Обучение на основе наблюдений. Линейная регрессия как задача контролируемого (индуктивного) обучения.

Лекция 4

Рычагов М.Н.,
профессор, д.ф.-м.н.

Регрессия с одной переменной

➢ **Задача о стоимости квартиры**

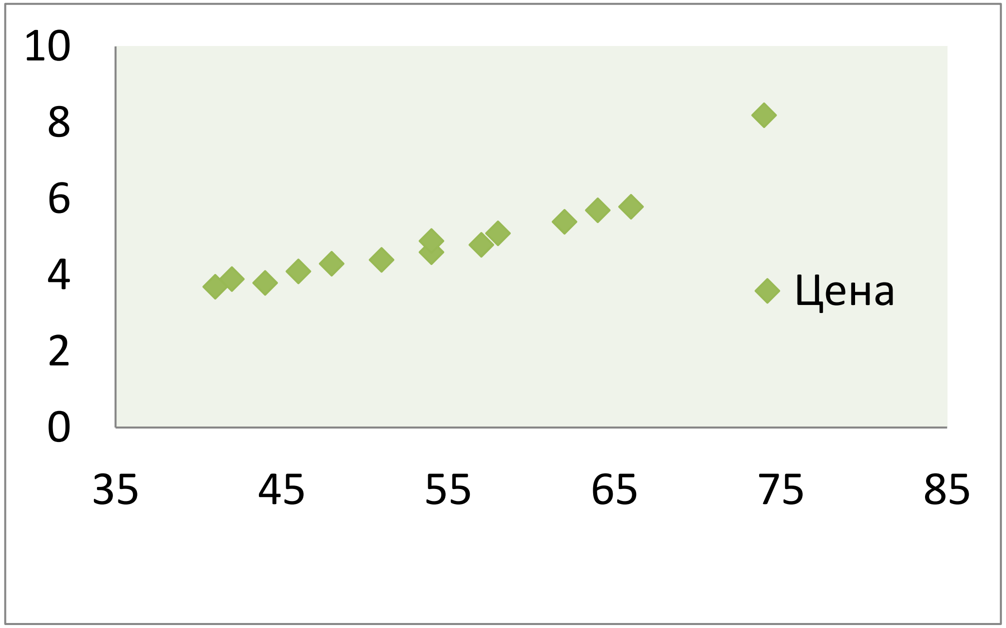
Площадь (м²) x	Цена (млн. руб) y
45	4,5
63	6,7
52	7,2
48	5,8
…	…

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

Управляемое обучение

Стоимость квартиры

Цена
(в 10^6 руб.)



Управляемое обучение

Для каждого входного элемента (примера) существует известное значение на выходе

Задача линейной регрессии

Предсказание выходного сигнала в пространстве вещественных чисел

Формирование обучающего набора

Обучающий набор по оценке стоимости квартир

Площадь, м^2 (x)	**Цена в 10^6 рублей (y)**
44	3,8
48	4,3
54	6,5
74	8,2
…	…

$m = 52$

Определение:

m - количество обучающих примеров (например, 52)

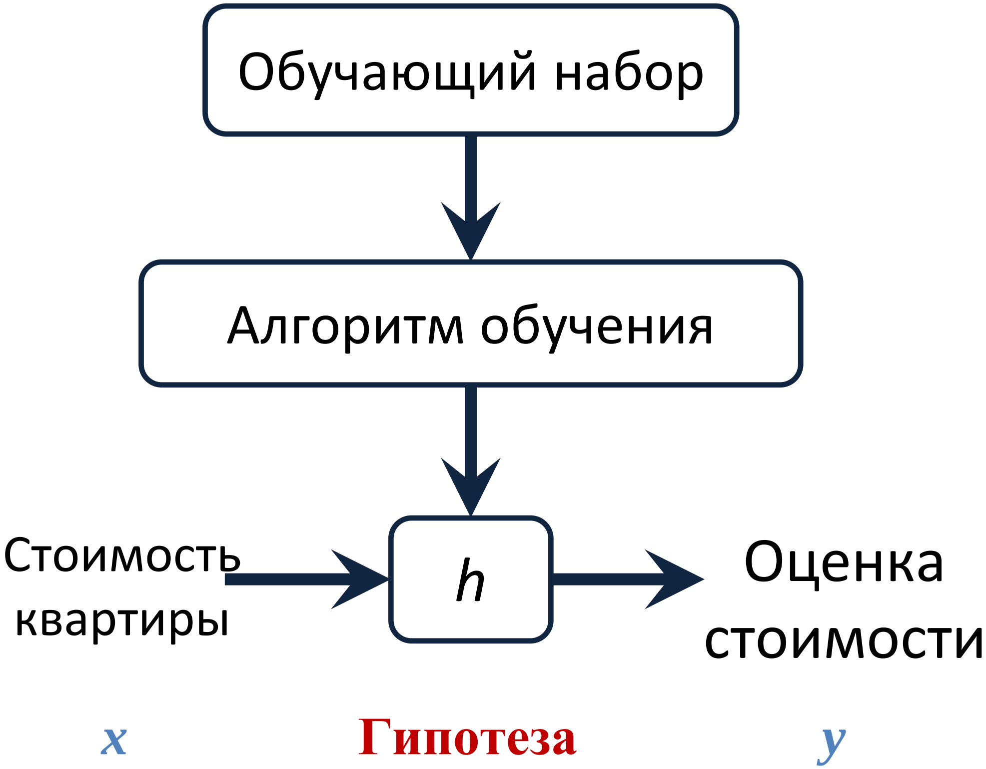
x's - входные значения / переменные

y's - выходные значения / “target” переменные

(x, y) – один пример из обучающего набора

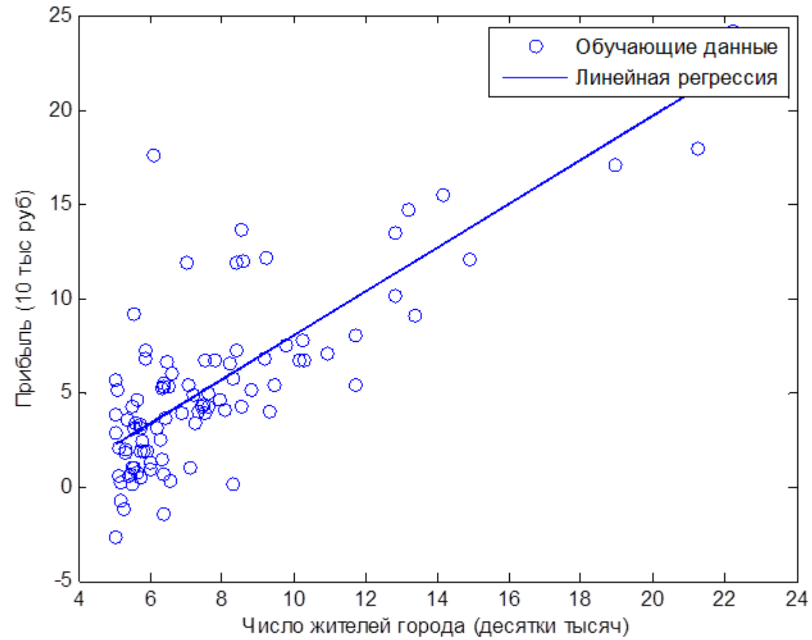
$(x^{(i)}, y^{(i)})$ – i-ый пример из обучающего набора

Представление гипотезы



Как представляется гипотеза *h* ?

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



Линейная регрессия с одной переменной
(на примере Лабораторной работы 1)

Постановка задачи

Обучающий набор

Площадь, м² (x)	**Цена в 10^6 рублей (y)**
44	3,8
48	4,3
54	6,5
74	8,2
…	…

$m = 52$

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

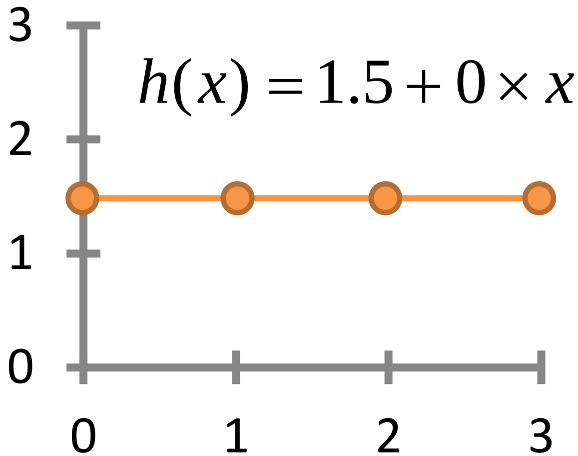
θ_i‘s: Параметры

Как выбрать θ_i ?

