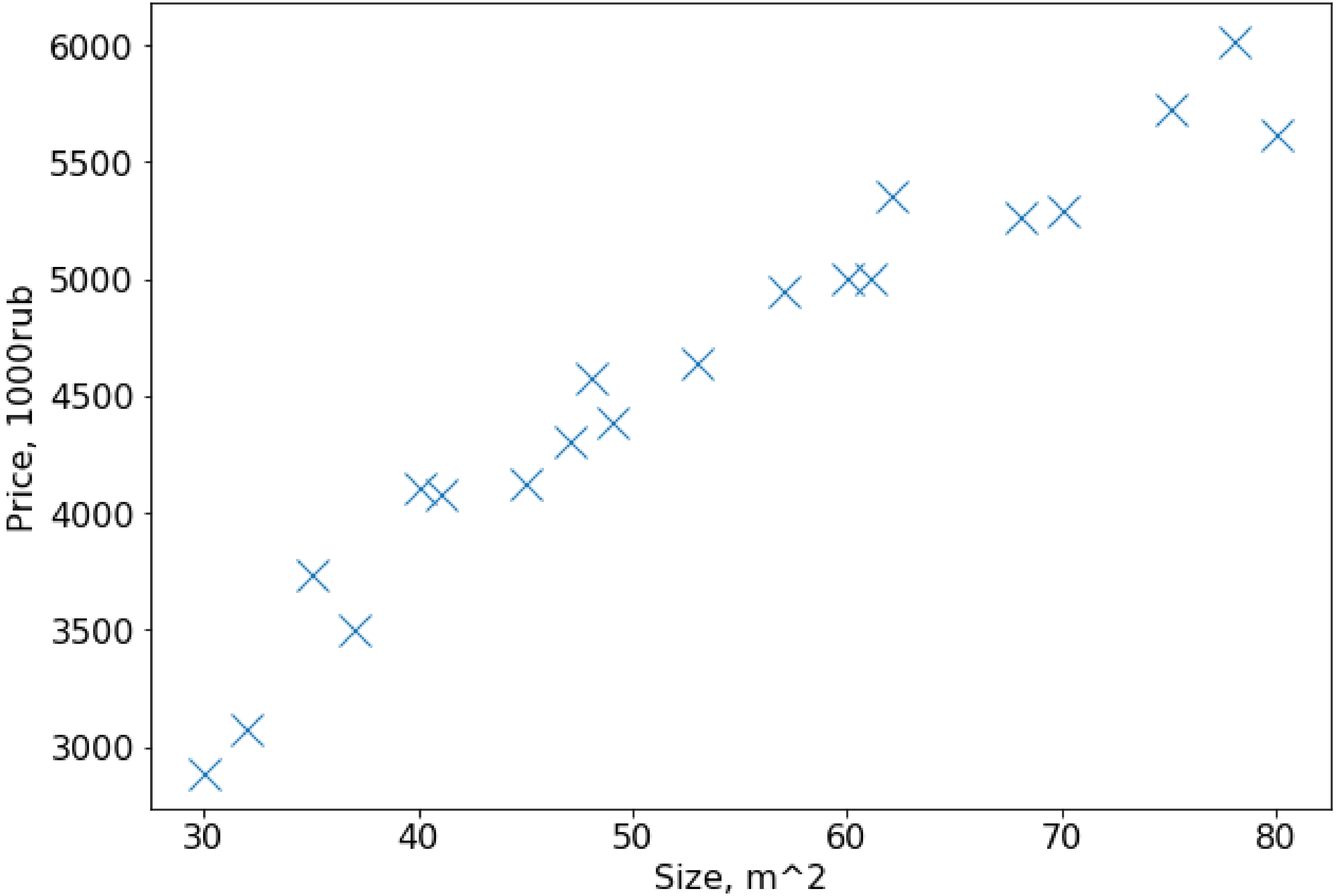
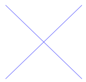
Машинное обучение, 1.2
Линейная регрессия

