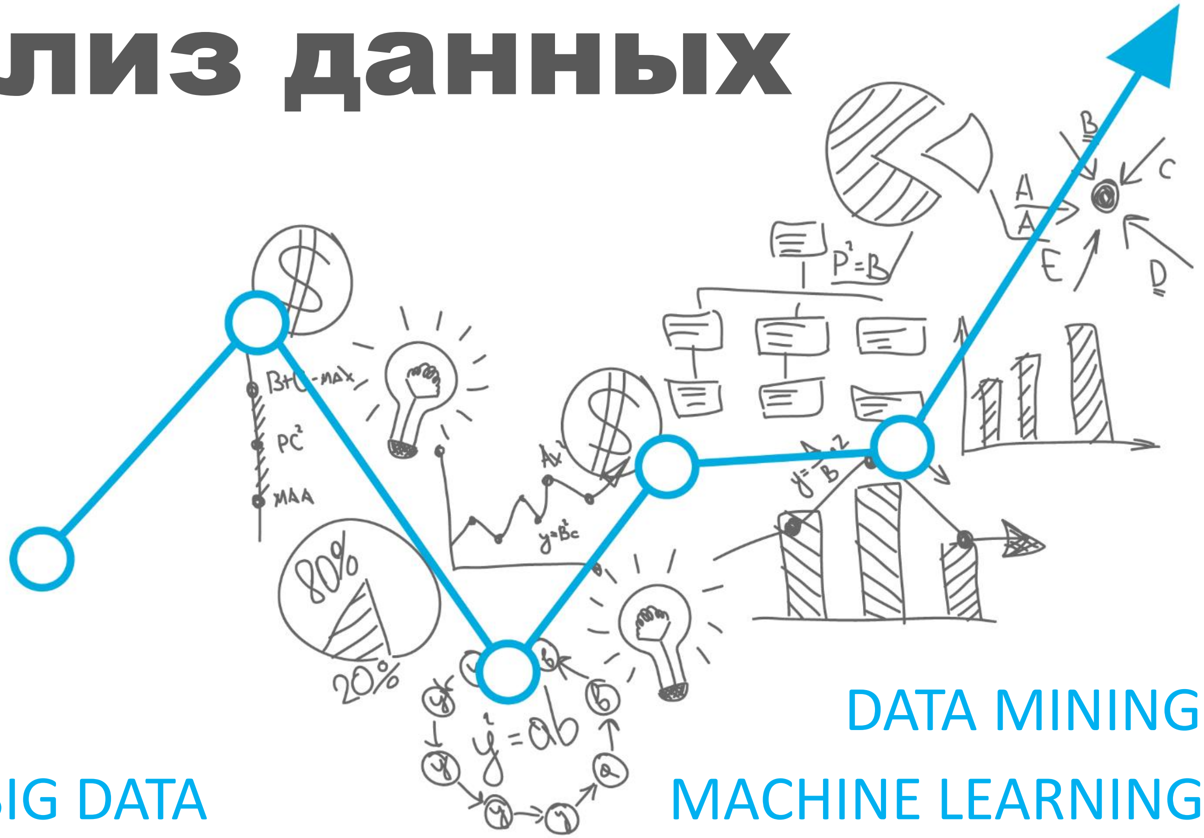
Интеллектуальный анализ данных



DATA MINING

BIG DATA

MACHINE LEARNING

Лекция 3. Линейная регрессия нескольких переменных

Материал прошлых лекций

Матрица $m \times n$

$$A = \underbrace{\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}}_{n} \left.\vphantom{\begin{bmatrix} A \\ A \\ A \end{bmatrix}}\right\} m \qquad \mathbb{R}^{m \times n}$$

Элемент матрицы A_{ij}

$$A = \begin{bmatrix} 1402 & 191 \\ 1371 & 821 \\ 147 & 1437 \end{bmatrix}$$

$$A_{11} = 1402$$
$$A_{12} = 191$$
$$A_{21} = 1371$$
$$A_{32} = 1437$$
$$A_{23} = undefined$$

Вектор – матрица размерами $n \times 1$

$$y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}$$

$$\mathbb{R}^n$$

Индексация на базе

нуля

$$y = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

единицы

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

Материал прошлых лекций

Сложение матриц
$$\begin{bmatrix}1 & 0\\ 2 & 5\\ 3 & 1\end{bmatrix}+\begin{bmatrix}4 & 0.5\\ 2 & 5\\ 0 & 1\end{bmatrix}=\begin{bmatrix}1+4 & 0+0.5\\ 2+2 & 5+5\\ 3+0 & 1+1\end{bmatrix}=\begin{bmatrix}5 & 0.5\\ 4 & 10\\ 3 & 2\end{bmatrix}$$

$$\begin{bmatrix}1 & 0\\ 2 & 5\\ 3 & 1\end{bmatrix}+\begin{bmatrix}5 & 9\\ 1 & 0.7\end{bmatrix}=error$$

Операция сложения матриц является *коммутативной*: $A+B=B+A$

Умножения матрицы на число
$$3\times\begin{bmatrix}1 & 0\\ 2 & 5\\ 3 & 9\end{bmatrix}=\begin{bmatrix}3\times 1 & 3\times 0\\ 3\times 2 & 3\times 5\\ 3\times 3 & 3\times 9\end{bmatrix}=\begin{bmatrix}3 & 0\\ 6 & 15\\ 9 & 27\end{bmatrix}=\begin{bmatrix}1 & 0\\ 2 & 5\\ 3 & 9\end{bmatrix}\times 3$$

$$\begin{bmatrix}4 & 0\\ 6 & 3\end{bmatrix}/4=\frac{1}{4}\times\begin{bmatrix}4 & 0\\ 6 & 3\end{bmatrix}=\begin{bmatrix}1 & 0\\ 1.5 & 3/4\end{bmatrix}$$

Комбинирование операций (приоритеты как в арифметике)
$$3\times\begin{bmatrix}1\\ 4\\ 2\end{bmatrix}+\begin{bmatrix}0\\ 1\\ 5\end{bmatrix}-\begin{bmatrix}2\\ 4\\ 8\end{bmatrix}/2=\begin{bmatrix}3\\ 12\\ 6\end{bmatrix}+\begin{bmatrix}0\\ 1\\ 5\end{bmatrix}-\begin{bmatrix}1\\ 2\\ 4\end{bmatrix}=\begin{bmatrix}2\\ 11\\ 7\end{bmatrix}$$