Гипотеза при различных значениях параметров

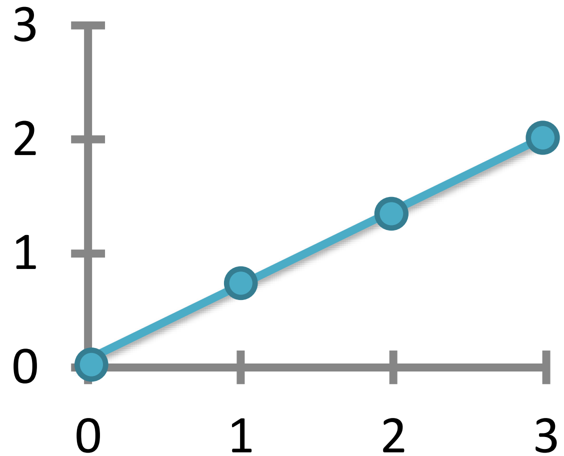
$$h_\theta(x) = \theta_0 + \theta_1 x$$

$h(x) = 1.5 + 0 \times x$



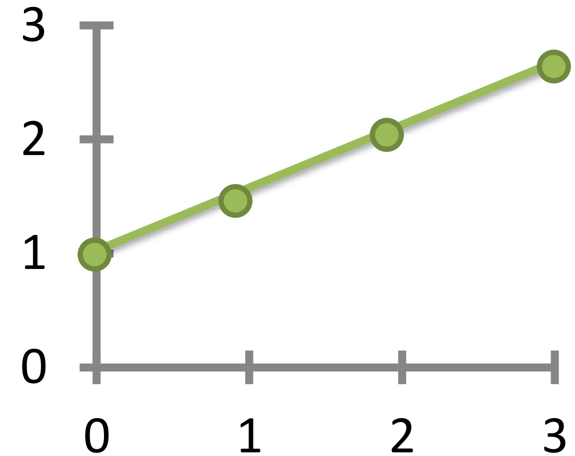
$\theta_0 = 1.5$
$\theta_1 = 0$

$h(x) = 0.5 \times x$



$\theta_0 = 0$
$\theta_1 = 0.5$

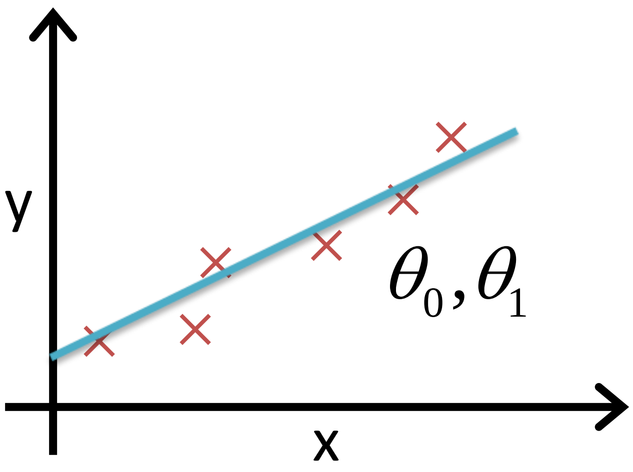
$h(x) = 1.0 + 0.5 \times x$



$\theta_0 = 1$
$\theta_1 = 0.5$

- Вопрос: случай независимых переменных, если гипотеза верна?

Функция стоимости



- Минимизировать по всем *m* относительно θ_0, θ_1
- **Функция стоимости**

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

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

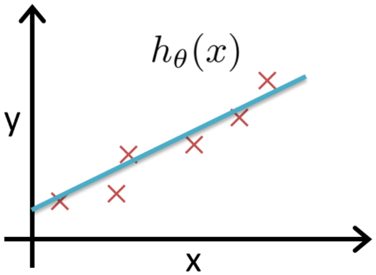
- Идея: выбрать θ_0, θ_1: $h_\theta(x)$ является наиболее близкой к *y* для каждого из обучающих примеров *(x, y)*

Сведение к однопараметрической задаче

Упрощенный случай

Гипотеза:

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



Параметры:

$$\theta_0, \theta_1$$

Функция стоимости:

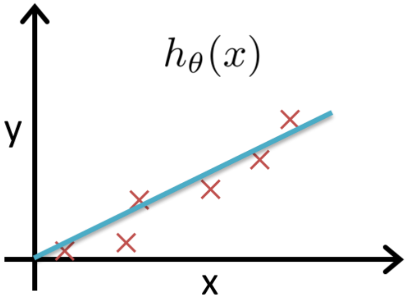
$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

Цель: $\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$

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

θ_1

$\theta_0 = 0$



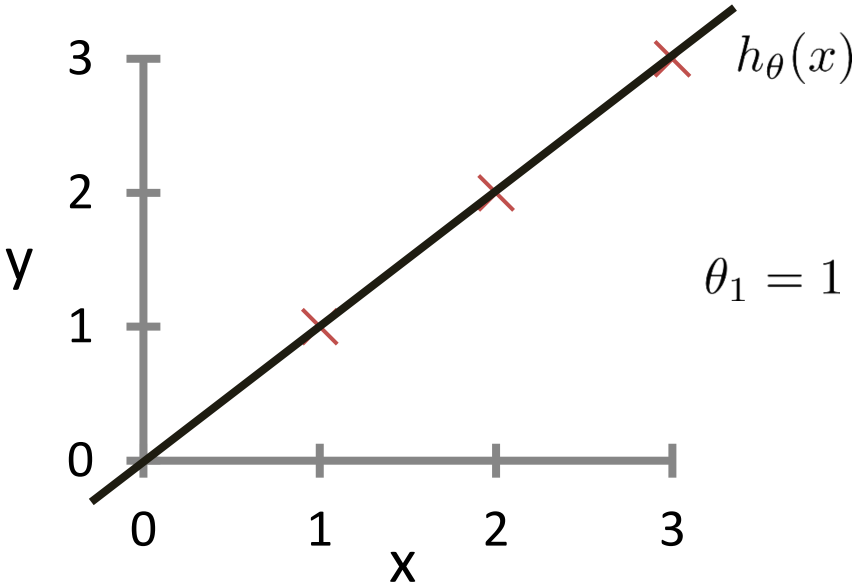
$$J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$\underset{\theta_1}{\text{minimize}}\ J(\theta_1)$$

Распределение $y=x$ и гипотеза $\Theta = (0, 1)$

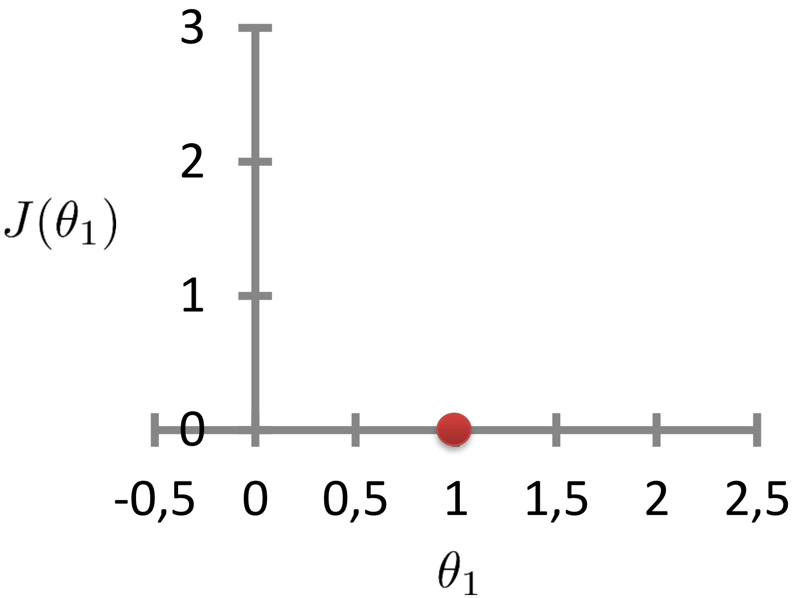
$h_\theta(x)$

(для фиксированного θ_1, это функция x)



$J(\theta_1)$

(функция параметра θ_1)

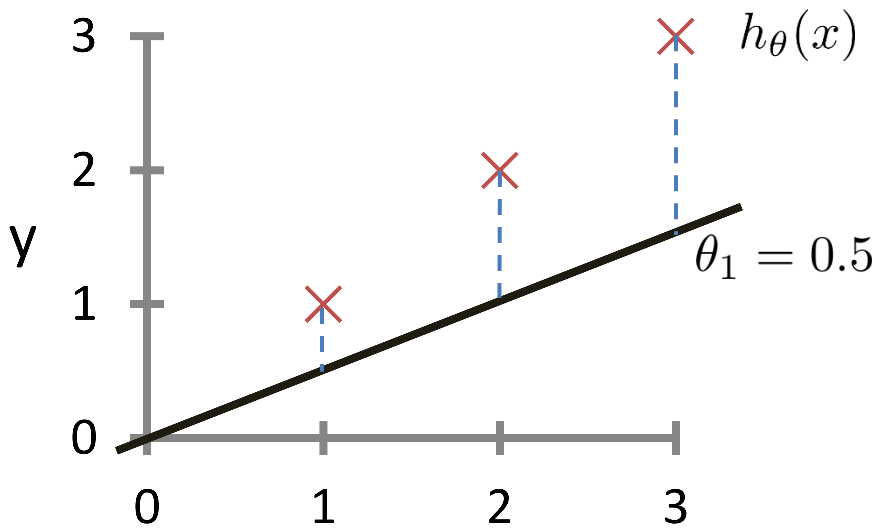


$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 = \frac{1}{2m}(0^2 + 0^2 + 0^2) = 0$$

Распределение $y=x$ и гипотеза $\Theta = (0, 0.5)$

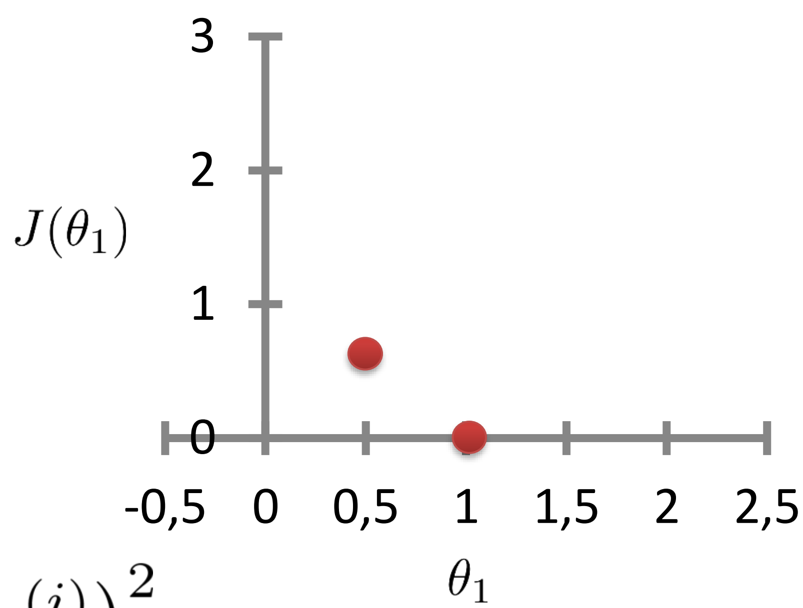
$h_\theta(x)$

(для фиксированного θ_1, это функция x)



$J(\theta_1)$

(функция параметра θ_1)