Задачка

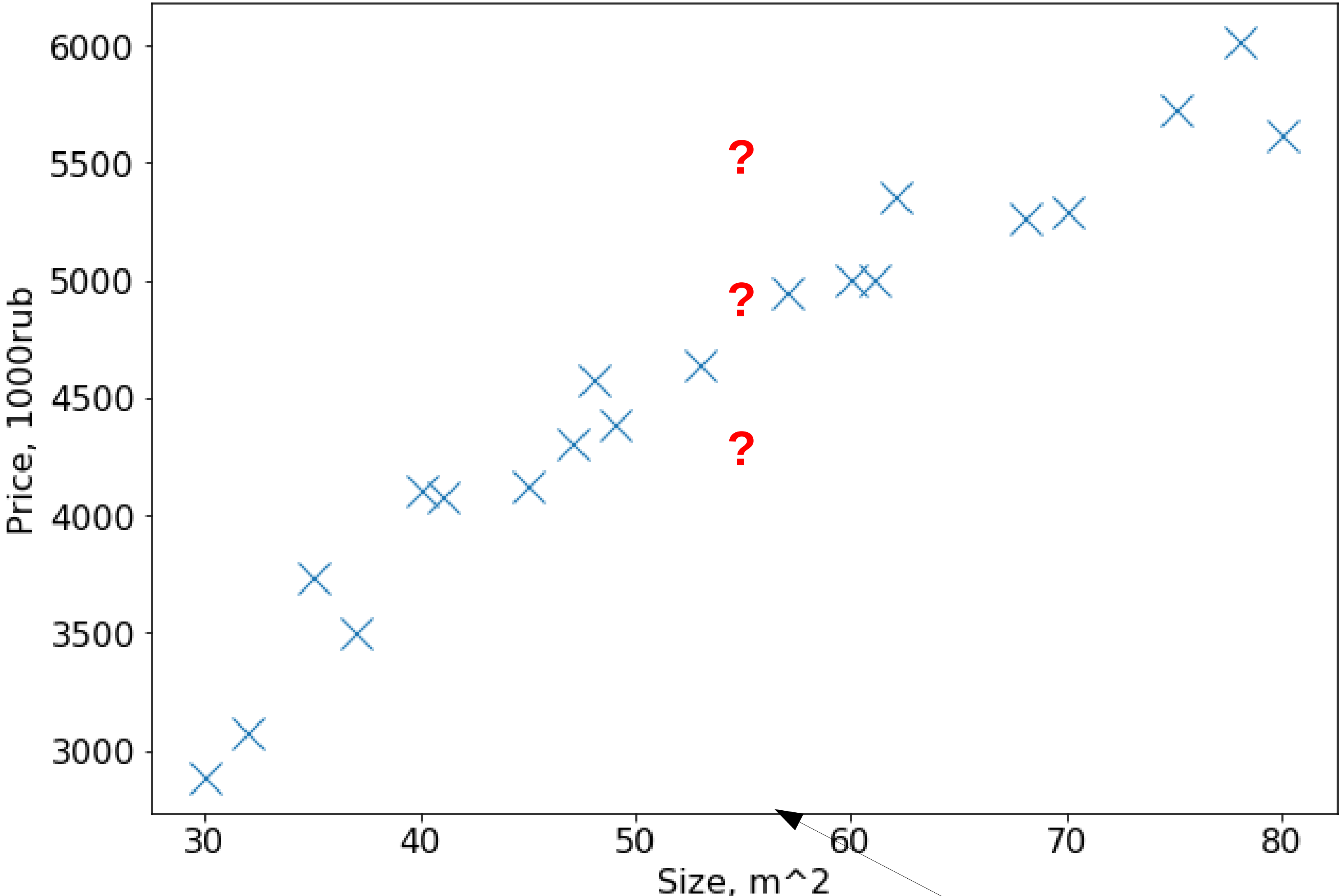


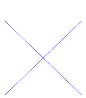


Данные
X — площадь квартиры
Y — цена

Задача:
Предсказывать цену там где она не указана

Задачка

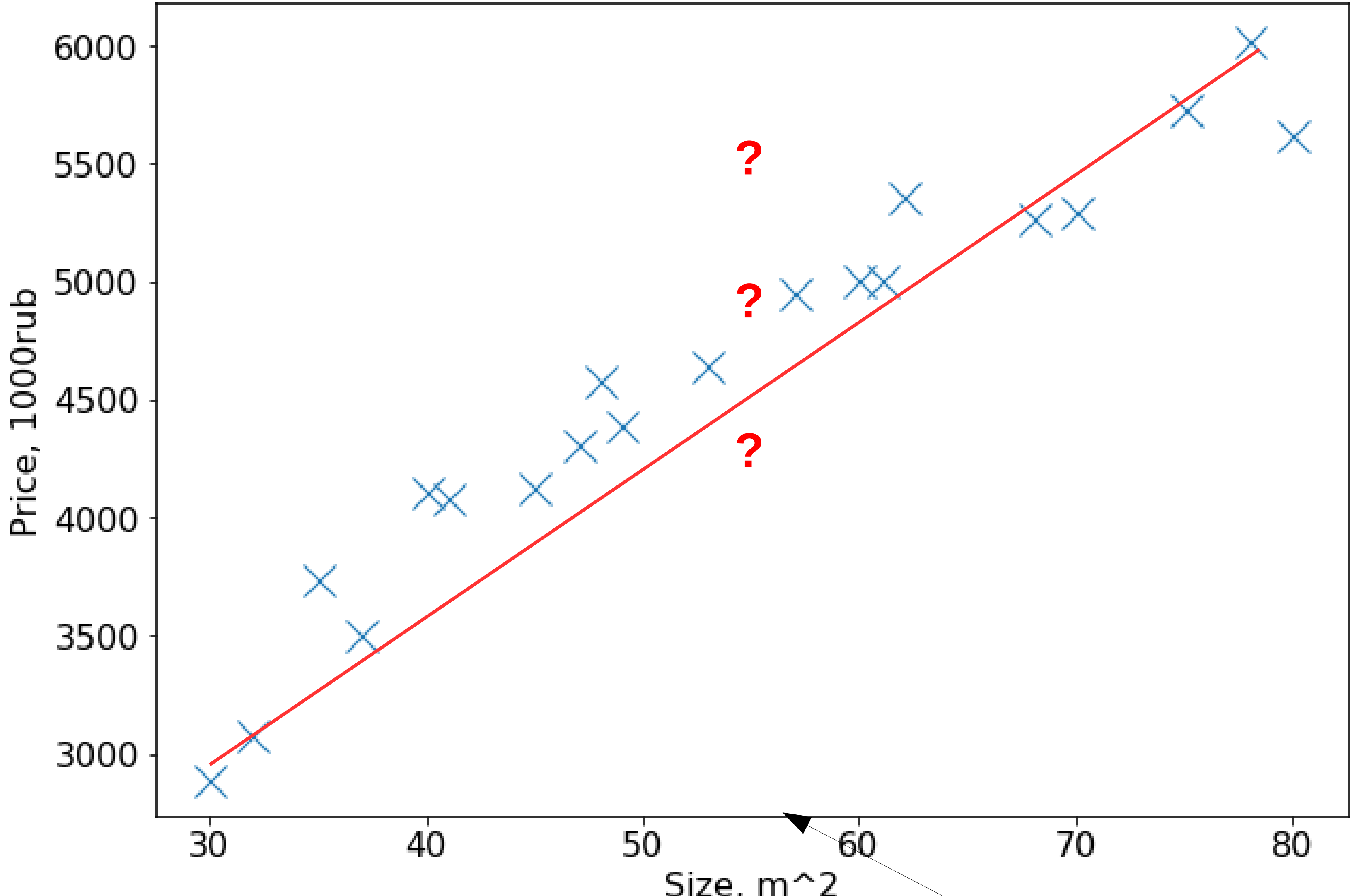


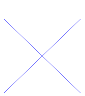


Данные
X — площадь квартиры
Y — цена

Задача:
Предсказывать цену там
где она не указана

Задачка

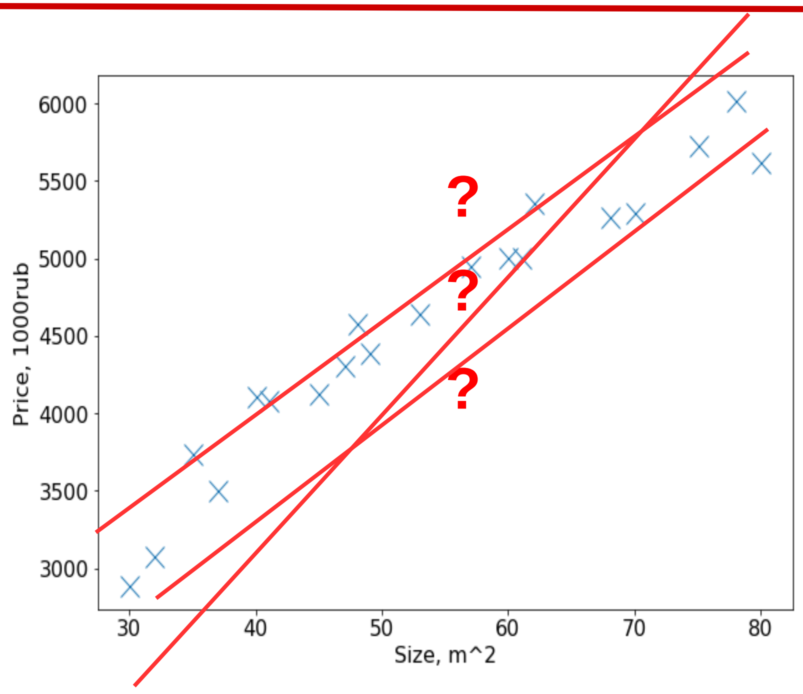




Данные
X — площадь квартиры
Y — цена

Задача:
Предсказывать цену там
где она не указана

Задачка



Гипотеза:

$$H(x|a,b) = ax + b$$

Фиксированный набор данных (x^i, y^i)

Оценка гипотезы:

$$I(a,b) = \frac{1}{2N} \sum_i \left(H\left(x^{(i)}|a,b\right) - y^{(i)}\right)^2$$

Хорошие a и b:

$$\min_{a,b} I(a,b) = \frac{1}{2N} \sum_i \left(ax^{(i)} + b - y^{(i)}\right)^2$$

Несколько переменных

Площадь	L от центра города	Кол-во комнат	Как давно был ремонт	Наличие паркинга

Цена

$$H(x|a_1 \ldots a_n, b) = a_1 x_1 + a_2 x_2 + \ldots + a_5 x_5 + b \qquad \theta = (a_1, a_2, \ldots, a_5, b)$$