Материал прошлых лекций

Умножение матрицы на вектор

$$A^{m\times n} \times x^{n} = y^{m}$$

$$\begin{bmatrix} \boxed{1 & 3} \\ 4 & 0 \\ 2 & 1 \end{bmatrix} \times \boxed{\begin{bmatrix} 1 \\ 5 \end{bmatrix}} = \begin{bmatrix} \boxed{16} \\ 4 \\ 7 \end{bmatrix}$$

$$y_i = \sum_{j=1}^{n} A_{ij} \cdot x_j\,, \qquad i = 1, \dots, m$$

$$\begin{bmatrix} 1 & 3 \\ \boxed{4 & 0} \\ 2 & 1 \end{bmatrix} \times \boxed{\begin{bmatrix} 1 \\ 5 \end{bmatrix}} = \begin{bmatrix} 16 \\ \boxed{4} \\ 7 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 3 \\ 4 & 0 \\ 2 & 1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 1\cdot 1 + 3\cdot 5 \\ 4\cdot 1 + 0\cdot 5 \\ 2\cdot 1 + 1\cdot 5 \end{bmatrix} = \begin{bmatrix} 16 \\ 4 \\ 7 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 3 \\ 4 & 0 \\ \boxed{2 & 1} \end{bmatrix} \times \boxed{\begin{bmatrix} 1 \\ 5 \end{bmatrix}} = \begin{bmatrix} 16 \\ 4 \\ \boxed{7} \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 1 & 5 \\ 0 & 3 & 0 & 4 \\ -3 & -2 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 1 \\ 3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1\cdot 1 + 2\cdot 3 + 1\cdot 2 + 5\cdot 1 \\ 0\cdot 1 + 3\cdot 3 + 0\cdot 2 + 4\cdot 1 \\ -1\cdot 1 + (-2)\cdot 3 + 0\cdot 2 + 0\cdot 1 \end{bmatrix} = \begin{bmatrix} 14 \\ 13 \\ -7 \end{bmatrix}$$

Материал прошлых лекций

Умножение матрицы на матрицу $A^{m\times n} \times B^{n\times o} = C^{m\times o}$

$$C_{ij} = \sum_{k=1}^{n} A_{ik} \cdot B_{kj}\,, \qquad i = 1, \dots, m, \qquad j = 1, \dots, o.$$

$$\begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 3 \\ 0 & 1 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} 1\cdot 1 + 3\cdot 0 + 2\cdot 5 & 1\cdot 3 + 3\cdot 1 + 2\cdot 2 \\ 4\cdot 1 + 0\cdot 0 + 1\cdot 5 & 4\cdot 3 + 0\cdot 1 + 1\cdot 2 \end{bmatrix} = \begin{bmatrix} 11 & 10 \\ 9 & 14 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 3 \\ 0 & 1 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} 11 & 10 \\ 9 & 14 \end{bmatrix} \qquad \begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 3 \\ 0 & 1 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} 11 & 10 \\ 9 & 14 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 3 \\ 0 & 1 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} 11 & 10 \\ 9 & 14 \end{bmatrix} \qquad \begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 3 \\ 0 & 1 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} 11 & 10 \\ 9 & 14 \end{bmatrix}$$

Материал прошлых лекций

Умножение матрицы на матрицу не является коммутативным!

$$A \times B \neq B \times A$$

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 2 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \\ 2 & 1 & 0 \\ 3 & 5 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 3 \\ 0 & 1 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} 11 & 10 \\ 9 & 14 \\ 2 & 7 \\ 8 & 16 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 3 \\ 0 & 1 \\ 5 & 2 \end{bmatrix} \times \begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \\ 2 & 1 & 0 \\ 3 & 5 & 1 \end{bmatrix} = error$$

Материал прошлых лекций

Умножение матрицы на матрицу **является ассоциативным**

$$A \times (B \times C) = (A \times B) \times C.$$

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix} \times \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} = ?$$

Вариант 1: $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}; \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 0 & 0 \end{bmatrix}$

Вариант 2: $\begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix} \times \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 4 & 2 \end{bmatrix}; \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} 0 & 0 \\ 4 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 0 & 0 \end{bmatrix}$

Материал прошлых лекций

Единичная матрица является такой матрицей, умножение которой на любую другую матрицу дает в результате вторую матрицу без изменений:

$$I \times A = A \times I = A$$

Аналог среди действительных чисел $1 \cdot z = z \cdot 1 = z$

$I^{m \times m}$

$$I^{2\times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$I^{3\times 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$I^{1\times 1} = [1]$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}^{4\times 4}$$

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1\cdot 1 + 2\cdot 0 & 1\cdot 0 + 2\cdot 1 \\ 3\cdot 1 + 4\cdot 0 & 3\cdot 0 + 4\cdot 1 \\ 5\cdot 1 + 6\cdot 0 & 5\cdot 0 + 6\cdot 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

Основы линейной алгебры

Материал прошлых лекций

Операция транспонирования матрицы A^T

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 5 & 9 \end{bmatrix} \qquad A^T = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 5 & 9 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 \\ 2 & 5 \\ 0 & 9 \end{bmatrix}$$

Обратная матрица – это такая матрица, при умножении на которую в результате получается единичная матрица:

$$A \times A^{-1} = A^{-1} \times A = I$$

$$A = \begin{bmatrix} 3 & 4 \\ 2 & 16 \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} 0.4 & -0.1 \\ -0.05 & 0.075 \end{bmatrix}$$

$$AA^{-1} = \begin{bmatrix} 3 & 4 \\ 2 & 16 \end{bmatrix} \begin{bmatrix} 0.4 & -0.1 \\ -0.05 & 0.075 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Вырожденные матрицы не имеют обратной! Пример:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Задача предсказания цен на недвижимость

Площадь, м²	Цена, тыс. м.к.
2104	460
1416	232
1534	315
852	178
…	…

$$X = \begin{bmatrix} 1 & 2104 \\ 1 & 1416 \\ 1 & 1534 \\ 1 & 852 \end{bmatrix}$$

y

Гипотеза

$$h_\theta(x) = -40 + 0.25x$$

$$H = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} -40 \\ 0.25 \end{bmatrix}$$

$h_\theta(x)$

Рассчитываем значения $h_\theta(x)$ для всей обучающей выборки за одну операцию умножения матрицы на вектор

$$\begin{bmatrix} 1 & 2104 \\ 1 & 1416 \\ 1 & 1534 \\ 1 & 852 \end{bmatrix} \times \begin{bmatrix} -40 \\ 0.25 \end{bmatrix} = \begin{bmatrix} 486 \\ 314 \\ 344 \\ 173 \end{bmatrix}$$

Задача предсказания цен на недвижимость

Площадь, м²	Цена, тыс. м.к.
2104	460
1416	232
1534	315
852	178
…	…

Гипотезы

$$h_1(x) = -40 + 0.25x$$
$$h_2(x) = 200 + 0.1x$$
$$h_3(x) = -150 + 0.4x$$

$$X = \begin{bmatrix} 1 & 2104 \\ 1 & 1416 \\ 1 & 1534 \\ 1 & 852 \end{bmatrix}$$

$$H = \begin{bmatrix} -40 & 200 & -150 \\ 0.25 & 0.1 & 0.4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2104 \\ 1 & 1416 \\ 1 & 1534 \\ 1 & 852 \end{bmatrix} \times \begin{bmatrix} -40 & 200 & -150 \\ 0.25 & 0.1 & 0.4 \end{bmatrix} = \begin{array}{c} \begin{bmatrix} 486 & 410 & 692 \\ 314 & 342 & 416 \\ 344 & 353 & 464 \\ 173 & 285 & 191 \end{bmatrix} \begin{array}{l} \leftarrow x_1 \\ \leftarrow x_2 \\ \leftarrow x_3 \\ \leftarrow x_4 \end{array} \\ \begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ h_1(x) & h_2(x) & h_3(x) \end{array} \end{array}$$