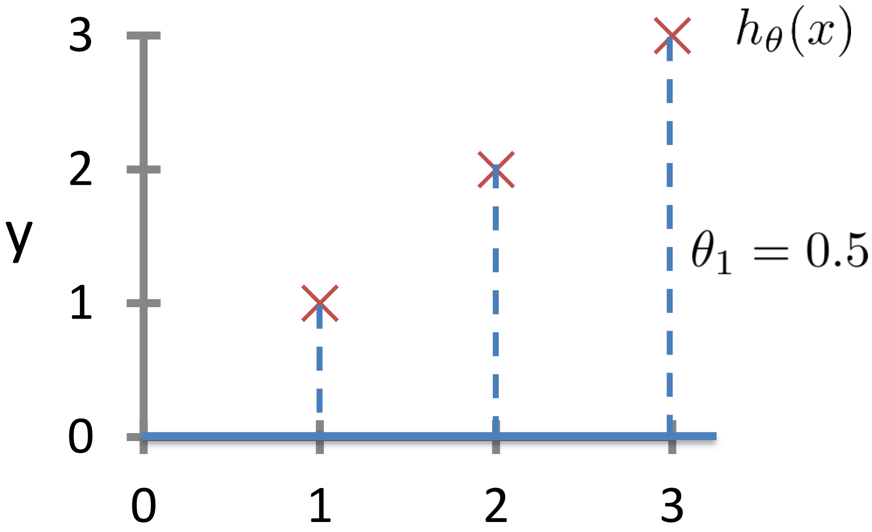
$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$= \frac{1}{2m}(0.5^2 + 1^2 + 1.5^2) = \frac{1}{2 \times 3}(3 \cdot 5) \approx 0.58$$

Распределение y=x и гипотеза $\Theta = (0, 0.0)$

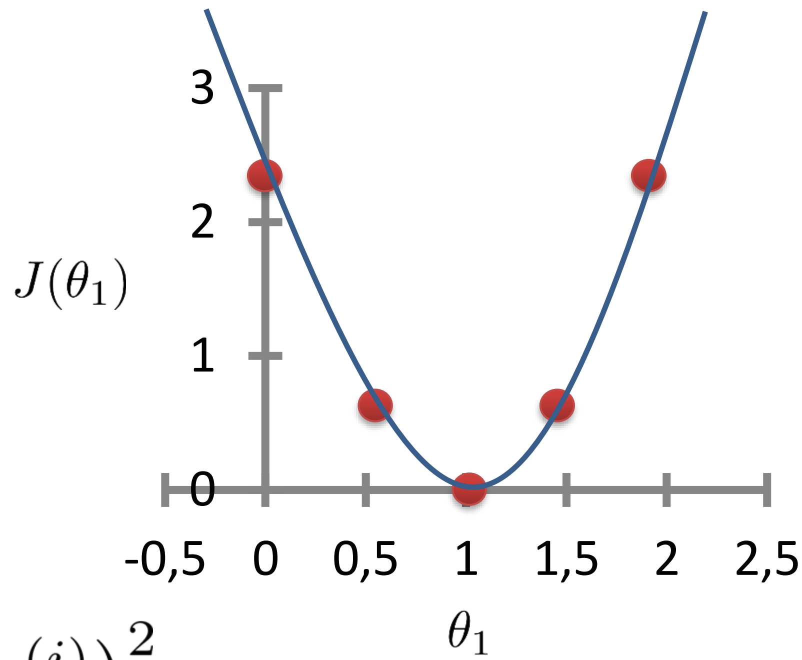
$h_\theta(x)$

(для фиксированного θ_1, это функция x)



$J(\theta_1)$

(функция параметра θ_1)



$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$= \frac{1}{2m}(1.0^2 + 2^2 + 3^2) = \frac{1}{2 \times 3} \times 14 \approx 2.3$$

Постановка задачи для градиентного спуска

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

Параметры: θ_0, θ_1

Ф-я стоимости: $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$

Цель: $\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$

Стратегия градиентного спуска

Имеем некоторую ф-ю стоимости $J(\theta_0, \theta_1)$

Обобщение: $J(\theta_0, \theta_1, \theta_2, \ldots, \theta_n)$

Намерение: $\min\limits_{\theta_0, \theta_1} J(\theta_0, \theta_1)$ или $J(\theta_0, \theta_1, \theta_2, \ldots, \theta_n)$

Метод:

- Начинаем с произвольных θ_0, θ_1
- Итерируем значения θ_0, θ_1, чтобы $\downarrow J(\theta_0, \theta_1)$

  до достижения минимума

Алгоритм градиентного спуска

Повторять итерации до сходимости

{

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \qquad (\text{for } j = 0 \text{ and } j = 1)$$

}

Верно: одновременное присвоение

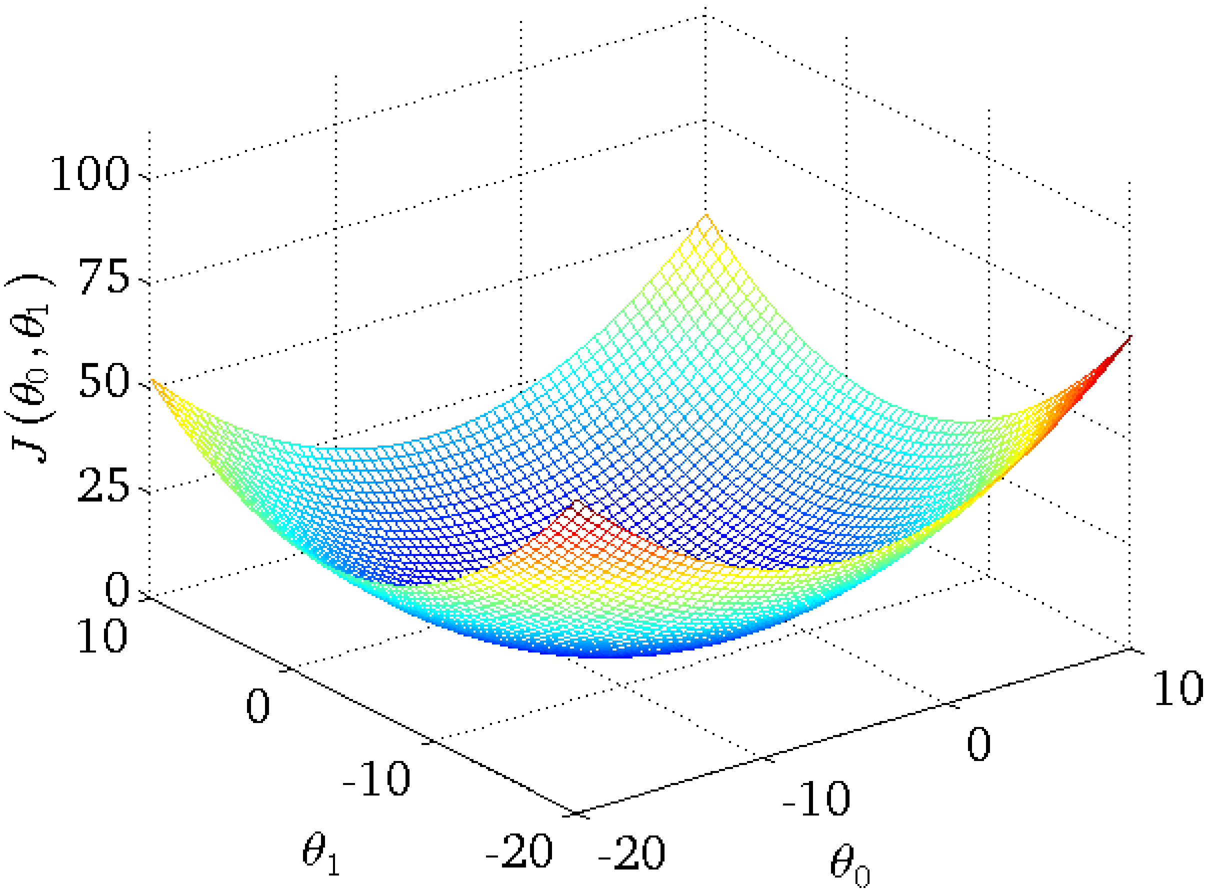
$$\begin{aligned}
&\text{temp0} := \theta_0 - \alpha \tfrac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) \\
&\text{temp1} := \theta_1 - \alpha \tfrac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) \\
&\theta_0 := \text{temp0} \\
&\theta_1 := \text{temp1}
\end{aligned}$$

Неверно:

$$\begin{aligned}
&\text{temp0} := \theta_0 - \alpha \tfrac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) \\
&\theta_0 := \text{temp0} \\
&\text{temp1} := \theta_1 - \alpha \tfrac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) \\
&\theta_1 := \text{temp1}
\end{aligned}$$

Графическое представление функции стоимости

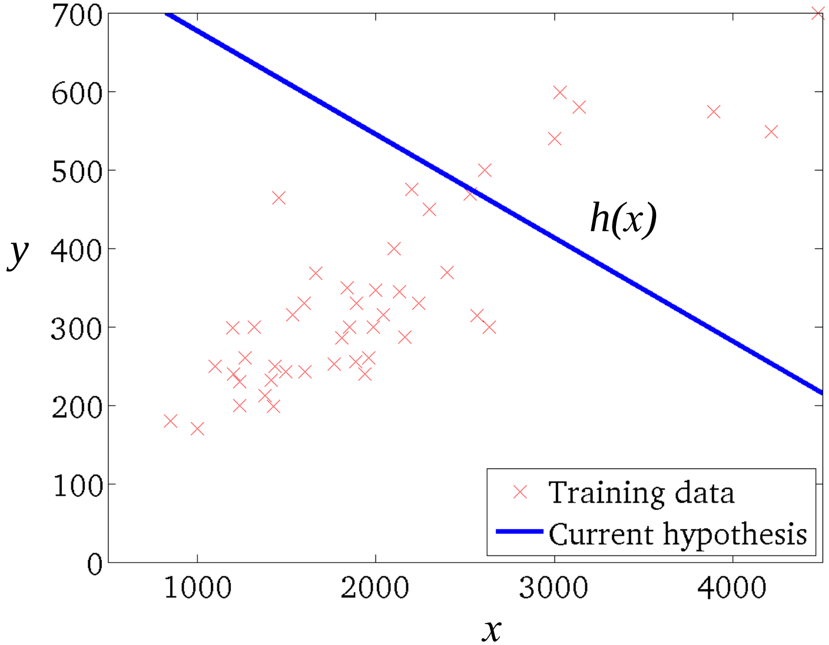
Выпуклая (чашеобразная) функция стоимости



