

Домашнее задание

Задача

Объект: клиент банка

Признаки: заработная плата, наличие собственности, сумма кредита и т. д.

Предсказать: был ли выдан кредит (1: выдан, 0: не выдан)

| | account | duration | credit_history | amount | savings | employment | guarantors | residence | age | ... | target |
|---|---------|----------|----------------|--------|---------|------------|------------|-----------|------|-----|--------|
| 0 | 1 | 0.18 | 2 | 1311.0 | 2 | 2 | 1 | 4 | 34.0 | | 1 |
| 1 | -1 | 0.18 | 0 | 1552.0 | 1 | 3 | 1 | 1 | 31.0 | | 1 |
| 2 | -1 | 0.12 | 2 | 451.0 | 1 | 3 | 1 | 4 | 21.0 | | 0 |
| 3 | -1 | 0.30 | 0 | 4036.0 | 0 | 1 | 1 | 3 | 25.0 | | 1 |
| 4 | 0 | 0.36 | 2 | 909.5 | 1 | 2 | 1 | 4 | 37.0 | | 0 |

Задача

Абсолютно аналогично решаются задачи с другими данными, например предсказание оттока

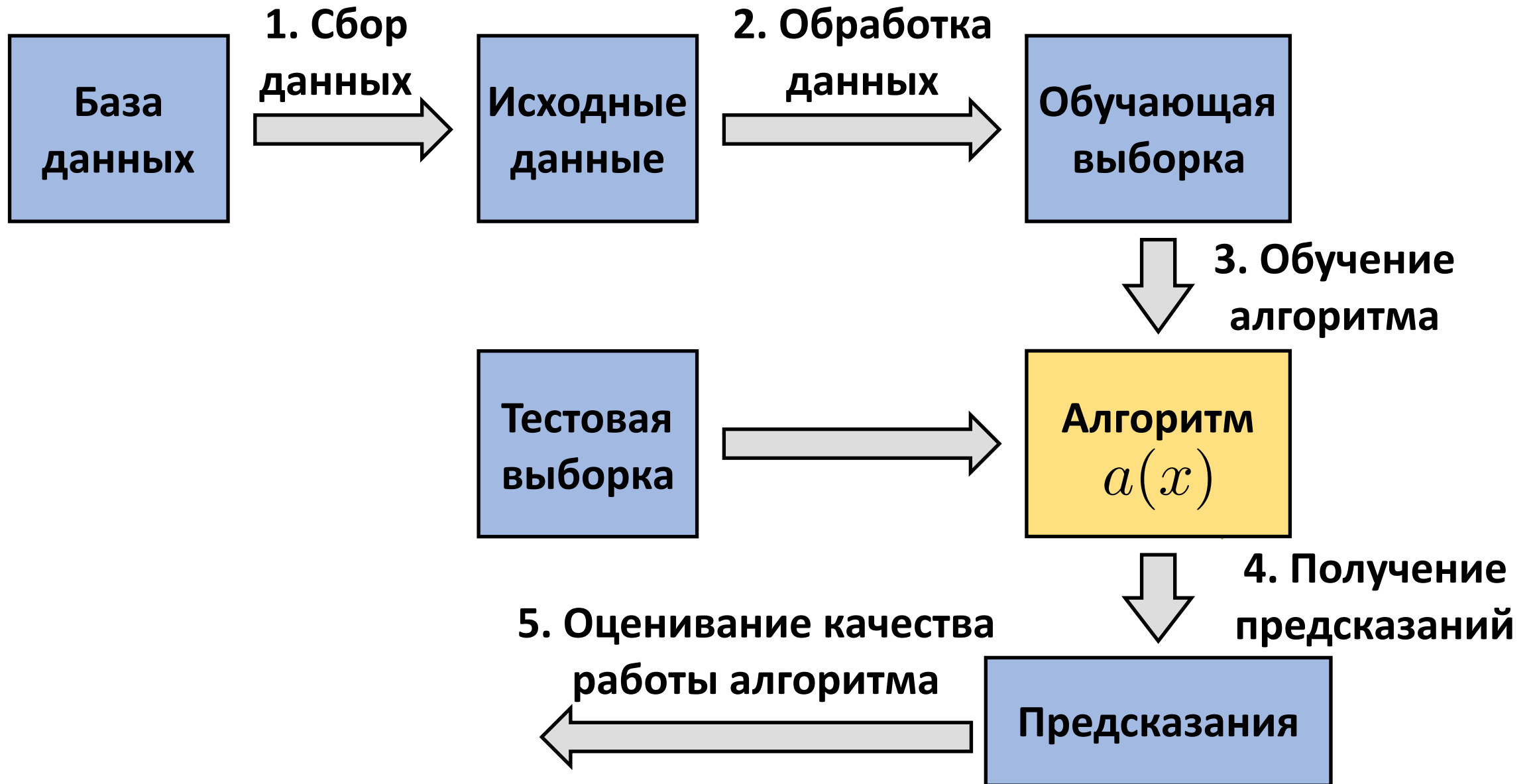
Объект: клиент банка

Признаки: заработная плата, наличие собственности, сумма кредита и т. д.

Предсказать: был ли выдан кредит (1: выдан, 0: не выдан)

| | account | duration | credit_history | amount | savings | employment | guarantors | residence | age | ... | target |
|---|---------|----------|----------------|--------|---------|------------|------------|-----------|------|-----|--------|
| 0 | 1 | 0.18 | 2 | 1311.0 | 2 | 2 | 1 | 4 | 34.0 | | 1 |
| 1 | -1 | 0.18 | 0 | 1552.0 | 1 | 3 | 1 | 1 | 31.0 | | 1 |
| 2 | -1 | 0.12 | 2 | 451.0 | 1 | 3 | 1 | 4 | 21.0 | | 0 |
| 3 | -1 | 0.30 | 0 | 4036.0 | 0 | 1 | 1 | 3 | 25.0 | | 1 |
| 4 | 0 | 0.36 | 2 | 909.5 | 1 | 2 | 1 | 4 | 37.0 | | 0 |

Схема работы машинного обучения



Линейные модели для задачи бинарной классификации

Линейная модель суммирует значения всех признаков с некоторыми весами и **сравнивает с порогом** (обычно ноль)

Веса при признаках — параметры, которые необходимо настраивать в процессе обучения

$$score = w_0 + w_1x_1 + \dots w_dx_d$$

d — число признаков

$$a(x) = \begin{cases} +1, & if \ score > 0 \\ -1, & if \ score < 0 \end{cases}$$

Измерение качества в классификации

| $a(x)$ | y | отклонение? |
|-----------|-----------|-------------|
| выдать | выдать | ? |
| выдать | не выдать | ? |
| не выдать | не выдать | ? |
| не выдать | не выдать | ? |

Измерение качества в классификации

| $a(x)$ | y | совпадение |
|-----------|-----------|------------|
| выдать | выдать | да |
| выдать | не выдать | нет |
| не выдать | не выдать | да |
| не выдать | не выдать | да |

Измерение качества в классификации

Доля правильных ответов (accuracy):

$$Accuracy = \frac{1}{\ell} \sum_{i=1}^{\ell} [a(x_i) = y_i]$$

ℓ — число прецедентов в выборке

* выражение $[x]$ равно единице, если x является верным утверждением, и нулю иначе

В примере:

$$Accuracy = \frac{1}{4} (1 + 0 + 1 + 1) = 0.75$$