Линейная регрессия нескольких переменных

Задача предсказания цен на недвижимость

Размер, м^2	Комнат	Этажей	Возраст дома	Цена
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178
…	…	…	…	…
x_1	x_2	x_3	x_4	y

m=47

n=4

$x_3^{(2)} = 2$

$$x^{(2)} = \begin{bmatrix} 1416 \\ 3 \\ 2 \\ 40 \end{bmatrix} \qquad x^{(i)} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

$x_j^{(i)}$ – j-й признак вектора признаков i-го элемента обучающей выборки

Задача предсказания цен на недвижимость

Размер, $м^2$	Комнат	Этажей	Возраст дома	Цена
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178
…	…	…	…	…

Гипотеза:

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \cdots + \theta_n x_n$$

$$h_\theta(x) = 80 + 0.1x_1 + 15x_2 - 0.5x_3 - 2x_4$$

Каждый дом имеет базовую стоимость 80, которая увеличивается в зависимости от площади (с коэффициентом 0.1), увеличивается от числа комнат (с коэффициентом 15), уменьшается с количеством этажей (по 0.5 за этаж) и уменьшается с возрастом дома (с коэффициентом 2)

Запись в матричной форме

$$x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \in \mathbb{R}^{n+1}$$

$$x_0^{(i)} = 1$$

Гипотеза:

$$\boxed{h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4}$$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix} \in \mathbb{R}^{n+1}$$

Гипотеза в матричной форме:

$$\boxed{h_\theta(x) = \theta^T x}$$

$$h_\theta(x) = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix}^T \times \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \theta_0 & \theta_1 & \theta_2 & \theta_3 & \theta_4 \end{bmatrix} \times \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h(x^{(i)}) - y^{(i)}\right)^2$$ – *функция стоимости*

$J(\theta)$ является функцией от $n+1$ переменных

Для ее минимизации можно применить метод градиентного спуска

Вычислить частные производные

$$\frac{\partial}{\partial \theta_j} J(\theta_0, \dots, \theta_n), \quad j = 0, \dots, n$$

Частные производные по всем параметрам $\theta_0, \dots, \theta_n$

$$\left\{\begin{aligned} &\frac{\partial}{\partial\theta_0}J(\theta) = \frac{1}{m}\sum_{i=0}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \\ &\frac{\partial}{\partial\theta_1}J(\theta) = \frac{1}{m}\sum_{i=0}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)x_1^{(i)} \\ &\dots \\ &\frac{\partial}{\partial\theta_n}J(\theta) = \frac{1}{m}\sum_{i=0}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)x_n^{(i)} \end{aligned}\right.$$

$$\frac{\partial}{\partial\theta_j}J(\theta) = \frac{1}{m}\sum_{i=0}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)x_j^{(i)}, \qquad j = 0, \dots, n$$

Алгоритм градиентного спуска.

Вход: $J(\theta_0 \; ,\ldots,\theta_n \;)$ – функция, $\theta_0^0,\ldots,\theta_n^0$– начальные значения переменных, α – темп обучения, ε – критерий остановки.

Выход: $\theta_0^*,\ldots,\theta_n^*$– найденные значения для локального минимума.

Действия:

для $j = 0, \ldots, n$:

$\quad\theta_j := \theta_j^0$

повторять {

$\quad$для $j = 0, \ldots, n$:

$$d\theta_j := \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)x^{(i)}$$

$\quad$для $j = 0, \ldots, n$:

$$\theta_j := \theta_j - \alpha\, d\theta_j$$

} пока $d\theta_j > \varepsilon,\ j = 0, \ldots, n$

для $j = 0, \ldots, n$:

$\quad\theta_j^* := \theta_j$

Конец алгоритма.

Линейная регрессия нескольких переменных

Задана функция стоимости для множественной линейной регрессии в виде:

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h(x^{(i)}) - y^{(i)}\right)^2$$

Какие из выражений ниже являются эквивалентными ей?

☑ $$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(\theta^T x^{(i)} - y^{(i)}\right)^2$$

☑ $$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(\left(\sum\nolimits_{j=0}^{n} \theta_j x_j^{(i)}\right) - y^{(i)}\right)^2$$

☒ $$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(\left(\sum\nolimits_{j=1}^{n} \theta_j x_j^{(i)}\right) - y^{(i)}\right)^2$$

☒ $$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(\left(\sum\nolimits_{j=0}^{n} \theta_j x_j^{(i)}\right) - \left(\sum\nolimits_{j=0}^{n} y_j^{(i)}\right)\right)^2$$

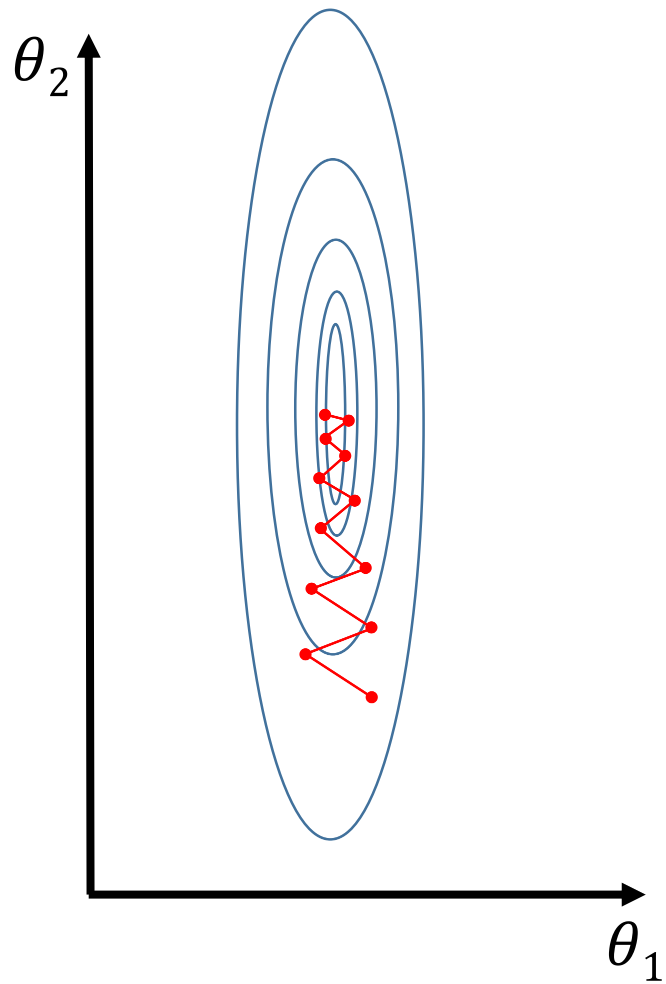
Нормализация данных

Задача предсказания цен на недвижимость

Размер, м^2	Комнат	…	Цена
2104	5	…	460
1416	3	…	232
1534	3	…	315
852	2	…	178
…	…	…	…

$x_1 \in [0, 3000]$ $x_2 \in [1, 5]$

Алгоритм градиентного спуска может не остановиться, поскольку ε для x_1 может быть мало, а для x_2 слишком велико

