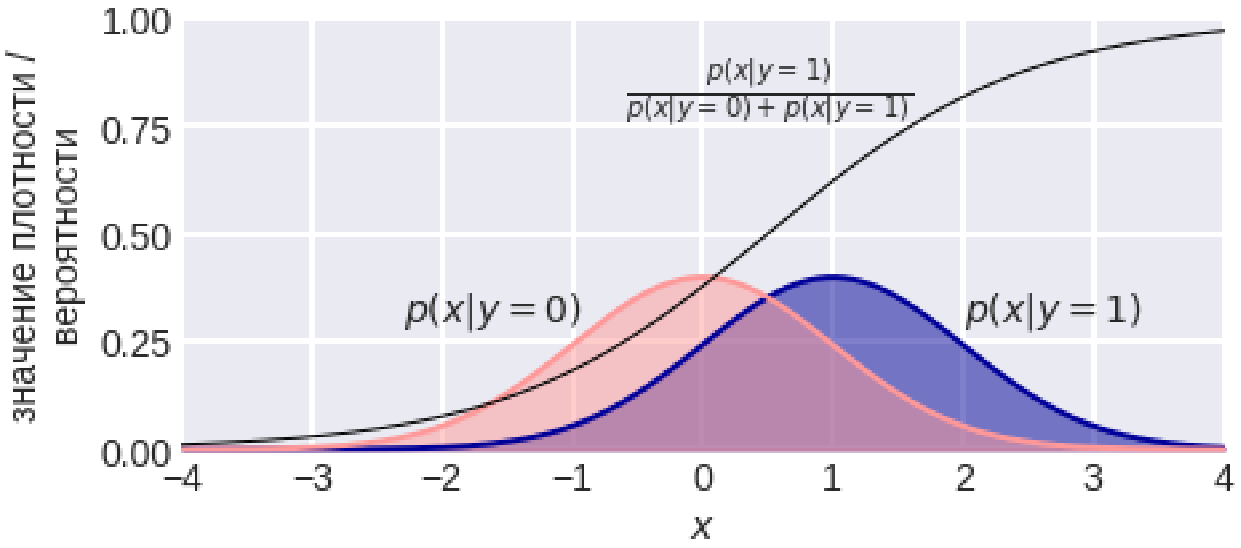
$z = w_0 + w_1 X_1 + \ldots + w_n X_n$ **– проекция на прямую (один признак)**

В однопризнаковом случае надо решить задачу классификации

Чем логистическая регрессия лучше регрессии



Откуда берётся сигмоида



$$p(x \mid y=t) = \frac{1}{(2\pi)^{n/2} \mid \Sigma \mid^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_t)^{\mathrm{T}} \Sigma^{-1} (x-\mu_t)\right)$$

нормальное распределение с одинаковыми матрицами ковариации

$$p(y=t \mid x) = \frac{p(x \mid y=t)\, p(y=t)}{p(x \mid y=0)\, p(y=0) + p(x \mid y=1)\, p(y=1)}$$

Откуда берётся сигмоида

считаем классы равновероятными

$$p(y=t\,|\,x)=\frac{\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu_t)^{\mathrm{T}}\Sigma^{-1}(x-\mu_t)\right)}{\sum_t\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu_t)^{\mathrm{T}}\Sigma^{-1}(x-\mu_t)\right)}=$$

$$=\frac{1}{1+\exp\left(+\frac{1}{2}(x-\mu_t)^{\mathrm{T}}\Sigma^{-1}(x-\mu_t)-\frac{1}{2}(x-\mu_{1-t})^{\mathrm{T}}\Sigma^{-1}(x-\mu_{1-t})\right)}=$$

$$=\frac{1}{1+\exp\left(-\frac{1}{2}\mu_t{}^{\mathrm{T}}\Sigma^{-1}x-\frac{1}{2}x^{\mathrm{T}}\Sigma^{-1}\mu_t+\frac{1}{2}\mu_t{}^{\mathrm{T}}\Sigma^{-1}\mu_t+\frac{1}{2}\mu_{1-t}{}^{\mathrm{T}}\Sigma^{-1}x+\frac{1}{2}x^{\mathrm{T}}\Sigma^{-1}\mu_{1-t}-\frac{1}{2}\mu_{1-t}{}^{\mathrm{T}}\Sigma^{-1}\mu_{1-t}\right)}=$$

$$=\sigma(w^{\mathrm{T}}x+w_0)$$

Обучение логистической регрессии

Метод максимального правдоподобия

$$L(w_0,\ldots,w_n)=\prod_{i:y_i=1}p(x_i)\prod_{i:y_i=0}(1-p(x_i))\to\max$$

здесь тоже выпуклая задача оптимизации

$$\log L=-\sum_{i:y_i=1}\log(1+e^{-z_i})-\sum_{i:y_i=0}\log(1+e^{+z_i})\equiv-\sum_i\log(1+e^{-y_i'z_i})$$