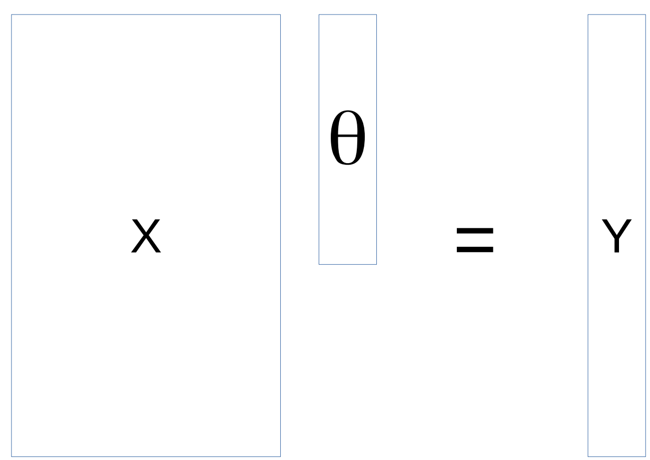
Площадь	L от центра города	Кол-во комнат	Как давно был ремонт	Наличие паркинга	1

Цена

$$H(x|\theta) = a_1 x_1 + a_2 x_2 + \ldots + a_5 x_5 + a_6 \cdot 1 \qquad \theta = (a_1, a_2, \ldots, a_5, a_6)$$

$$H(x^{(i)}|\theta) = (x^{(i)}, \theta)$$

или



$$X\theta = Y$$

Несколько переменных

Оценка гипотезы:

$$I(\theta)=\frac{1}{2N}\sum_i \left(H\left(x^{(i)}|\theta\right)-y^{(i)}\right)^2$$

Хорошие а

$$min_\theta I(\theta)=\frac{1}{2N}\sum_i \left(\theta\, x^{(i)}-y^{(i)}\right)^2$$

Как решать такую задачу?

$$min_{\theta} I(\theta) = \frac{1}{2N} \sum_i \left(\theta x^{(i)} - y^{(i)}\right)^2$$