Представление гипотезы и функции стоимости. I

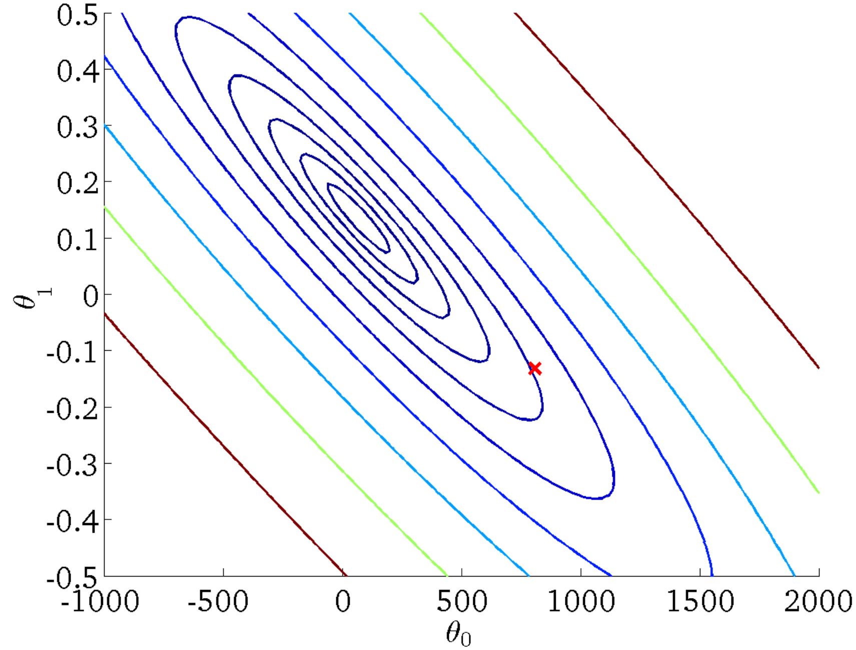
$h_\theta(x)$

(для фиксированных θ_0, θ_1 - это функция x)



$J(\theta_0, \theta_1)$

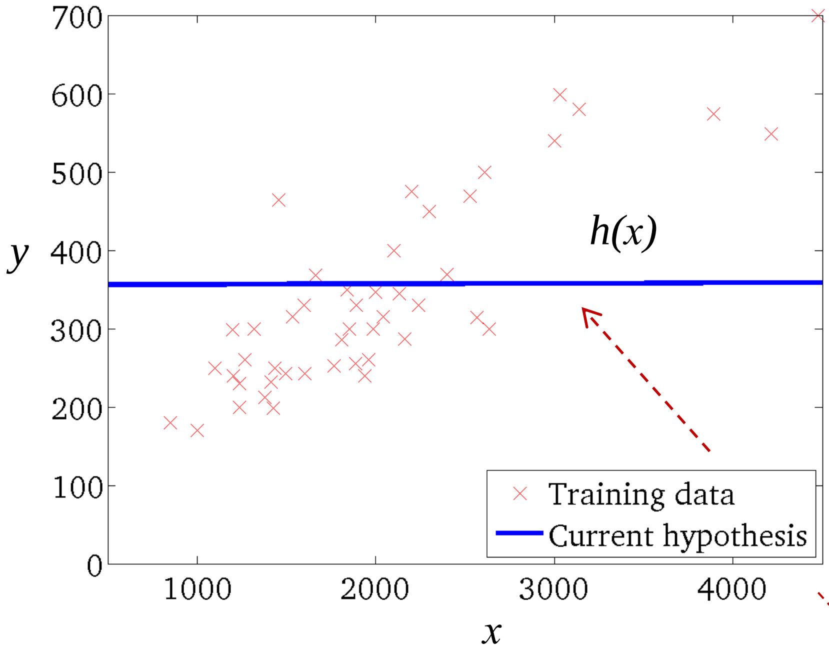
(функция, зависящая от θ_0, θ_1)



Представление гипотезы и функции стоимости. II

$h_\theta(x)$

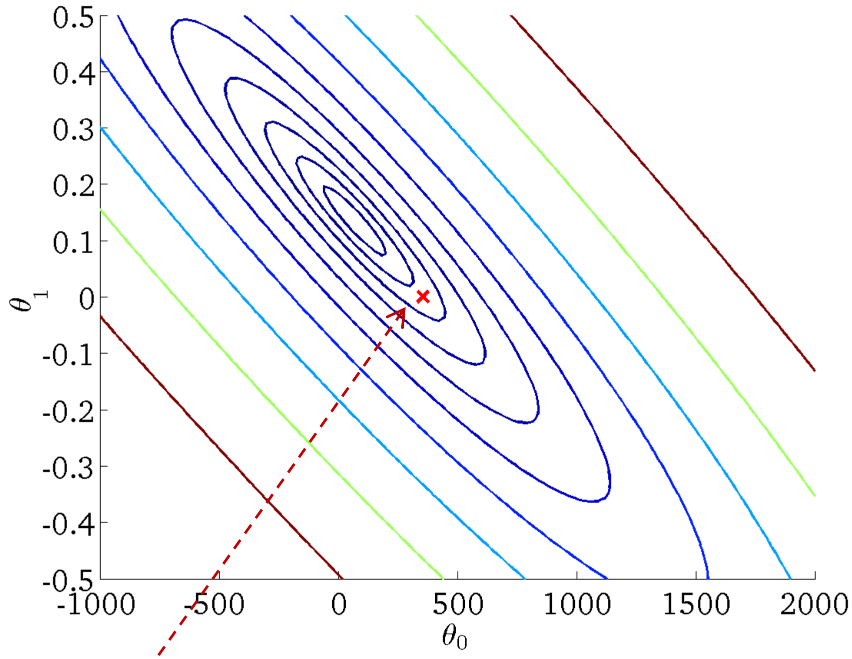
(для фиксированных θ_0, θ_1 - это функция x)



$h(x)=360+0\times x$

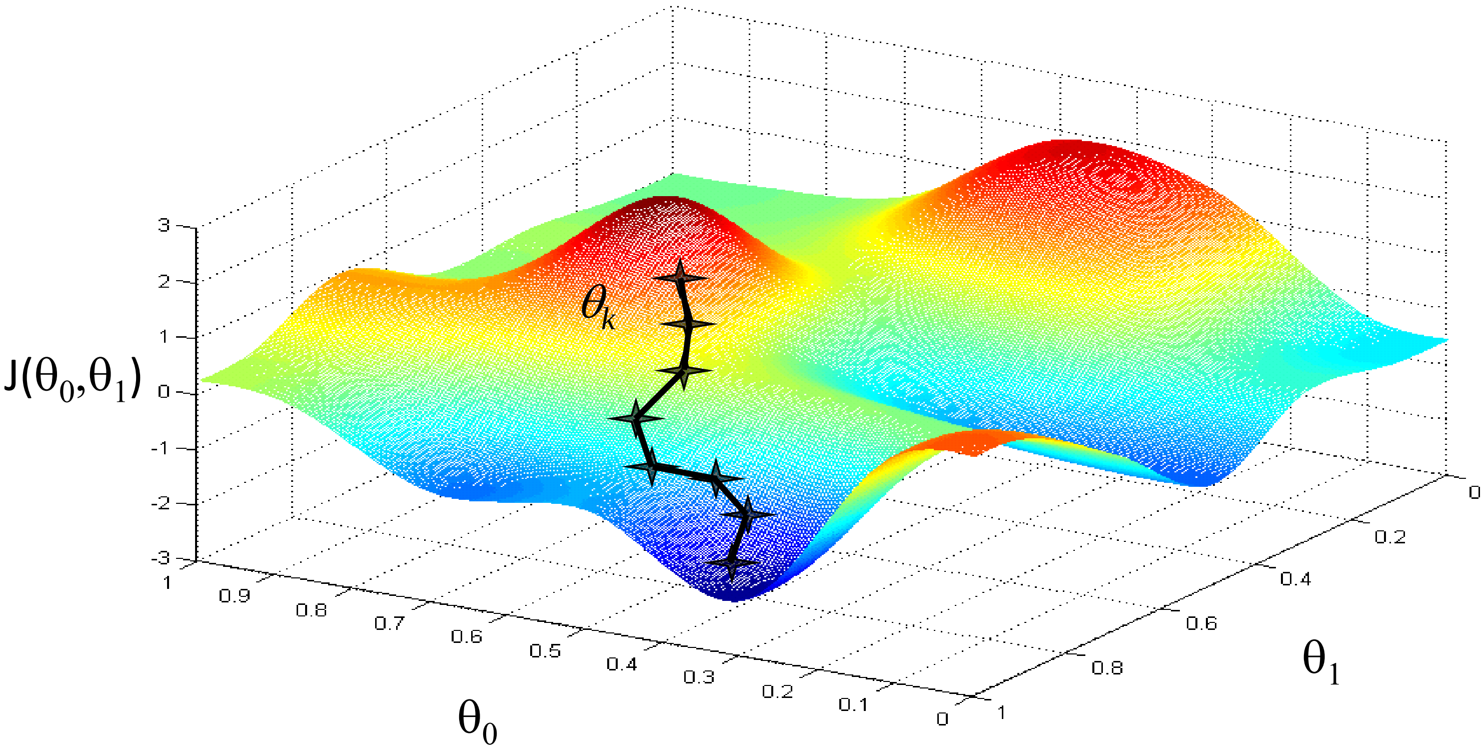
$J(\theta_0, \theta_1)$

(функция, зависящая от θ_0, θ_1)



