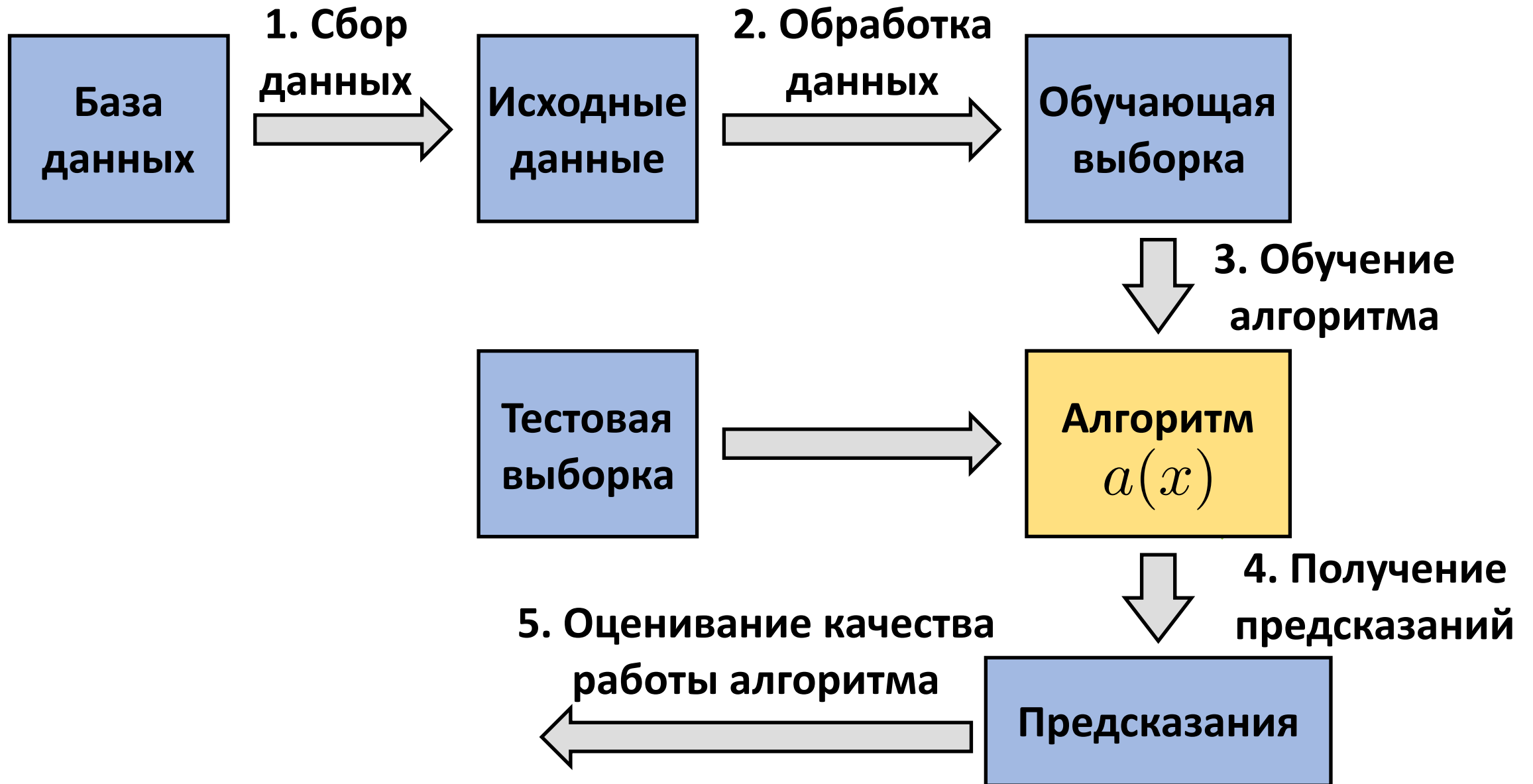


Машинное обучение:
напоминание

Схема работы машинного обучения



Напоминание: обучающая выборка

Обучающая выборка

объект
 x_i

Площадь	Год постройки	Число комнат	Цена
45	1995	1	7000000
60	2005	2	9900000
35	2010	1	5500000