$$\theta = \left(a_{0,}\ldots, a_n\right)$$

Градиентный спуск

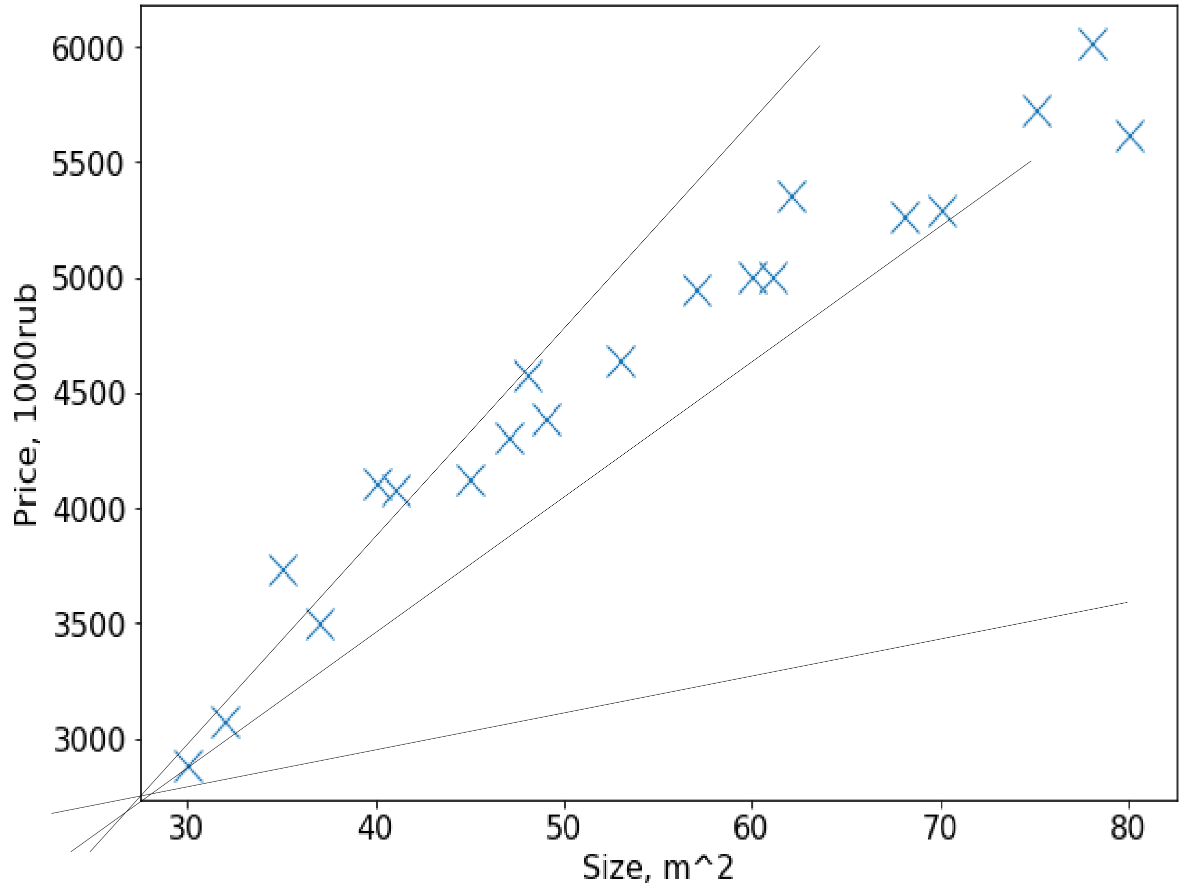
Упрощеный
одномерный случай

$$H(x|a) = ax$$

$$\alpha = (a)$$

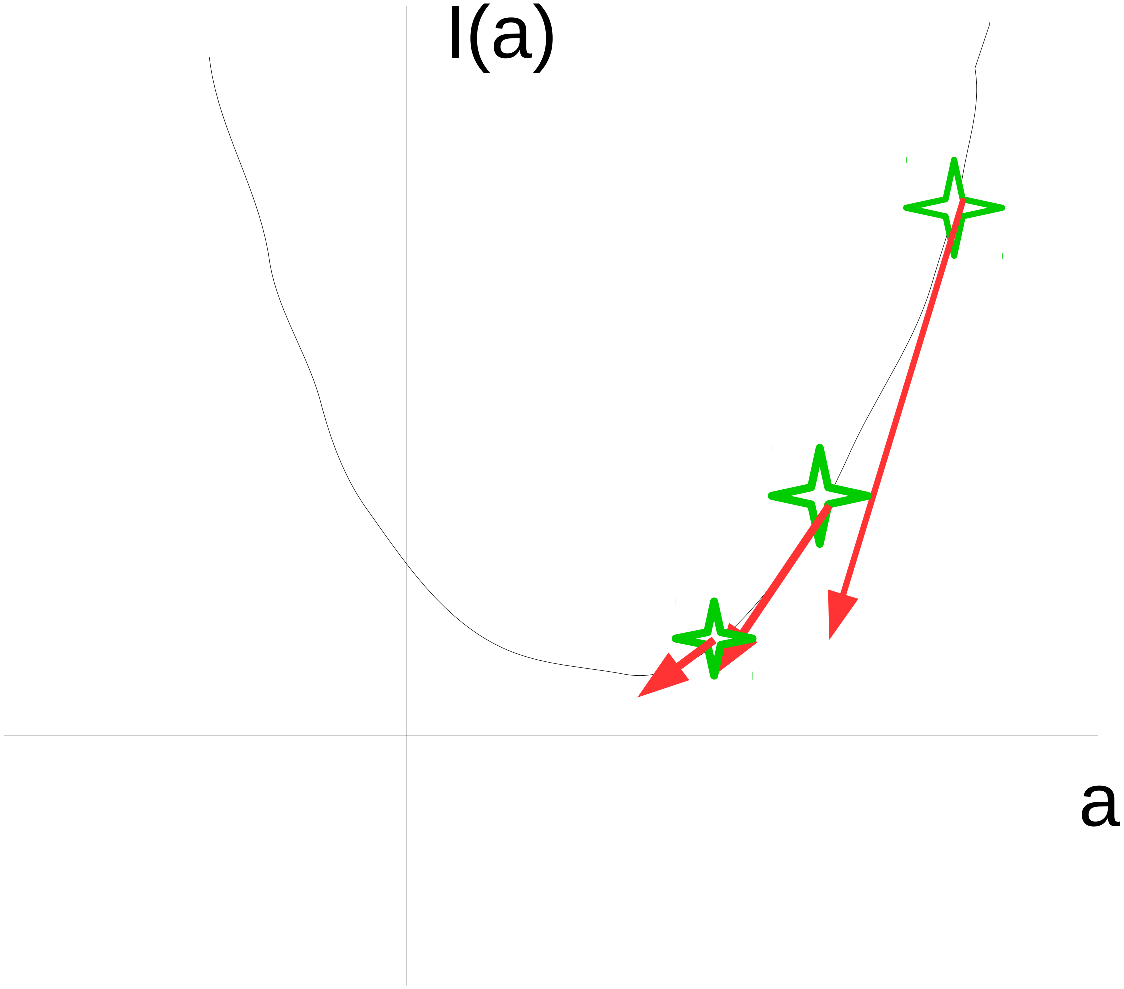
$$min_a I(a) = \frac{1}{2N} \sum_i (ax^{(i)} - y^{(i)})^2$$

Парабола I(a)



Градиентный спуск

$$min_a I(a) = \frac{1}{2N} \sum_i \left(ax^{(i)} - y^{(i)}\right)^2 \longrightarrow a = \frac{\sum x^i y^i}{\sum (x^i)^2}$$

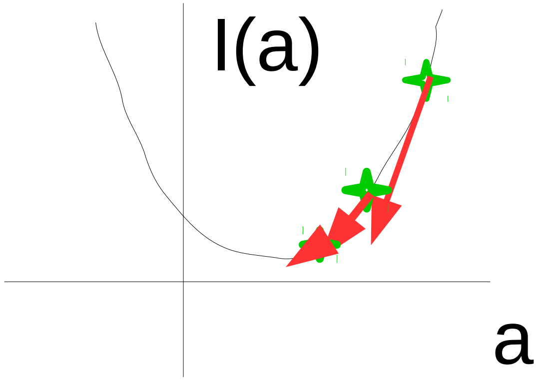


Повторяй пока что-то происходит

1. Берем случайное значение a
2. Считаем производную I(a) по a в точке a
3. Смещаем a в сторону в которую указывет производная