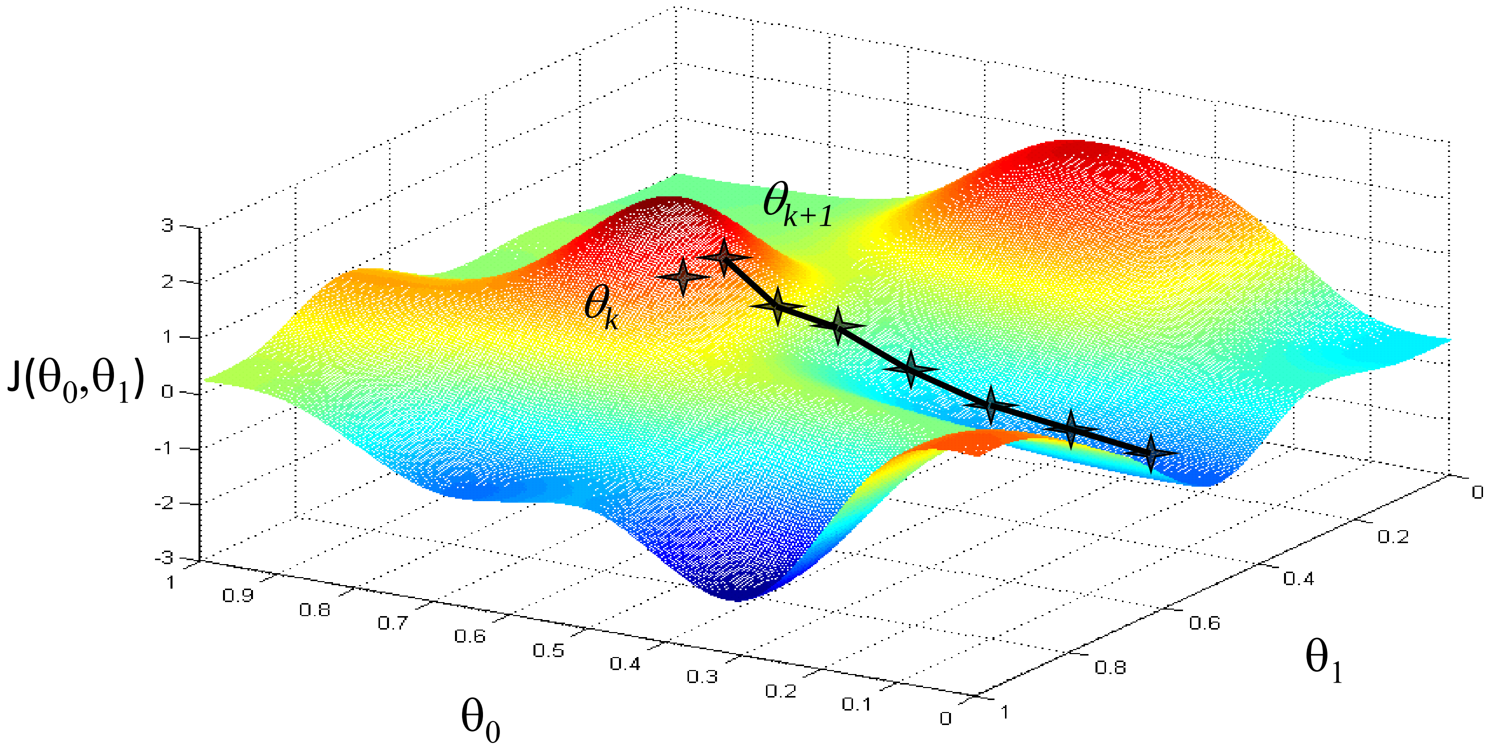
$\theta_0 = 360$
$\theta_1 = 0$

Сходимость алгоритма (1/2)



Сходимость алгоритма от начального значения θ_k

Сходимость алгоритма (2/2)



Сходимость алгоритма от начального значения θ_{k+1}

Итерационный процесс

Повторение до сходимости $\{$

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

$\}$

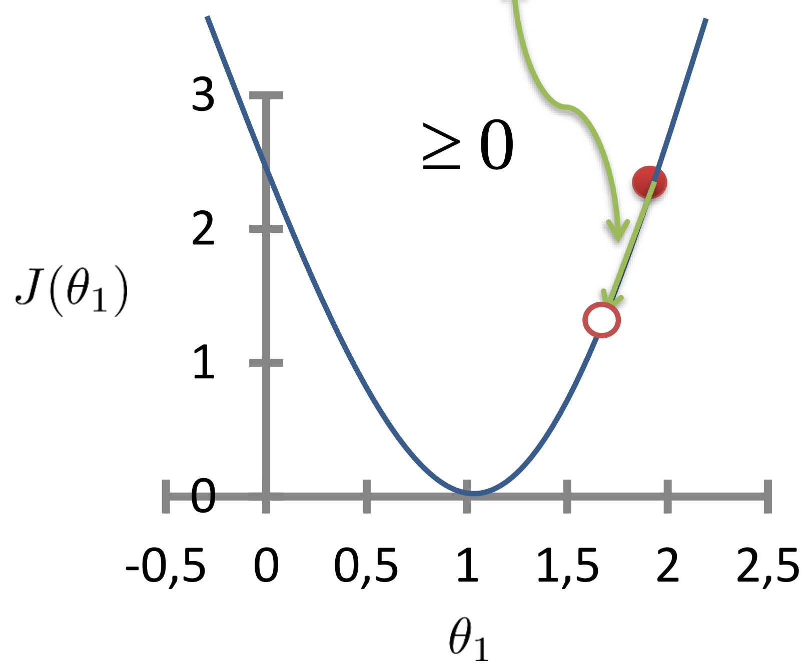
Скорость обучения

Указание:

Одновременное присвоение

$j = 0$ and $j = 1)$

$$\theta_j := \theta_j - \alpha \boxed{\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)}$$



Итерационный процесс

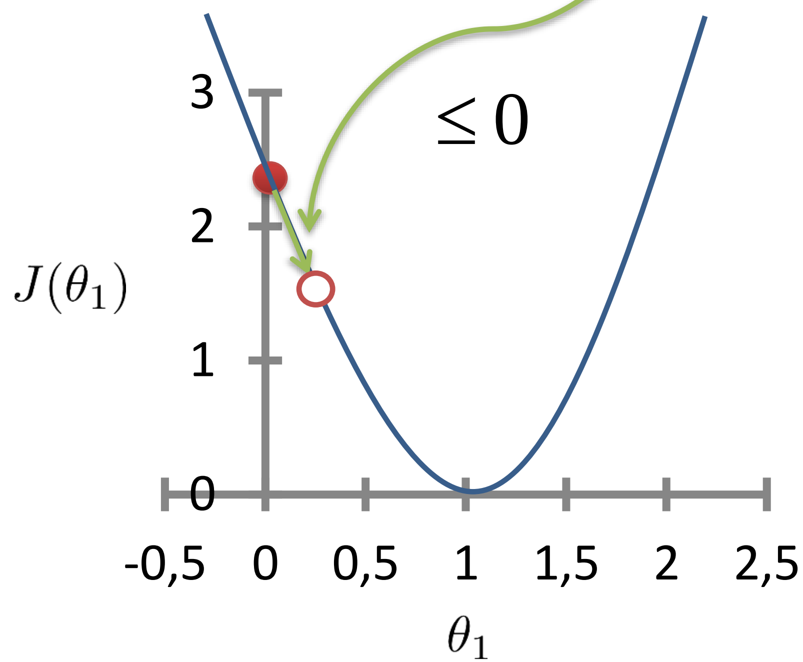
$$\theta_j := \theta_j - \alpha \boxed{\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)}$$

Повторение до сходимости $\{$

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

$\}$

Скорость обучения



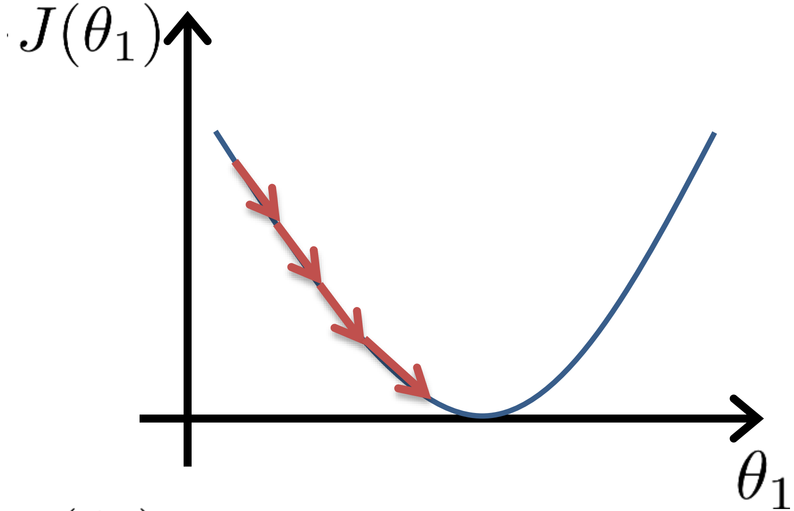
<u>Указание:</u>

Одновременное присвоение

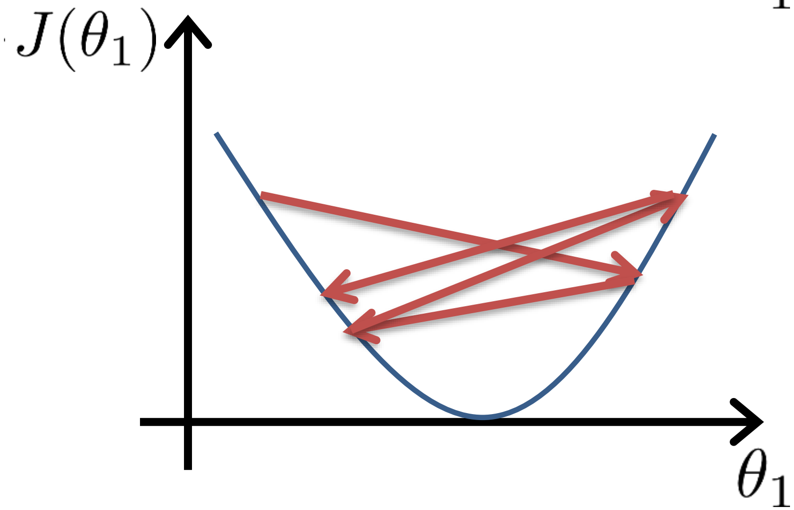
$j = 0$ and $j = 1$)

Скорость «спуска» (итерационного процесса)

$$\theta_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_1)$$



Если α слишком малая величина, градиентный спуск может быть очень медленным.