ответ
 y_i

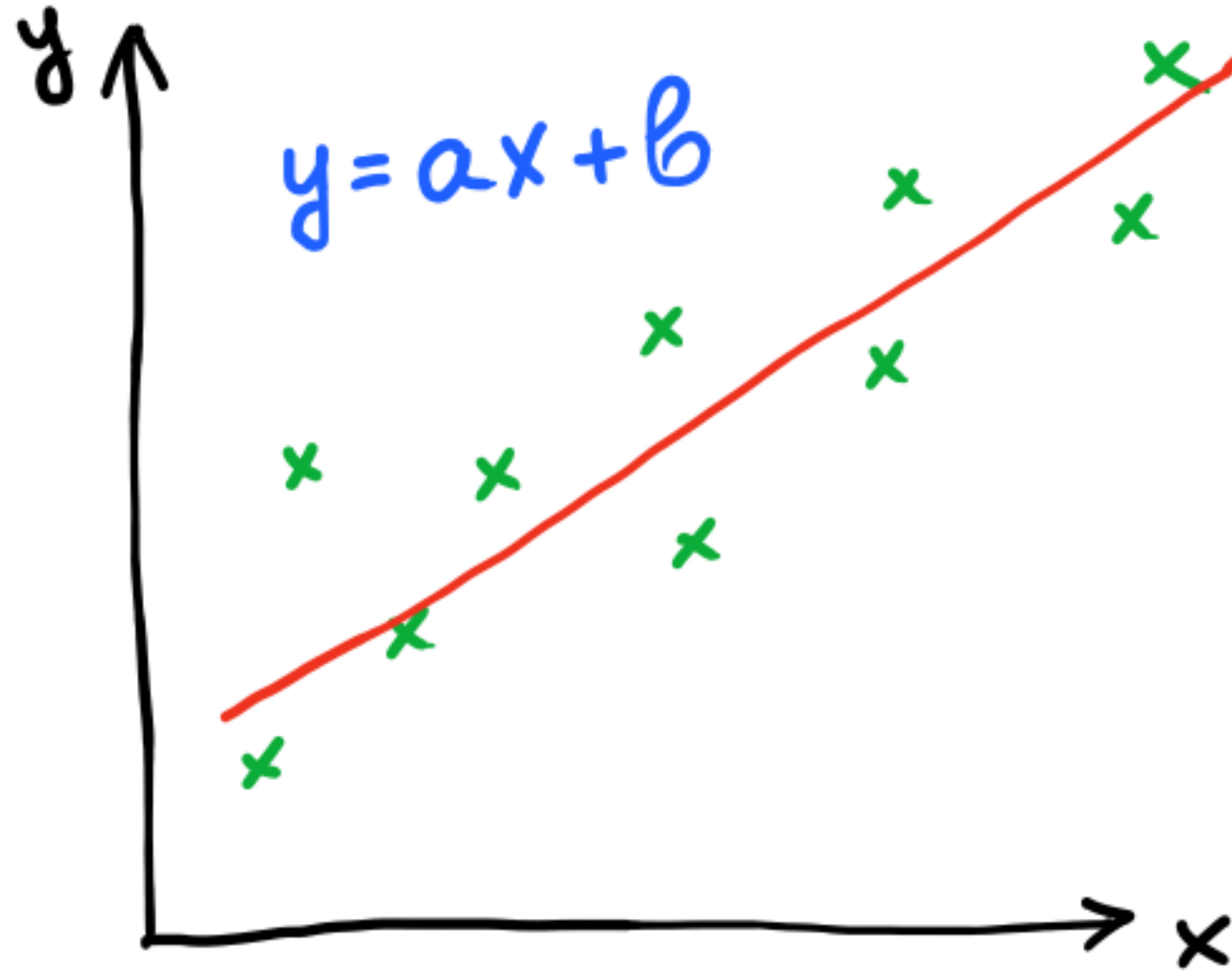
Обобщающая способность алгоритма

- Обучаем алгоритм на **обучающих** данных, измеряем качество на **тестовых**:



Линейная модель для регрессии

x	y
1	2
3	5
-1	-2
5	?