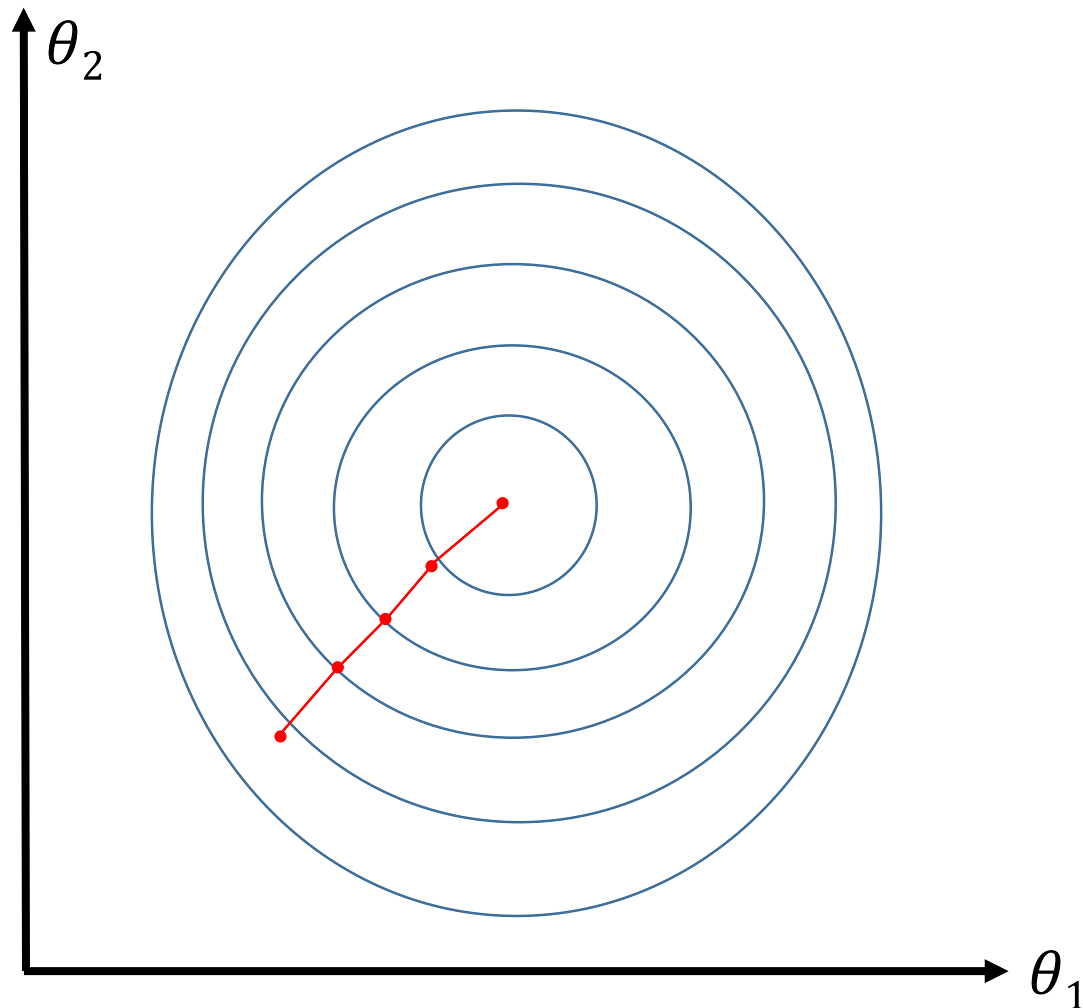


Масштабирование исходных данных

$$x_1 = \frac{\text{площадь}}{3000}$$

$$x_2 = \frac{\textit{кол}-\textit{во комнат}}{5}$$

$$0 \le x_i \le 1$$



Нормализация среднего значения

$$\acute{x}_i = \frac{x_i - \mu_i}{s_i}$$

$$s_i = \max(x_i) - \min(x_i)$$

μ_i – среднее значение переменной

$$-0.5 \le x_i \le 0.5$$

$$x_2 = \frac{кол-во\ комнат - 2}{5}$$

$$x_1 = \frac{\text{площадь} - 1500}{3000}$$

$$\mathbf{-1 \le x_i \le 1}$$

$$\mathbf{0 \le x_i \le 3}$$

$$\mathbf{-2 \le x_i \le 0.9}$$

$$\mathbf{-10 \le x_i \le 200}$$

$$\mathbf{-0.001 \le x_i \le 0.00001}$$

Нормализация данных

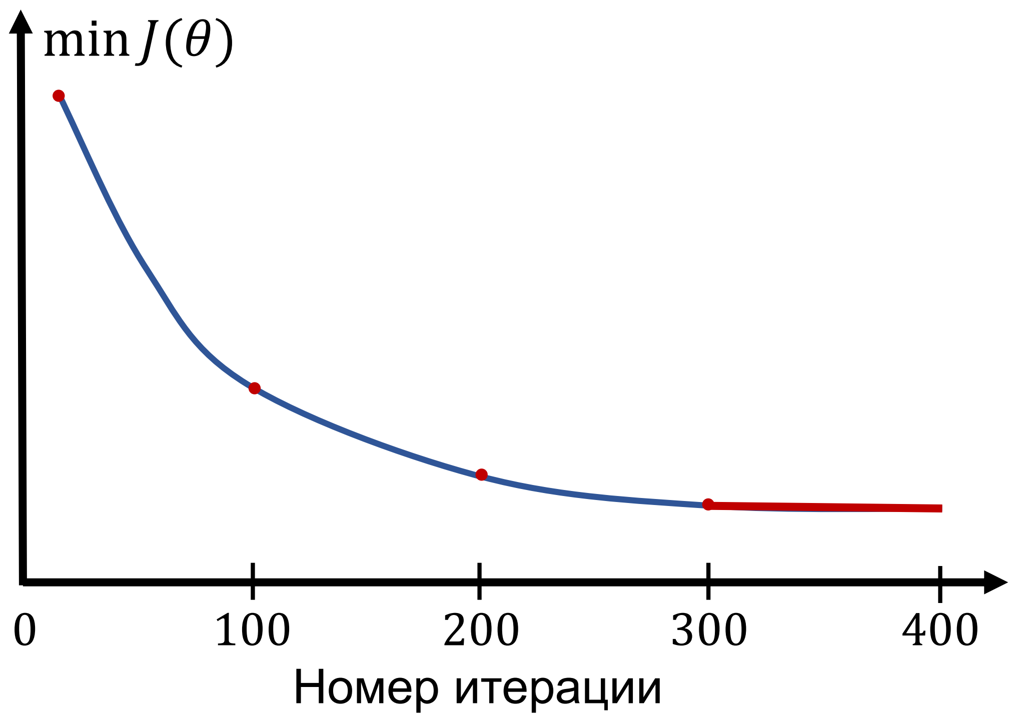
Используется алгоритм обучения для задачи предсказания стоимости дома. Одна из переменных в обучающей выборке означает возраст дома и принимает значения от 30 до 50 со средним значением 38. Какое из следующих выражений следует использовать для нормализации этой переменной?