Если α выбрана достаточно большой, градиентный спуск может пропускать минимум. Возможны проблемы со сходимостью.

Градиентный спуск в задаче линейной регрессии

Повторение до сходимости {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

}

<u>Указание:</u>

Одновременное присвоение

$j = 0$ and $j = 1$)

Модель линейной регрессии

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

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

Градиентный спуск в задаче линейной регрессии

$$\frac{\partial}{\partial\theta_j} J(\theta_{0,}\theta_1) = \frac{\partial}{\partial\theta_j}\frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

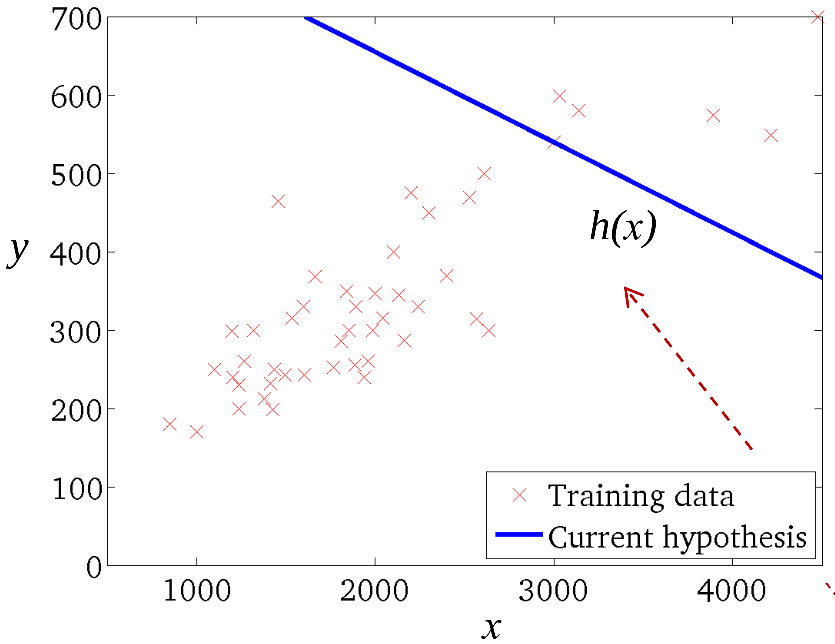
$$= \frac{\partial}{\partial\theta_j}\frac{1}{2m}\sum_{i=1}^{m}\left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)^2$$

$$\begin{cases} j=0: \dfrac{\partial}{\partial\theta_0} J(\theta_{0,}\theta_1) = \dfrac{1}{m}\displaystyle\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right) \\ j=1: \dfrac{\partial}{\partial\theta_1} J(\theta_{0,}\theta_1) = \dfrac{1}{m}\displaystyle\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)\cdot x^{(i)} \end{cases}$$

Итерационный процесс (1/9)

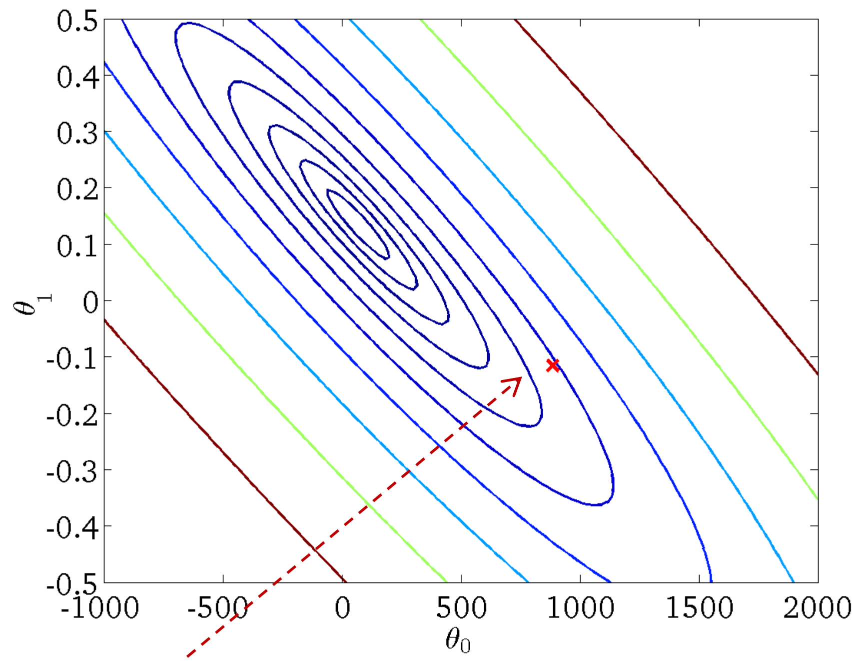
$h_\theta(x)$

(для фиксированных θ_0, θ_1 - это функция x)



$J(\theta_0, \theta_1)$

(функция, зависящая от θ_0, θ_1)



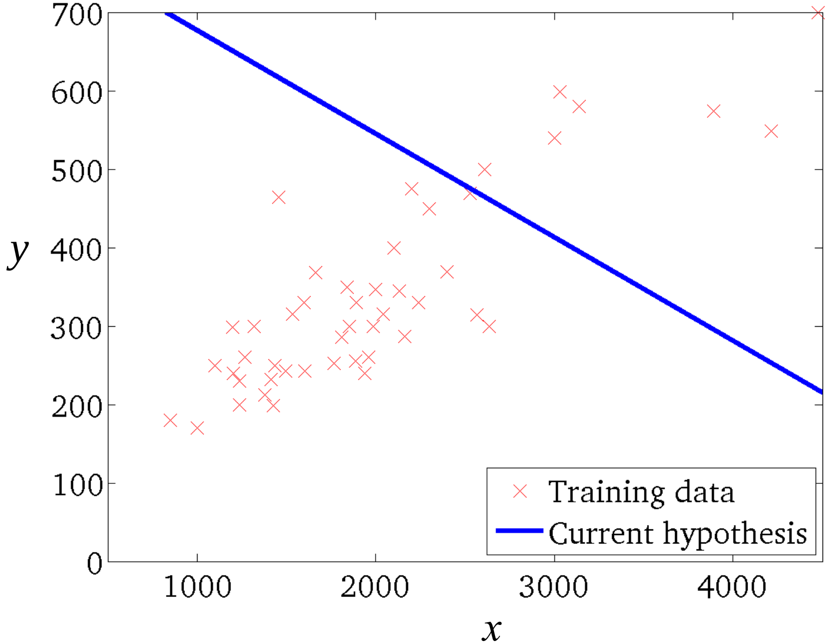
$h(x) = -900 - 0.1 \times x$

θ_0 = - 900
θ_1 = - 0.1

Итерационный процесс (2/9)

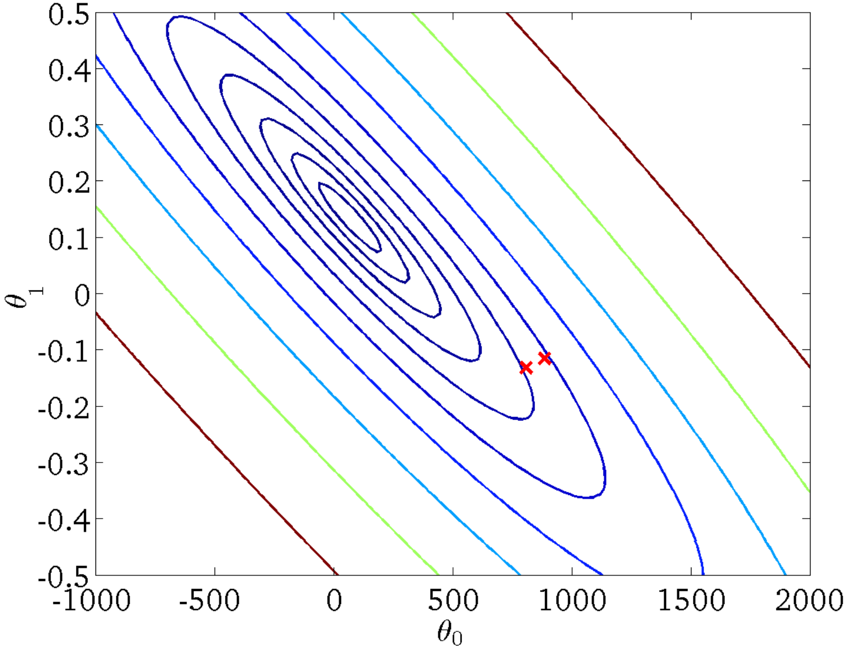
$$h_\theta(x)$$

(для фиксированных θ_0, θ_1 - это функция x)



$$J(\theta_0, \theta_1)$$

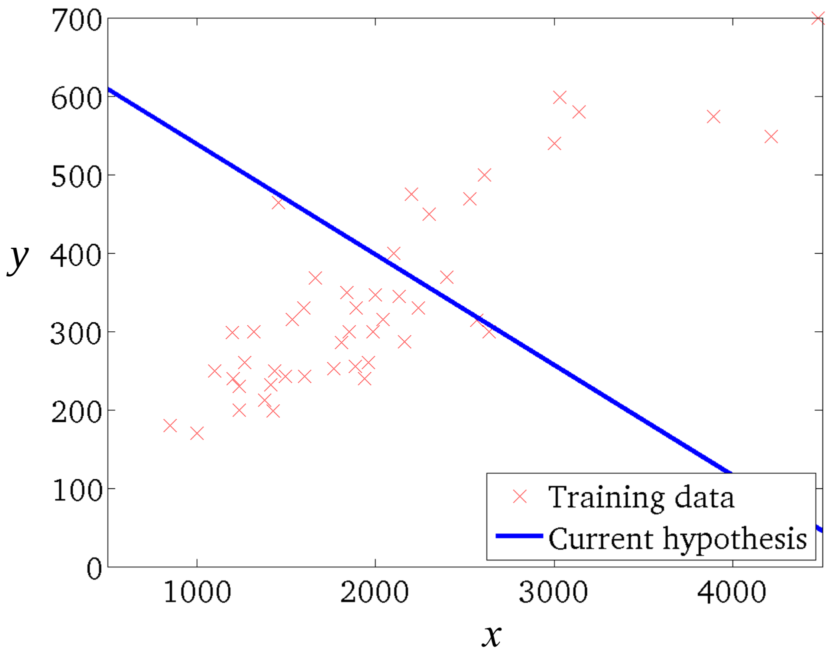
(функция, зависящая от θ_0, θ_1)