Линейные модели для задачи регрессии

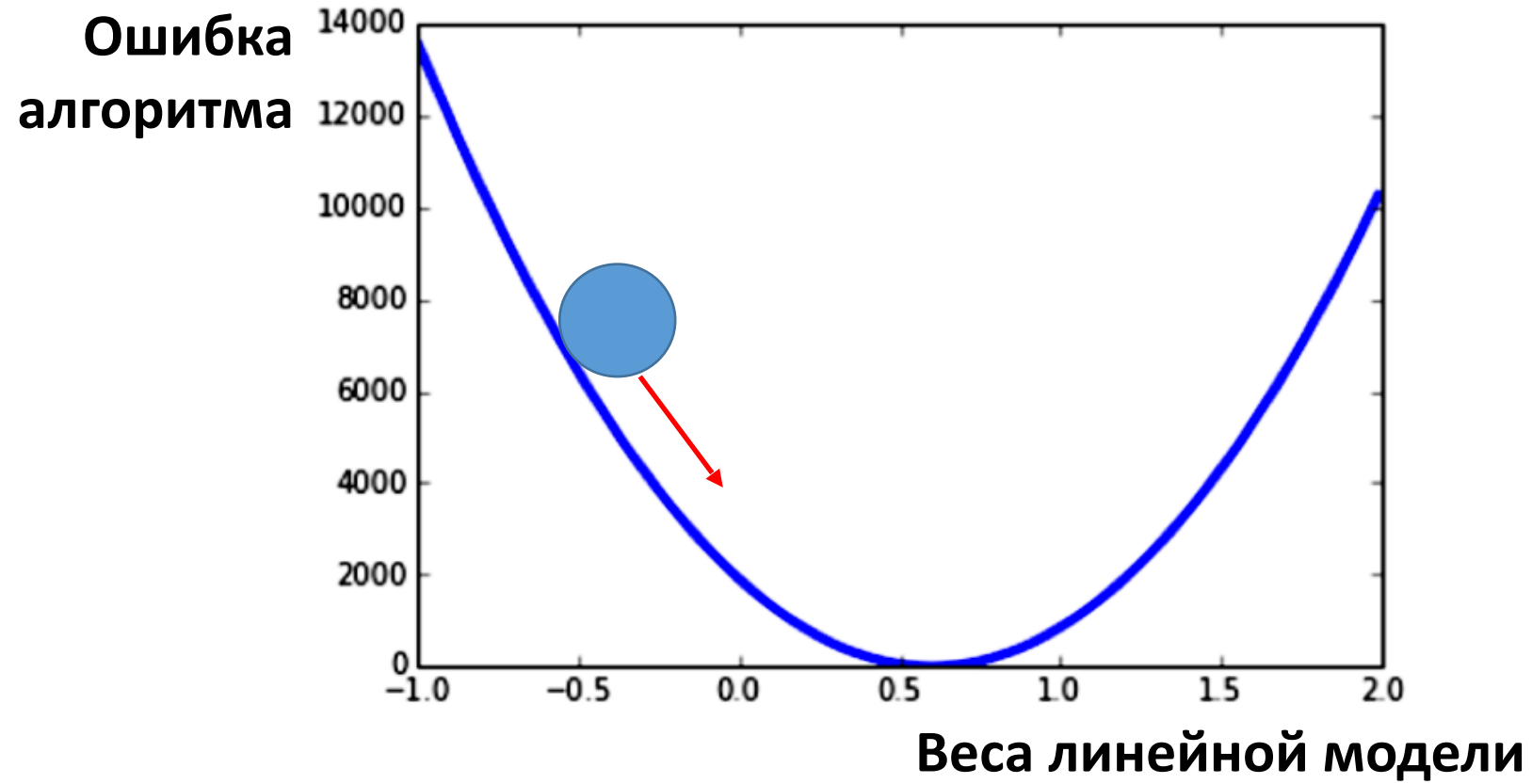
Линейная модель суммирует значения всех признаков с некоторыми весами