- ☒ $x_{\mathrm{i}} = \text{возраст дома}$
- ☒ $x_{\mathrm{i}} = \dfrac{\text{возраст дома}}{50}$
- ☒ $x_{\mathrm{i}} = \dfrac{\text{возраст дома} - 38}{50}$
- ☑ $x_{\mathrm{i}} = \dfrac{\text{возраст дома} - 38}{20}$

Выбор темпа обучения

Слишком малый α приводит к долгому времени сходимости алгоритма
Слишком большое значение α может привести к тому, что алгоритм не закончит работу

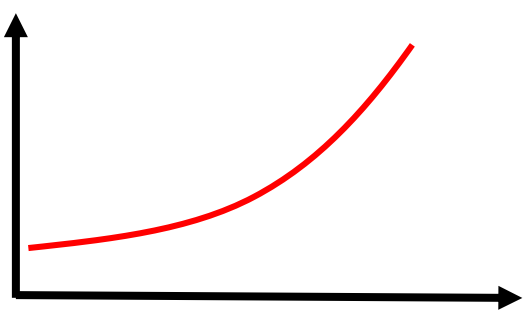
Проверка правильности выбора значения α

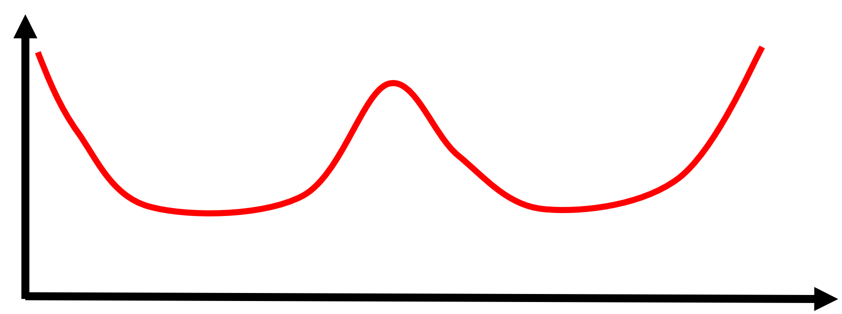


Значение $J(\theta)$ <u>должно уменьшаться</u> на каждом шаге

$\alpha = \ldots, 0.0001, 0.001, 0.01, 0.1, \ldots$

Выбираем значение α немного меньше, чем наибольшее из «правильных»



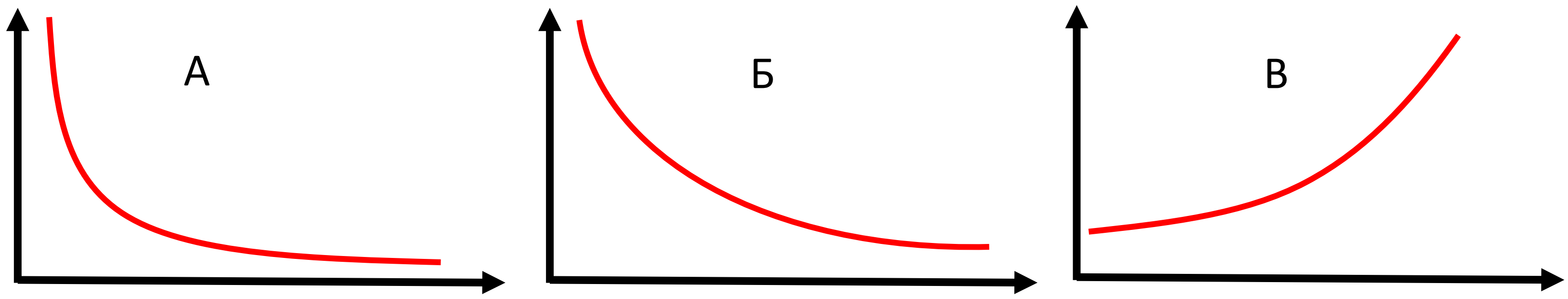


Выбор темпа обучения

Предположим, что алгоритм градиентного спуска был запущен три раза со значениями:

$\alpha=0.01$ $\alpha=0.1$ $\alpha=1$

Определите, какому из графиков ниже (А, Б и В) соответствует какое из значений α.



- [ ] А для $\alpha=0.01$, Б для $\alpha=0.1$, В для $\alpha=1$
- [x] А для $\alpha=0.1$, Б для $\alpha=0.01$, В для $\alpha=1$
- [ ] А для $\alpha=1$, Б для $\alpha=0.01$, В для $\alpha=0.1$
- [ ] А для $\alpha=1$, Б для $\alpha=0.1$, В для $\alpha=0.01$

Выбор и конструирование переменных

Задача предсказания стоимости дачного участка

Переменные:
1. длина участка вдоль улицы (frontage)
2. глубина от улицы (depth)

Первый вариант решения – построить линейную регрессионную модель:

$$h_\theta(x) = \theta_0 + \theta_1 \cdot frontage + \theta_2 \cdot depth$$

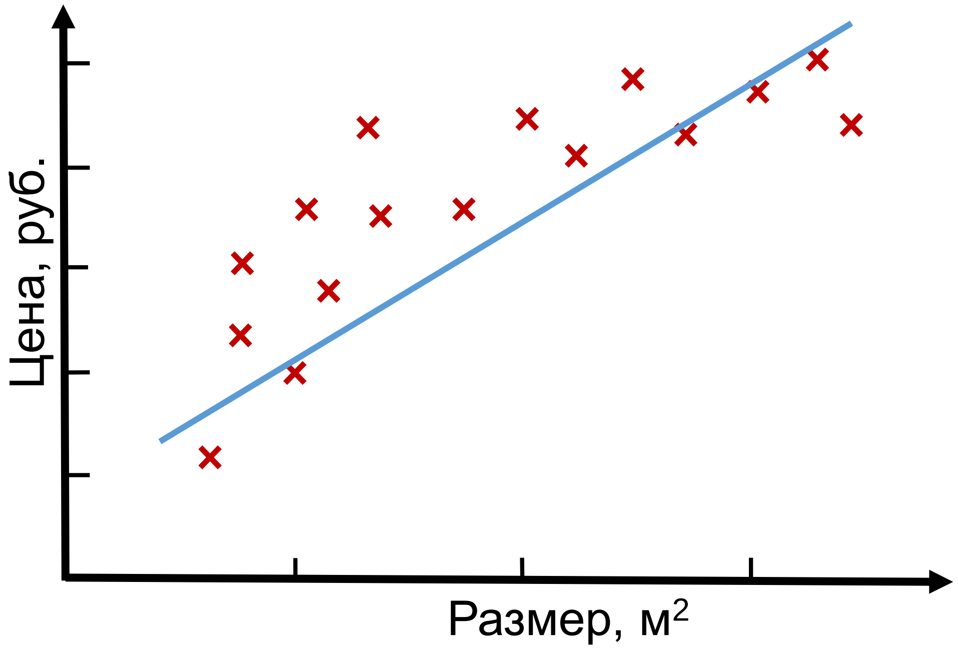
Проще!