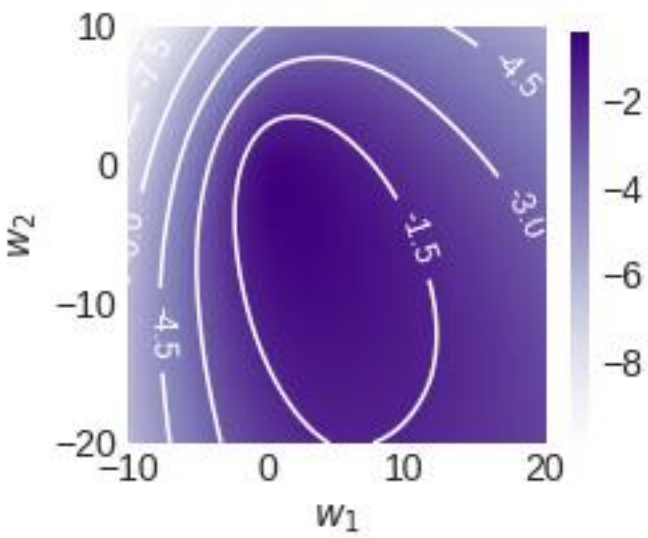
$$\nabla_w\log L=\sum_{i:y_i=1}\frac{1}{1+e^{-w^{\mathrm{T}}x_i}}e^{-w^{\mathrm{T}}x_i}x_i-\sum_{i:y_i=0}\frac{1}{1+e^{+w^{\mathrm{T}}x_i}}e^{+w^{\mathrm{T}}x_i}x_i=$$

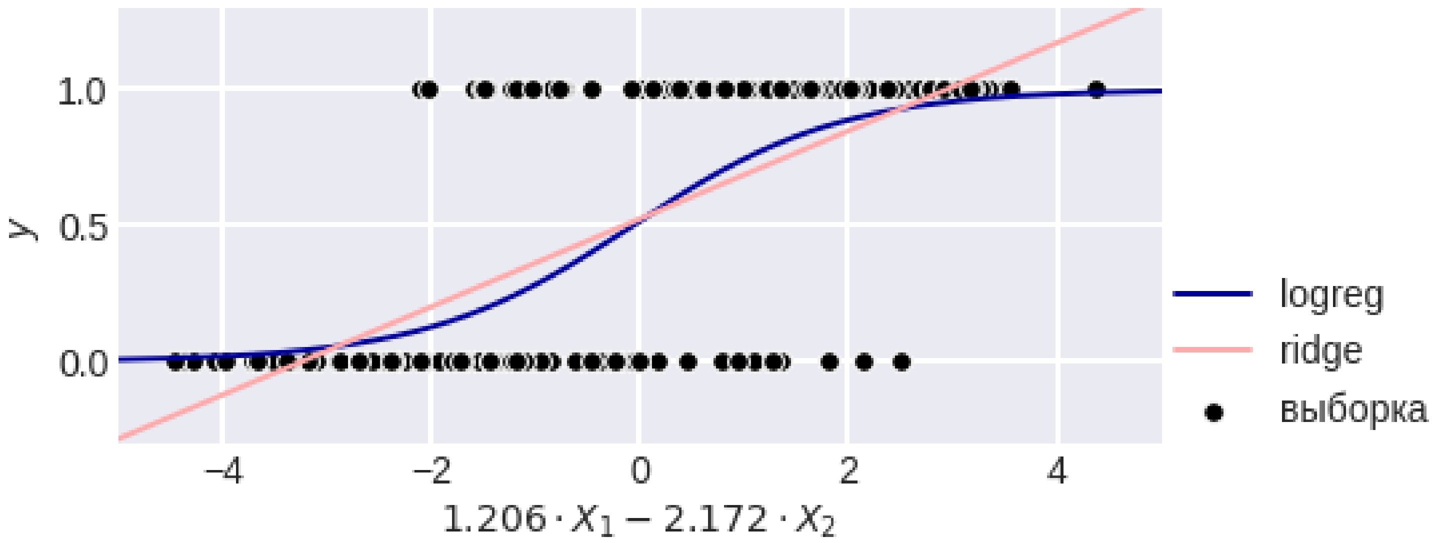
$$=\sum_i\frac{y_i'x_i}{1+e^{+y_i'w^{\mathrm{T}}x_i}}=\sum_i y_i'x_i\sigma(-y_i'w^{\mathrm{T}}x_i)$$

где (для удобства записи)

$$y_i'=2y_i-1$$

Качество логистической регрессии
– логарифм правдоподобия
(потом будет соответствующая функция ошибки logloss)





метод SGD

$$w \leftarrow w + \eta\sigma(-y_i' w^{\mathrm{T}} x_i) y_i' x_i$$

Запомним!

Многоклассовая логистическая регрессия
Multiclass logistic regression (multinomial regression)

в glmnet такой «симметричный вариант»

$$P(Y=k\,|\,x)=\frac{e^{w_{0k}+w_{1k}X_1+\ldots+w_{nk}X_n}}{\sum\limits_{j=1}^{l}e^{w_{0j}+w_{1j}X_1+\ldots+w_{nj}X_n}}$$

Если

$$\mathrm{softmax}(a_1,\ldots,a_l)=\frac{1}{Z}[e^{a_1},\ldots,e^{a_l}],$$

где $Z=e^{a_1}+\ldots+e^{a_l}$

тогда

$$P(Y=k\,|\,x)=\mathrm{softmax}(w(1)^{\mathrm{т}}x,\ldots,w(l)^{\mathrm{т}}x)$$

Реализация в scikit-learn

`sklearn.linear_model.LogisticRegression`

`penalty="l2"`	**Тип регуляризации** Не все солверы поддерживают все типы
`dual=False`	**Переход к двойственной задачи**
`C=1.0`	**Обратная величина к коэффициенту регуляризации**
`fit_intercept=True`	**Свободный член**
`class_weight=None`	**Веса классов**
`solver="warn"`	**Солвер** `"newton-cg", "lbfgs", "liblinear", "sag", "saga"`
`warm_start=False`	**Использовать ли предыдущие начальные условия**
`l1_ratio`	**Формализация штрафов для ElasticNet**

```
tol=0.0001, intercept_scaling=1, random_state=None, max_iter=100,
multi_class='warn', verbose=0, n_jobs=None
```