Итерационный процесс (3/9)

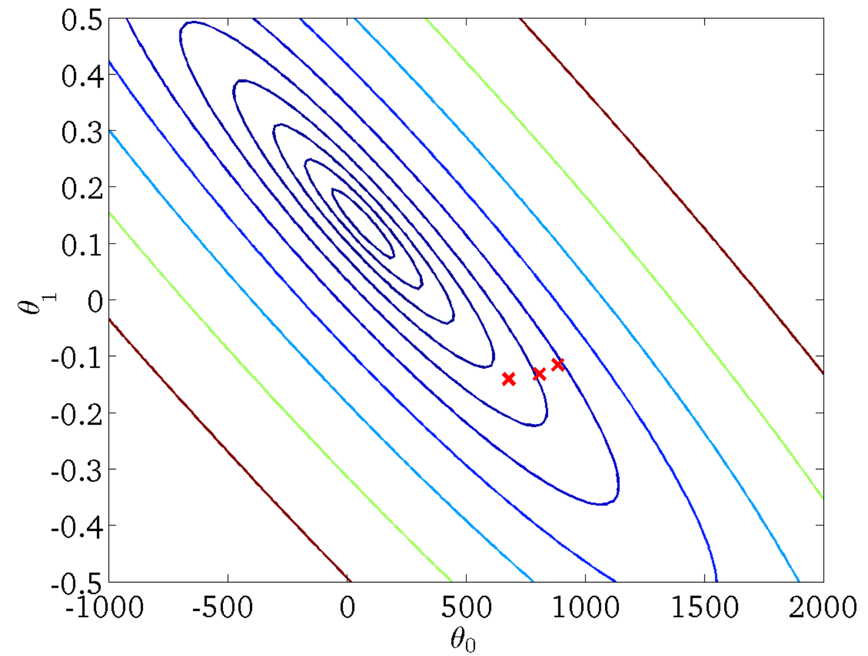
$h_\theta(x)$

(для фиксированных θ_0, θ_1 - это функция x)



$J(\theta_0, \theta_1)$

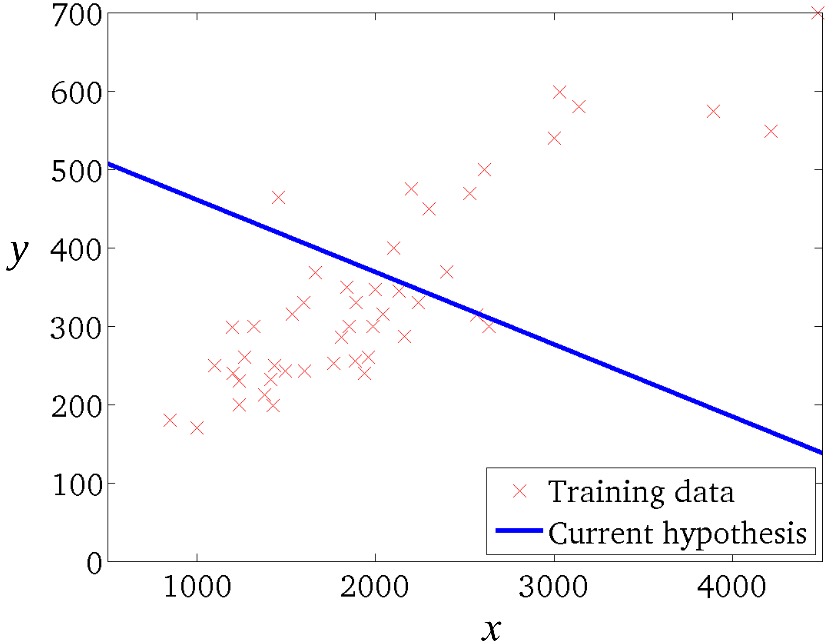
(функция, зависящая от θ_0, θ_1)



Итерационный процесс (4/9)

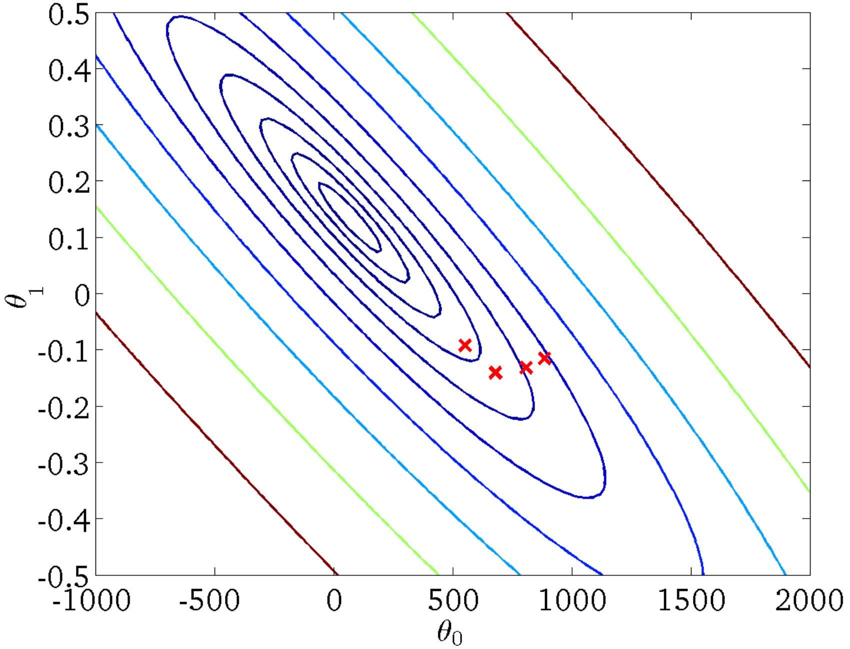
$h_\theta(x)$

(для фиксированных θ_0, θ_1 - это функция x)



$J(\theta_0, \theta_1)$

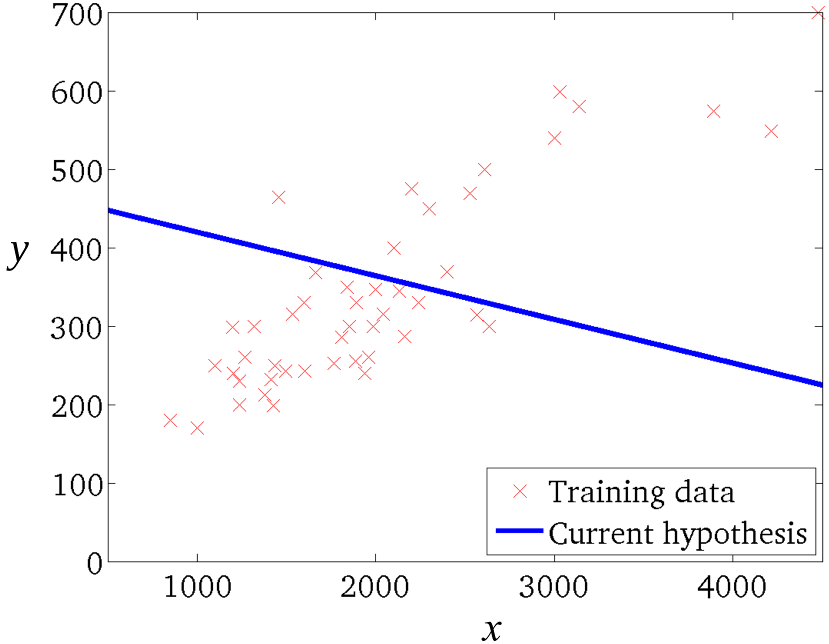
(функция, зависящая от θ_0, θ_1)



Итерационный процесс (5/9)