Второй вариант решения – сконструировать новую переменную

$$x = frontage \cdot depth$$

$$h_\theta(x) = \theta_0 + \theta_1 x$$

Задача предсказания стоимости дома
(конструируем переменные для усложнения модели)



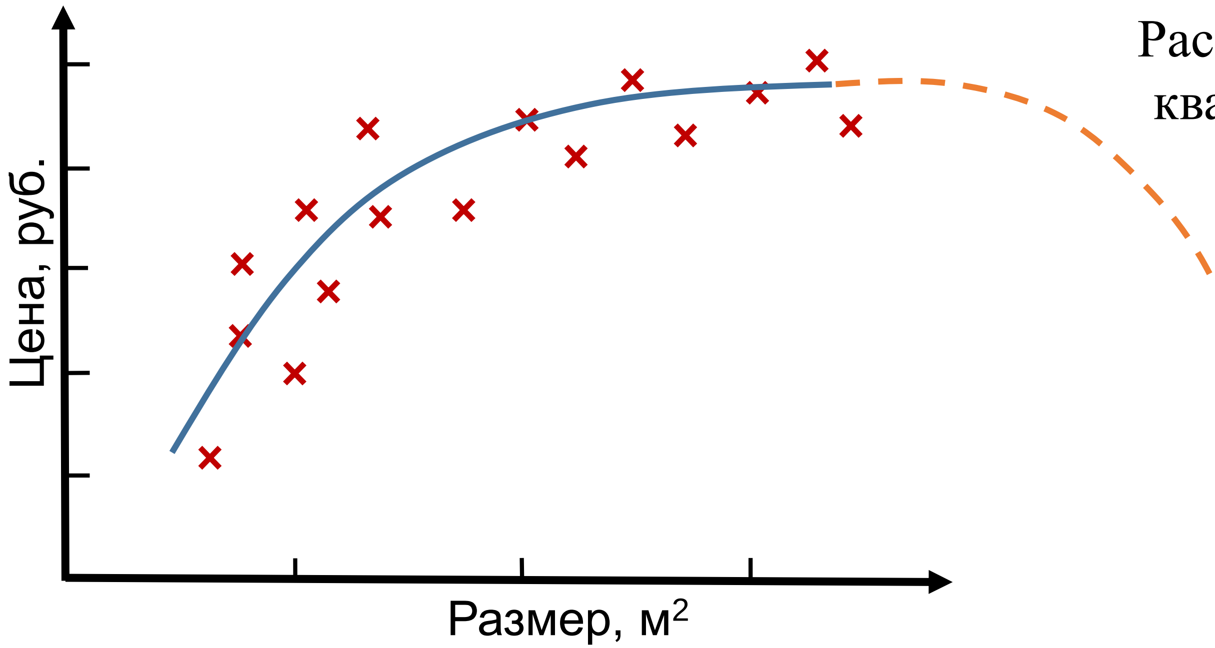
x – размер дома (м^2)

y – цена (руб)

Простейшее решение:

$$h_\theta(x) = \theta_0 + \theta_1 x$$

Выбор и конструирование переменных



Распределение похоже на квадратичную функцию

$$x_1 = (\text{размер})$$

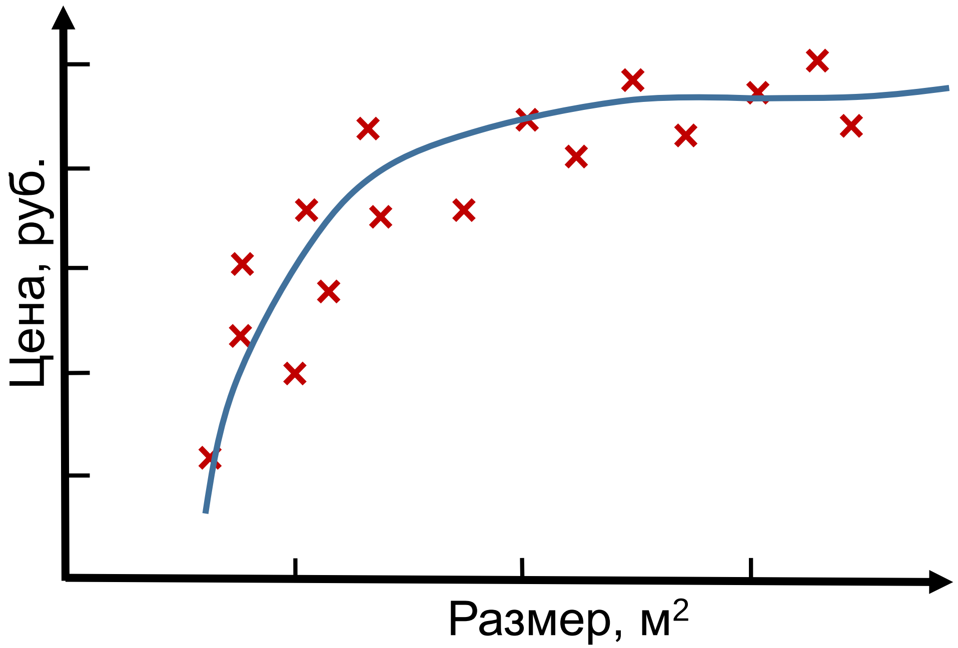
$$x_2 = (\text{размер})^2$$

Регрессия по двум переменным:

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

Решается градиентным спуском!

Выбор и конструирование переменных



Можно избавиться от недостатка квадратичной функции использую кубическую

$$x_1 = (\text{размер})$$

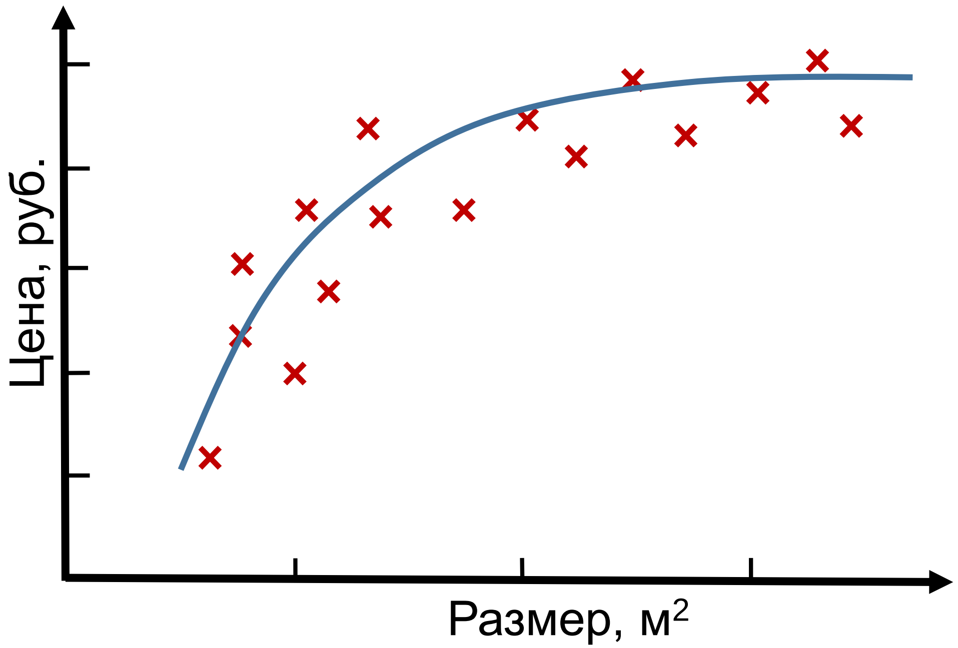
$$x_2 = (\text{размер})^2$$

$$x_3 = (\text{размер})^3$$

Регрессия по трем переменным:

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_2 x_3$$

Выбор и конструирование переменных



Можно использовать практически любые функции для конструирования новых переменных

$$x_1 = (\text{размер})$$

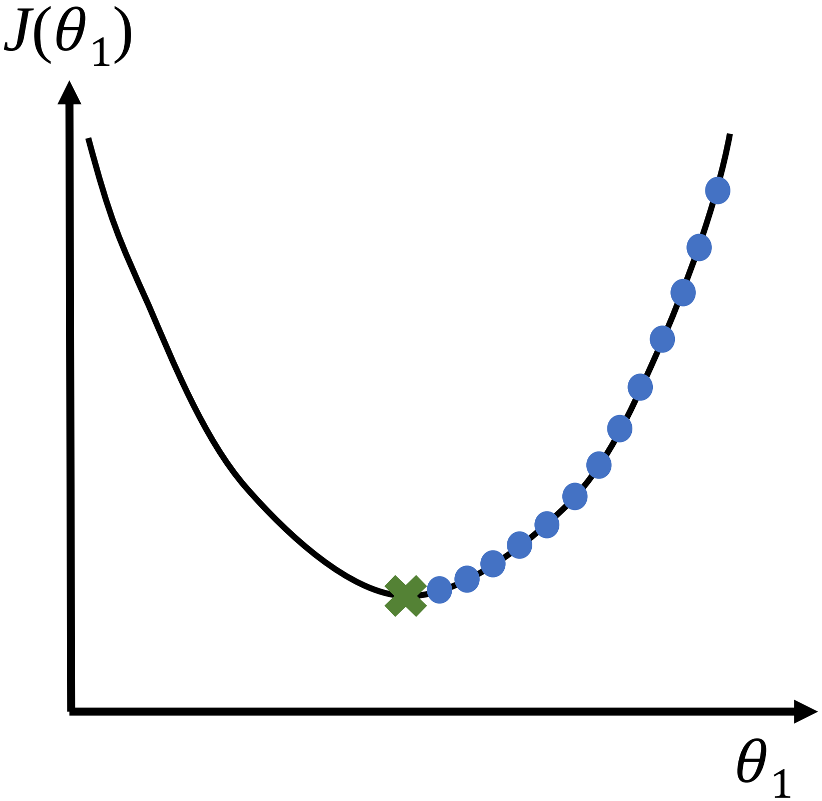
$$x_2 = \sqrt{(\text{размер})}$$

Регрессия по двум переменным:

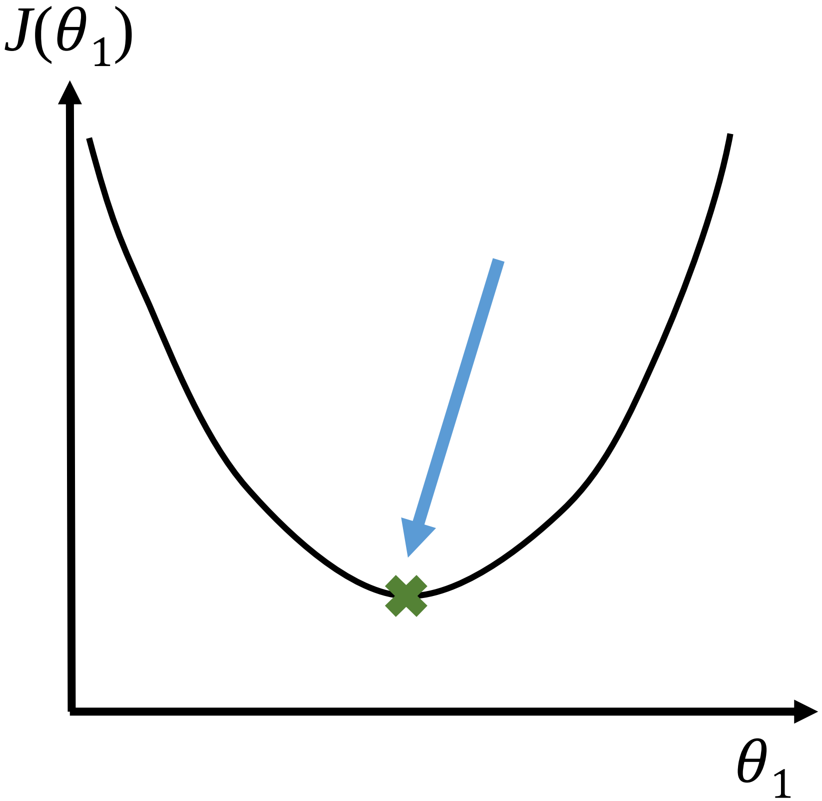
$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 = \theta_0 + \theta_1(\text{размер}) + \theta_2\sqrt{(\text{размер})}$$

Получаем модель без недостатков квадратичной, но уже с двумя переменными

Градиентный спуск
Последовательное приближение к минимуму за множество итераций



Нормальные уравнения
Аналитическое вычисление оптимальных значений параметров θ за один шаг