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

$$a(x) = w_0 + w_1x_1 + \dots w_dx_d$$

d — число признаков

Обучение линейной модели



Метод k ближайших соседей

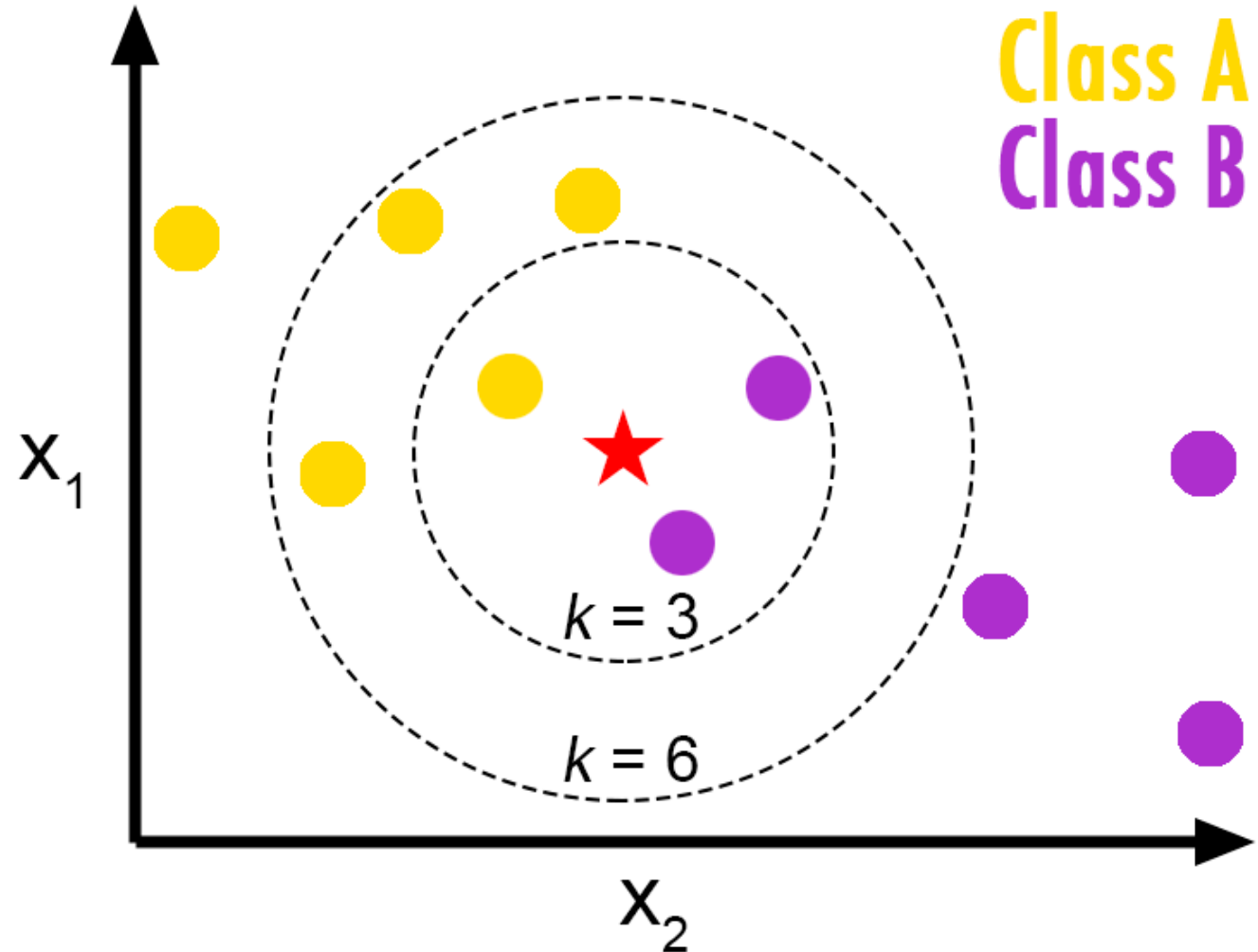
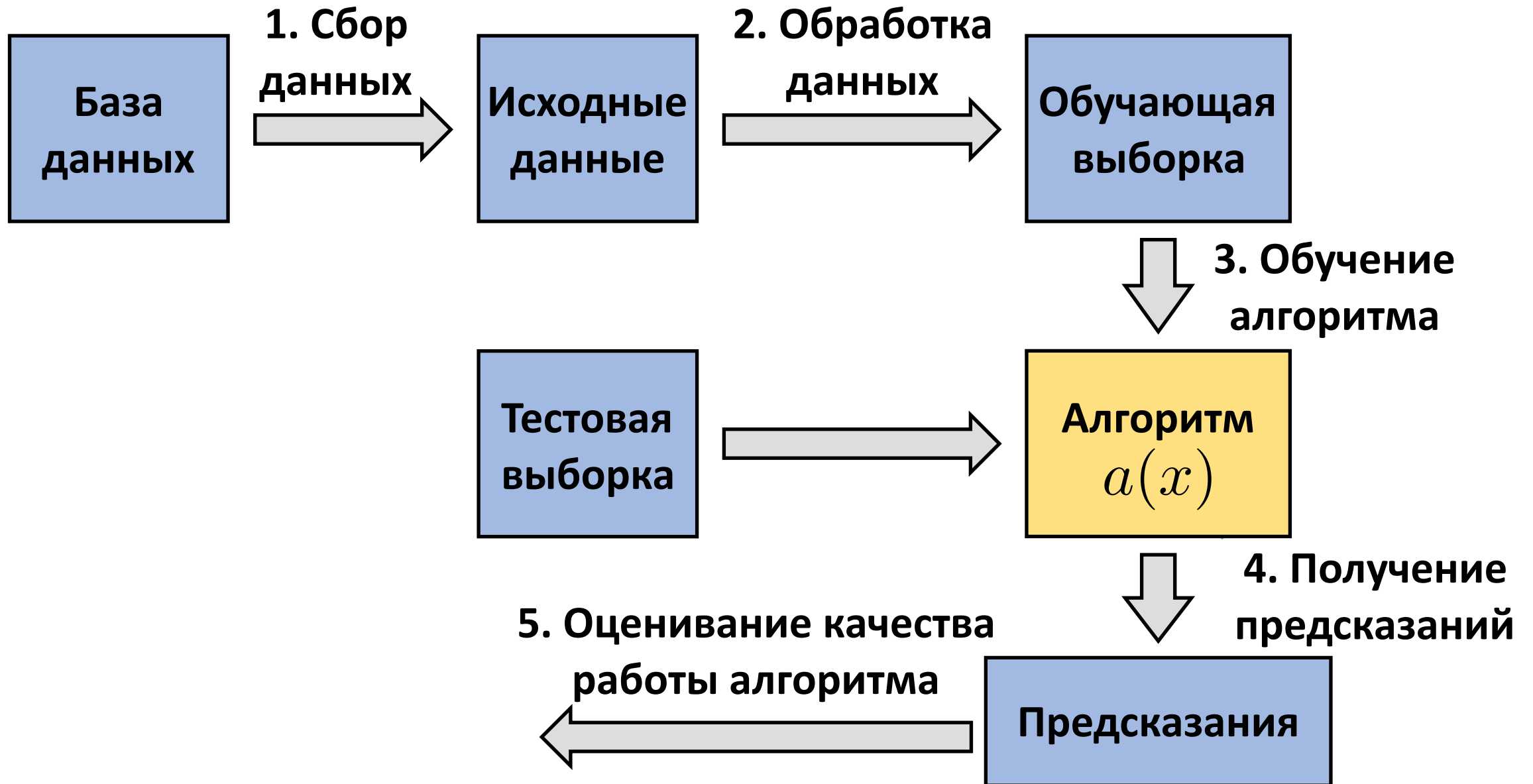


Схема работы машинного обучения



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 регрессии

$a(x)$	y	отклонение?
11	10	?
9	10	?
20	10	?
1	10	?

Измерение качества в регрессии

$a(x)$	y	отклонение
11	10	1
9	10	-1
20	10	10
1	10	-9

Измерение качества в регрессии

$a(x)$	y	$ a(x) - y $
11	10	1
9	10	1
20	10	10
1	10	9

Измерение качества в регрессии

Среднее абсолютное отклонение, или MAE (Mean Absolute Error):

$$MAE = \frac{1}{\ell} \sum_{i=1}^{\ell} |a(x_i) - y_i|$$

ℓ — число объектов в выборке

В примере:

$$MAE = \frac{1}{4} (1 + 1 + 10 + 9) = 5.25$$

Схема работы машинного обучения