Градиентный спуск

$$min_a I(a) = \frac{1}{2N} \sum_i (ax^{(i)} - y^{(i)})^2$$



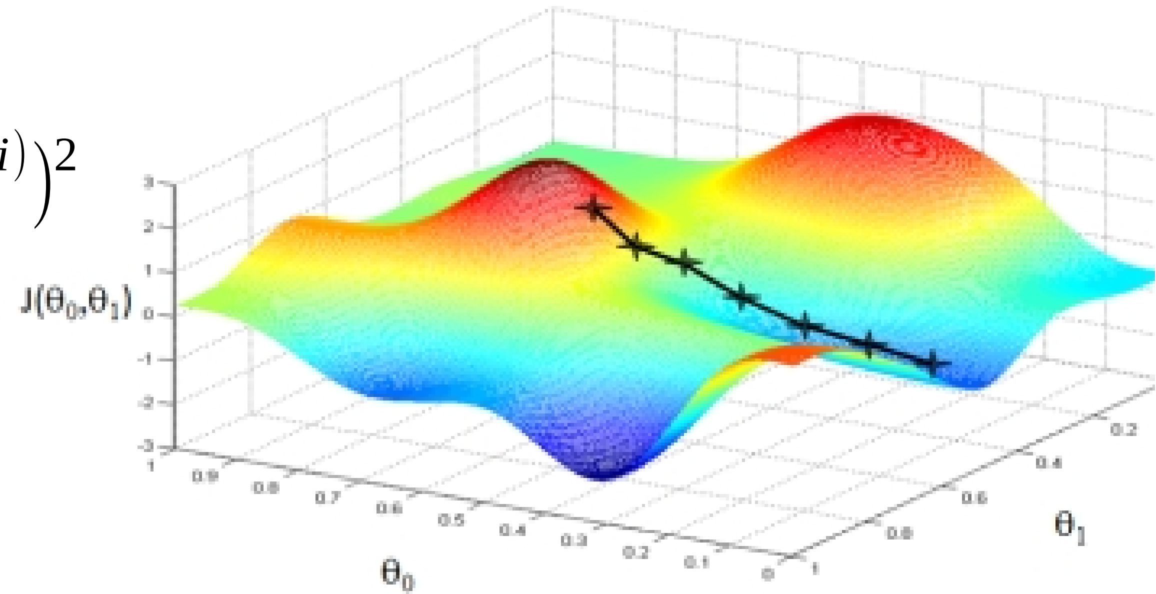
1) Начиаем с произвольного значения а : $a = 239$
2) Считаем производную

$$I'(a)_a = \frac{1}{N} \sum_i (ax^{(i)} - y^{(i)}) x^i$$

3) Сдвигаем а в нужную сторону

$$a_{t+1} = a_t - \alpha \frac{1}{N} \sum_i (a_t x^{(i)} - y^{(i)}) x^i$$, где α малая величина

Градиентный спуск



$$min_\theta I(\theta)=\frac{1}{2N}\sum_i (\theta x^{(i)}-y^{(i)})^2$$

$$\theta=(a_{0,},\ldots,a_n)$$

1) Начиаем с произвольного значения : $\theta=(0.1,0.25,\ldots,0.04)$

2) Считаем производные по всем а

$$I'(\theta)_{a_i}=\left[\frac{1}{2N}\sum_j(\theta x^{(j)}-y^{(j)})^2\right]'_{a_i}=\left[\frac{1}{2N}\sum_j\left(\sum_i(a_i x_i^{(j)})-y^{(j)}\right)^2\right]'_{a_i}$$

$$\left[\frac{1}{2N}\sum_j\left(\sum_i(a_i x_i^{(j)})-y^{(j)}\right)^2\right]'_{a_i}=\frac{1}{N}\sum_j\left(\sum_i(a_i x_i^{(j)})-y^{(j)}\right)x_i^j=\frac{1}{N}\sum_j(\theta x^j-y^{(j)})x_i^j$$

3) Сдвигаем а в нужную сторону

$a_i=a_i-\alpha\frac{1}{N}\sum_j(\theta x^{(j)}-y^{(j)})x_{,i}^i$ где α малая величина

Градиентный спуск, Алгоритм

1)Начиаем с произвольного значения : $\theta=(0.1, 0.25, \ldots, 0.04)$

2)Пока $\sum_i (a_{i_t}-a_{i_{t+1}})^2 > \epsilon$

$$a_i = a_i - \alpha \frac{1}{N} \sum_j (\theta x^{(j)} - y^{(j)}) x_i^i,$$ где α малая величина

Градиентный спуск

Устранение недостатков:

1. Пока никак не устранить
2. Вычисление

$$a_i = a_i - \alpha \frac{1}{N} \sum_j \left(\theta x^{(j)} - y^{(j)}\right) x_i^j$$