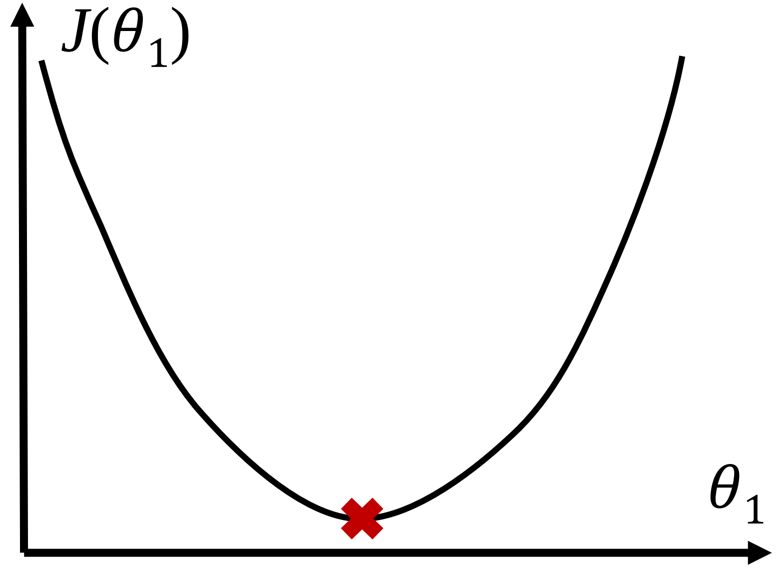


Нормальные уравнения

Функция стоимости линейной регрессии

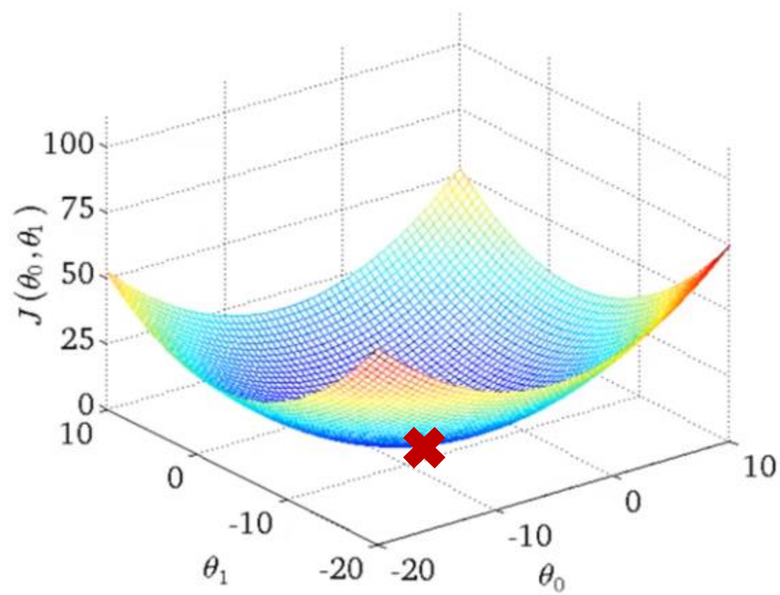
$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x_i) - y_i)^2$$

(квадратичная функция)



Условие минимума функции $J(\theta)$

$$\frac{\partial}{\partial\theta}J(\theta) = 0$$



Решение уравнения дает значение θ^*, в котором достигается минимум $J(\theta)$ в случае одного параметра θ

Нормальные уравнения

Гипотеза

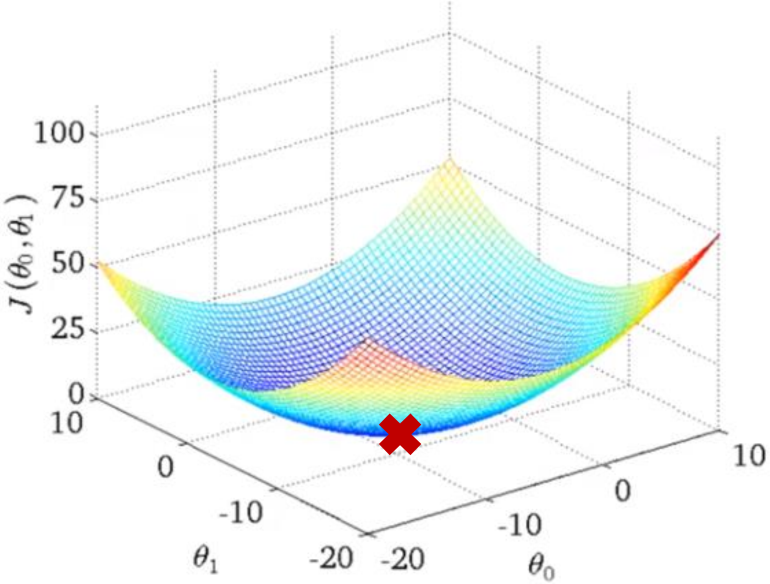
$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \cdots + \theta_n x_n$$

Функция стоимости

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x_i) - y_i)^2$$

Условие минимума функции $J(\theta)$

$$\frac{\partial}{\partial\theta}J(\theta) = 0 \qquad \theta \in \mathbb{R}^{n+1}$$



$$\begin{cases} \dfrac{\partial}{\partial\theta_0}J(\theta_0, \dots, \theta_n) = 0 \\ \dfrac{\partial}{\partial\theta_1}J(\theta_0, \dots, \theta_n) = 0 \\ \dots \\ \dfrac{\partial}{\partial\theta_n}J(\theta_0, \dots, \theta_n) = 0 \end{cases} \Rightarrow \boxed{\theta = (X^T X)^{-1} X^T y}$$

$X \in \mathbb{R}^{m \times n+1}$

$\theta \in \mathbb{R}^{n+1}$

$y \in \mathbb{R}^m$

Нормальные уравнения

Задача предсказания цен на недвижимость

Размер, м^2	Комнат	Этажей	Возраст дома	Цена
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178
…	…	…	…	…

m

1

n=4

$$X = \begin{bmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1534 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 36 \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} \qquad y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \\ \dots \end{bmatrix}$$

Просто!

Вычисление $(\mathbb{R}^{n+1\times n+1})^{-1}$ очень долго для больших n

$(n \gg 1000)$

$O(n^3)$

$$\theta = (X^T X)^{-1} X^T y$$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \dots \\ \theta_n \end{bmatrix}$$

Градиентный спуск	Нормальные уравнения
➢ Нужно выбирать приемлемое значение α	➢ Нет необходимости выбирать значение α
➢ Требуется множество итераций	➢ Нет множества итераций
➢ Требуется нормализация параметров	➢ Не требуется нормализация
➢ Работает устойчиво даже при больших n	➢ Требуется вычисление обратной матрицы $(\mathbb{R}^{n+1\times n+1})^{-1}$ а) очень медленно при больших n б) некоторые матрицы бывают необратимыми (не вычисляется обратная)

Причины необратимости $X^T X$

1) Избыточные переменные

$x_1 =$ (размер в метрах)

~~$x_2 =$ (размер в сантиметрах)~~

~~$x_1 = 100 \cdot x_2$ — линейная зависимость~~

✓ Избавиться от избыточных переменных (устранить, линейную зависимость между переменными)

2) Слишком мало данных

$m < n$ Эквивалентно решению системы из m уравнений с n неизвестными

✓ Убрать малозначимые переменные, добиться, чтобы $m \geq n$

✓ Использовать ***регуляризацию*** (будем изучать позже)

Нормальные уравнения

Имеем обучающую выборку из m=23 промеров с n=5 переменными. Решаем задачу регрессии с использованием нормального уравнения $\theta = (X^T X)^{-1} X^T y$.

Какие будут размерности матриц и векторов θ, X, y?

- ☒ $X \in \mathbb{R}^{23\times5}$ $y \in \mathbb{R}^{23\times1}$ $\theta \in \mathbb{R}^{5\times1}$
- ☑ $X \in \mathbb{R}^{23\times6}$ $y \in \mathbb{R}^{23\times1}$ $\theta \in \mathbb{R}^{6\times1}$
- ☒ $X \in \mathbb{R}^{23\times5}$ $y \in \mathbb{R}^{23\times1}$ $\theta \in \mathbb{R}^{5\times5}$
- ☒ $X \in \mathbb{R}^{23\times6}$ $y \in \mathbb{R}^{23\times6}$ $\theta \in \mathbb{R}^{6\times6}$

Имеем обучающую выборку из m=50 промеров с n=15 переменными. Необходимо решить задачу линейной регрессии от нескольких переменных.
Что лучше выбрать: алгоритм градиентного спуска или нормальные уравнения и почему?

- ☑ Нормальные уравнения, поскольку предоставляют эффективный способ получить решение сразу
- ☒ Градиентный спуск, поскольку он всегда сходится к оптимальному значению θ
- ☒ Нормальные уравнения, поскольку градиентный спуск может не найти оптимальное решение
- ☒ Градиентный спуск, потому что $(X^T X)^{-1}$ будет вычисляться очень долго