Трудное вычисление $\sum_j \left(\theta x^{(j)} - y^{(j)}\right) x_i^j$

Градиентный спуск

Не будем считать честно, попробуем оценить:

$$\sum_j \left(\theta x^{(j)} - y^{(j)}\right) x_i^j$$

Возьмем не все x^j а небольшое подмножество J

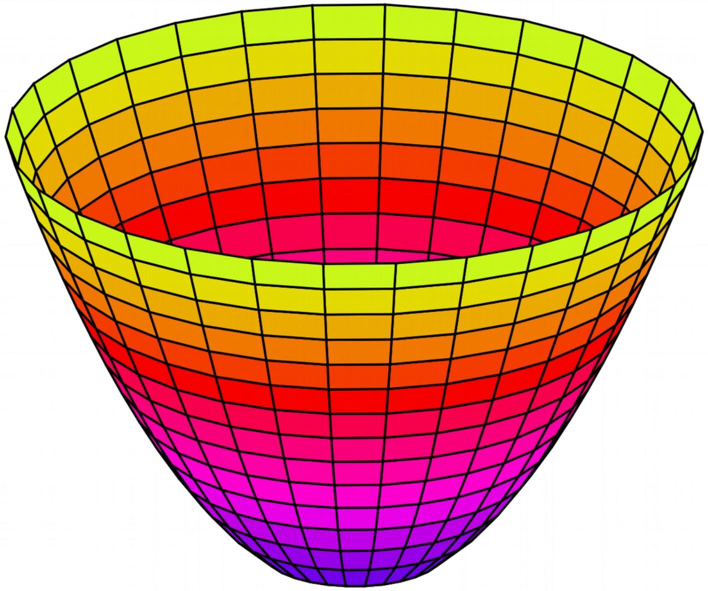
Оценим:

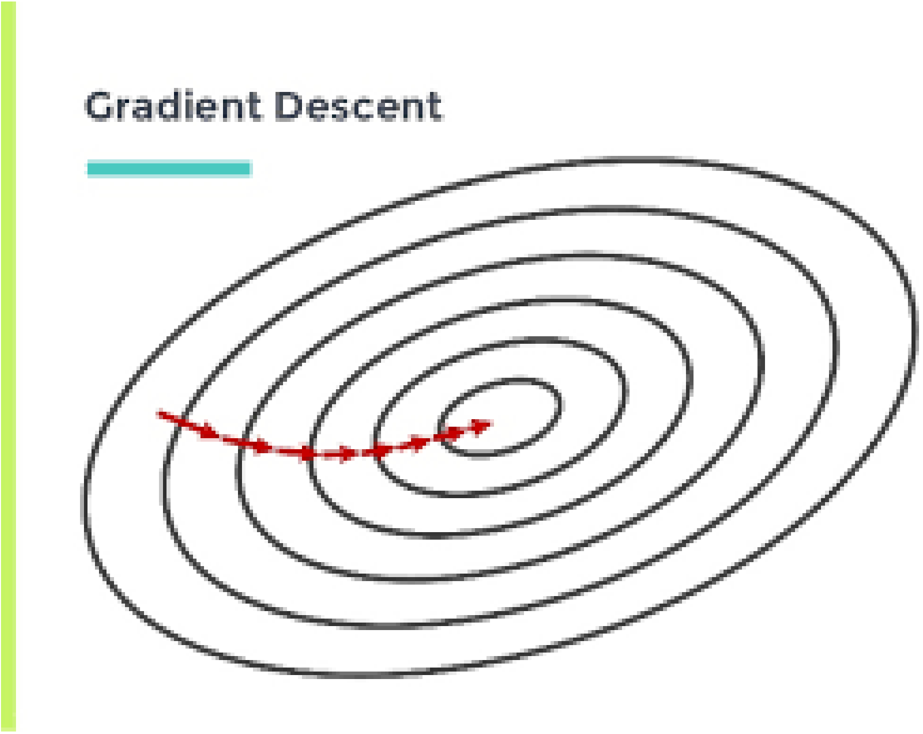
$$\sum_{all\ j} \left(\theta x^{(j)} - y^{(j)}\right) x_i^j \approx \sum_{j \in J} \left(\theta x^{(j)} - y^{(j)}\right) x_i^j$$

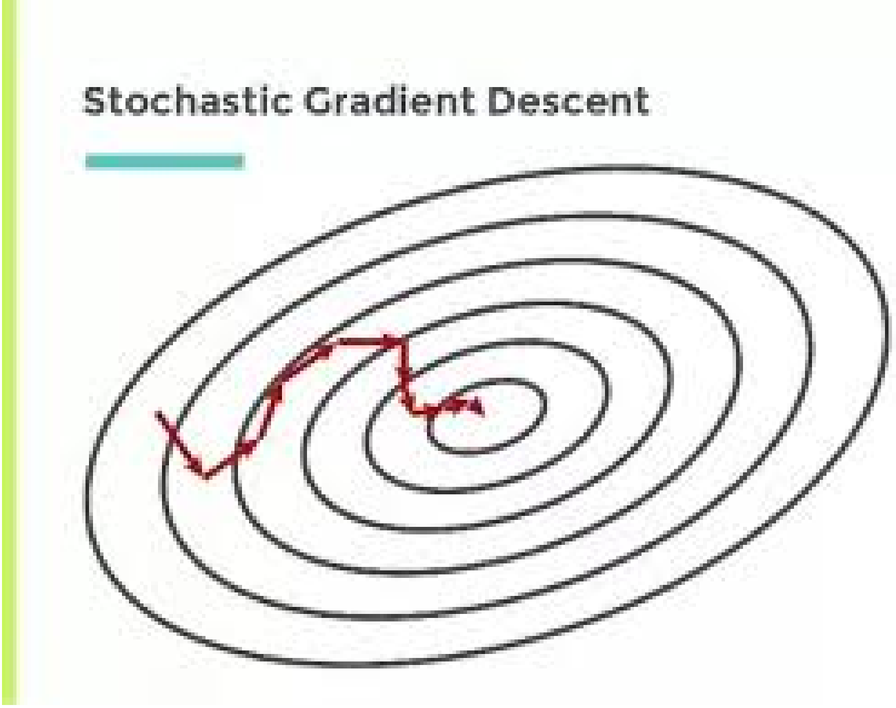
Всего примеров в данных может быть несколько миллионов, а мы можем брать подмножества J размером 10 или 100.

На каждом шаге нужно брать случайное подмножество

Градиентный спуск



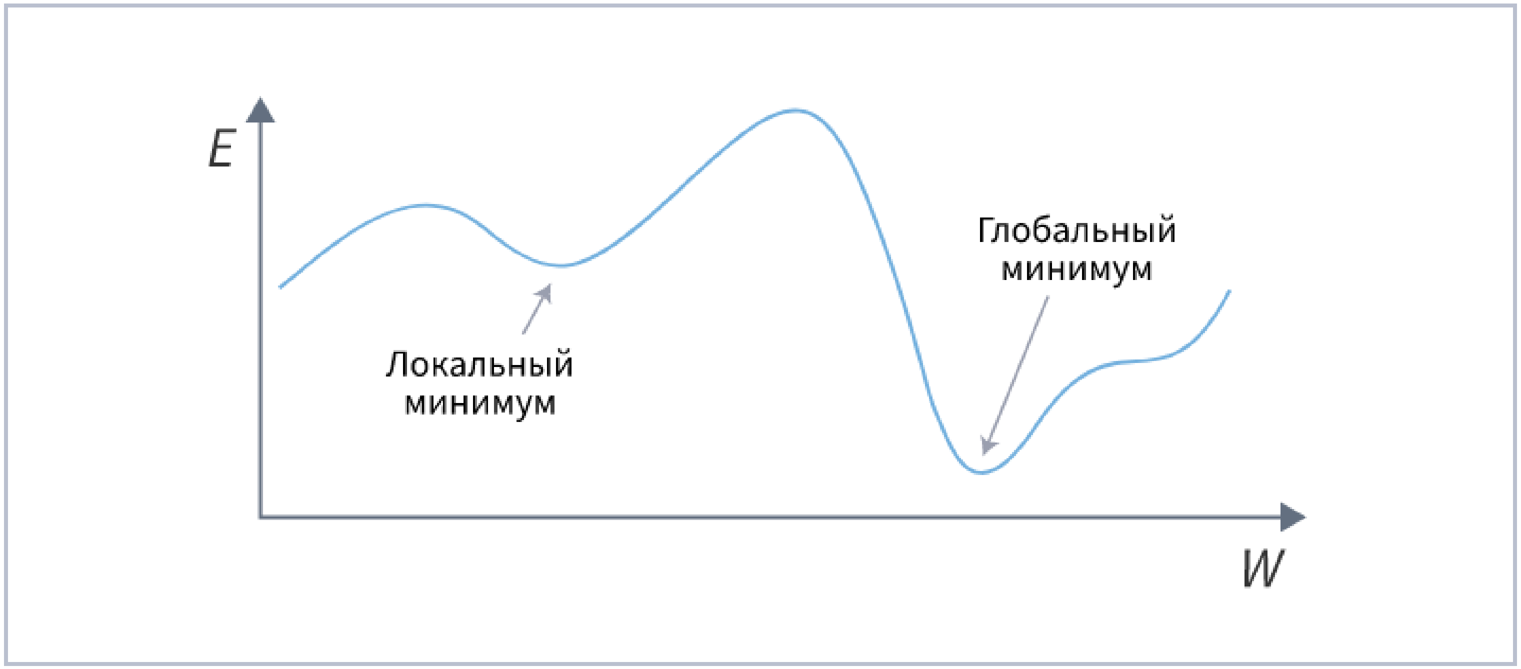




Градиентный спуск

Недостатки:

1. Не всегда находит оптимальное решение



2. Вычисление

$$a_i = a_i - \alpha \frac{1}{N} \sum_j \left(\theta x^{(j)} - y^{(j)}\right) x_i^j$$

занимает много времени на больших данных.

МОЁ ХОББИ: ЭКСТРАПОЛИРОВАТЬ