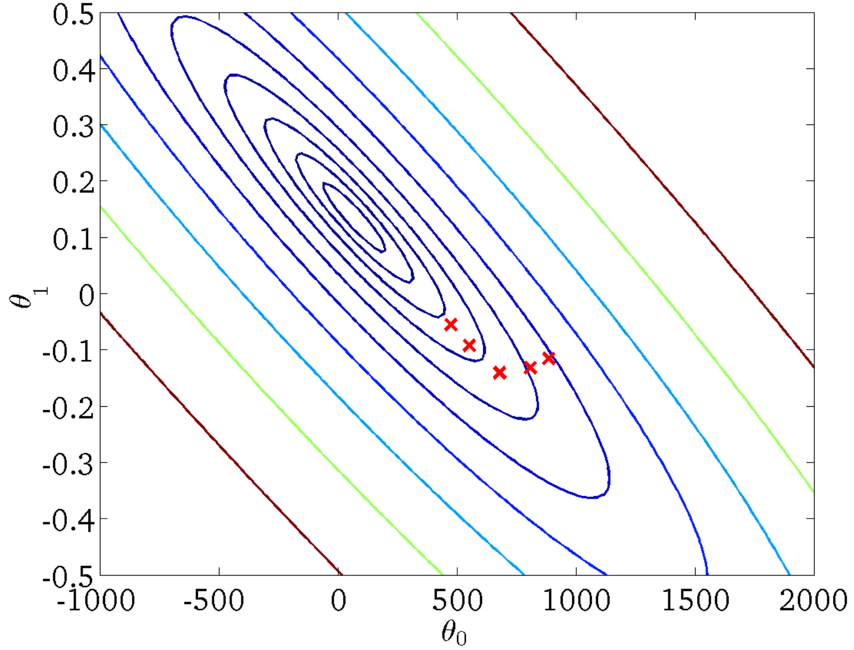
$$h_\theta(x)$$

(для фиксированных θ_0, θ_1 - это функция x)



$$J(\theta_0, \theta_1)$$

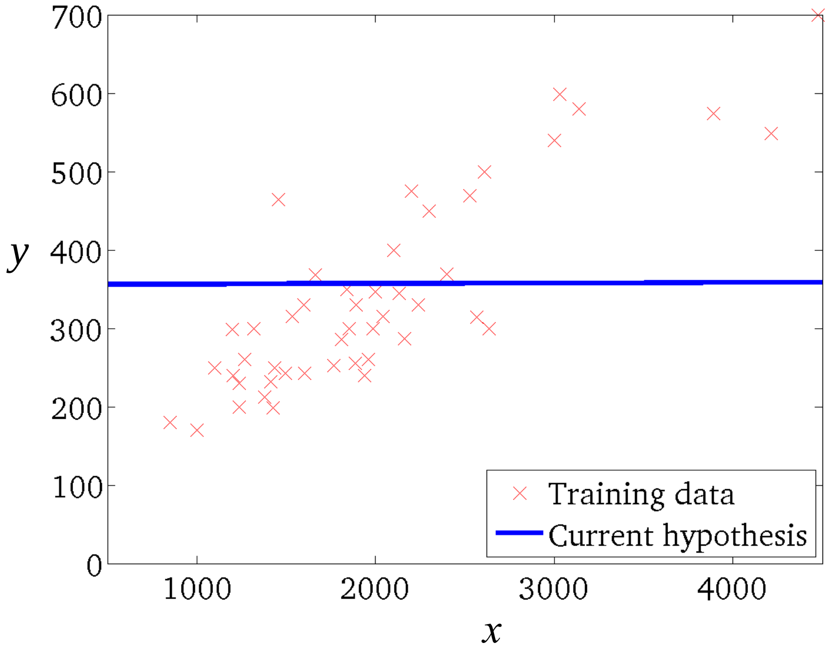
(функция, зависящая от θ_0, θ_1)



Итерационный процесс (6/9)

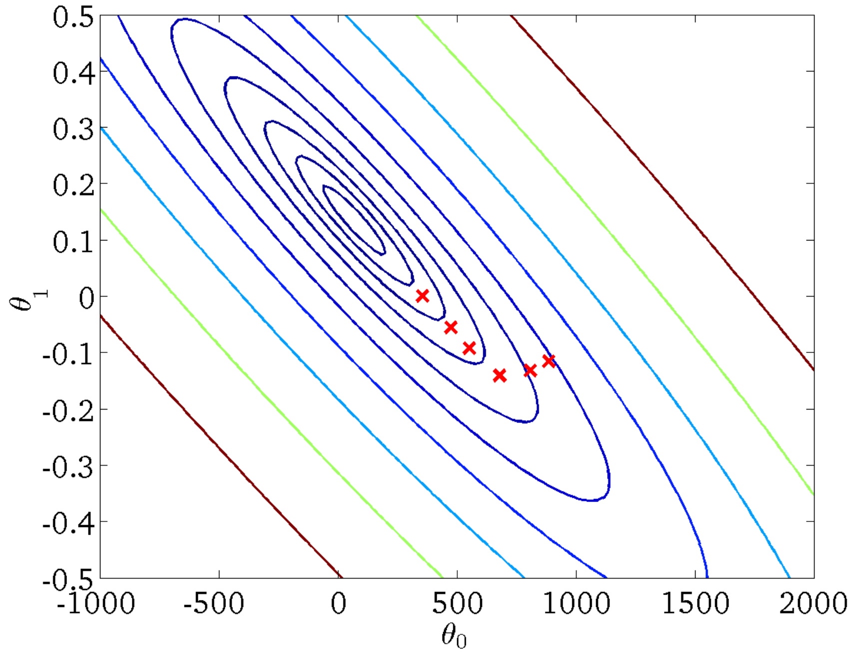
$$h_\theta(x)$$

(для фиксированных θ_0, θ_1 - это функция x)



$$J(\theta_0, \theta_1)$$

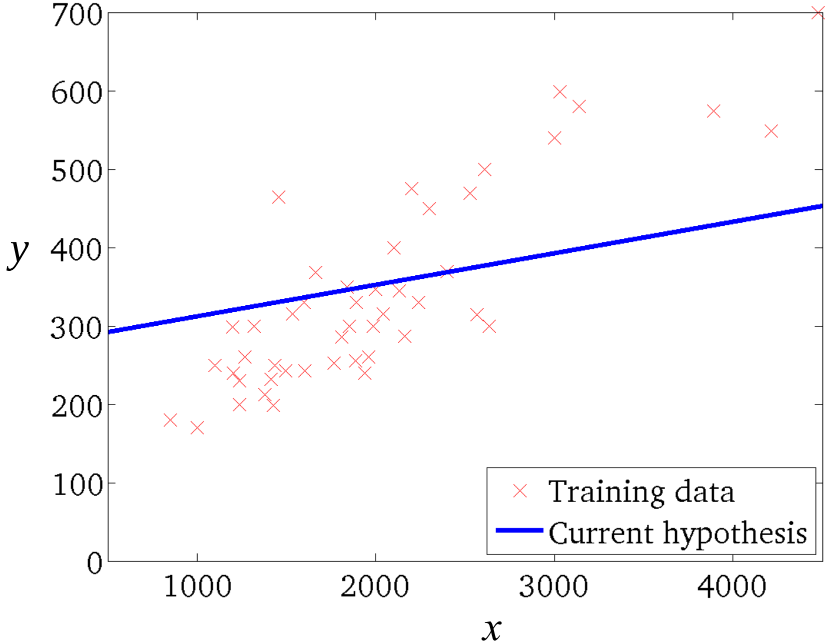
(функция, зависящая от θ_0, θ_1)



Итерационный процесс (7/9)

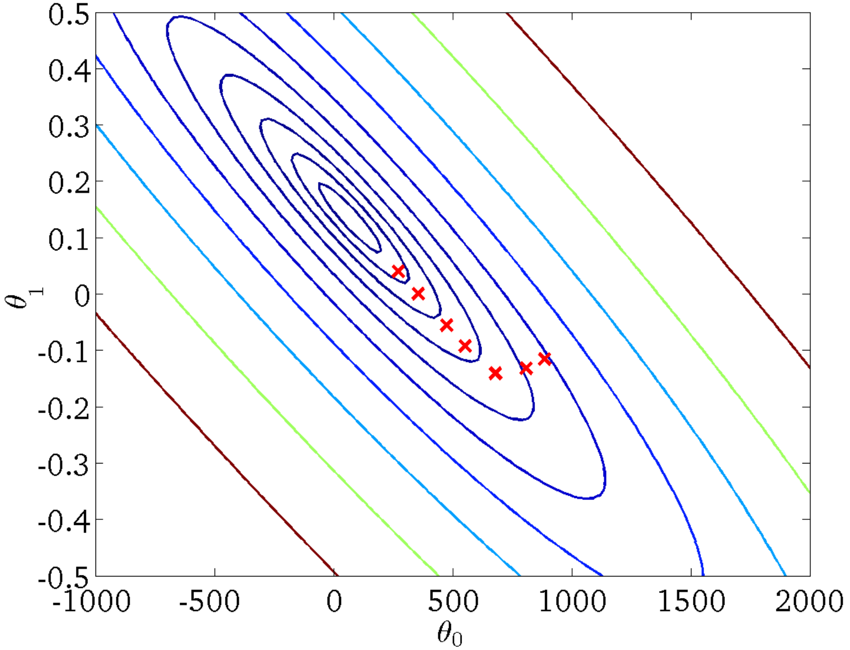
$h_\theta(x)$

(для фиксированных θ_0, θ_1 - это функция x)



$J(\theta_0, \theta_1)$

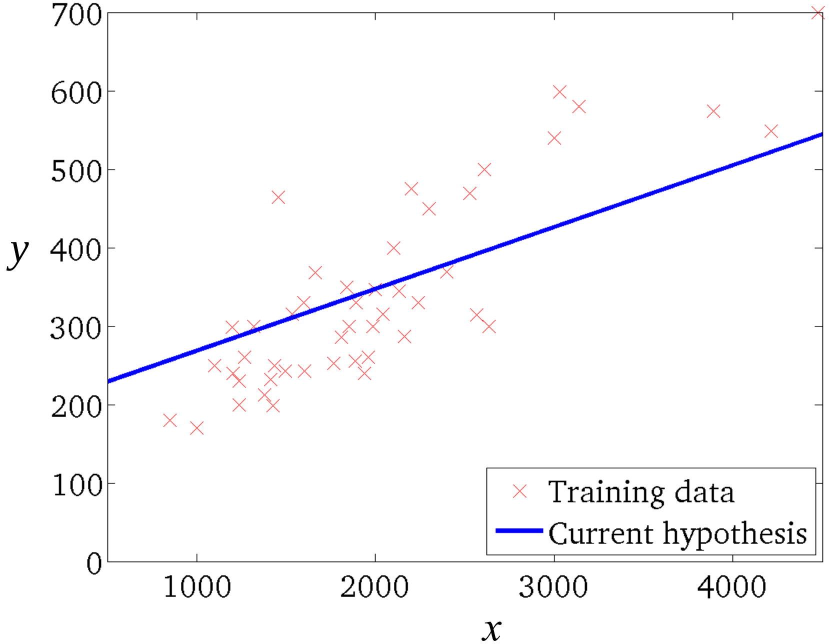
(функция, зависящая от θ_0, θ_1)



Итерационный процесс (8/9)

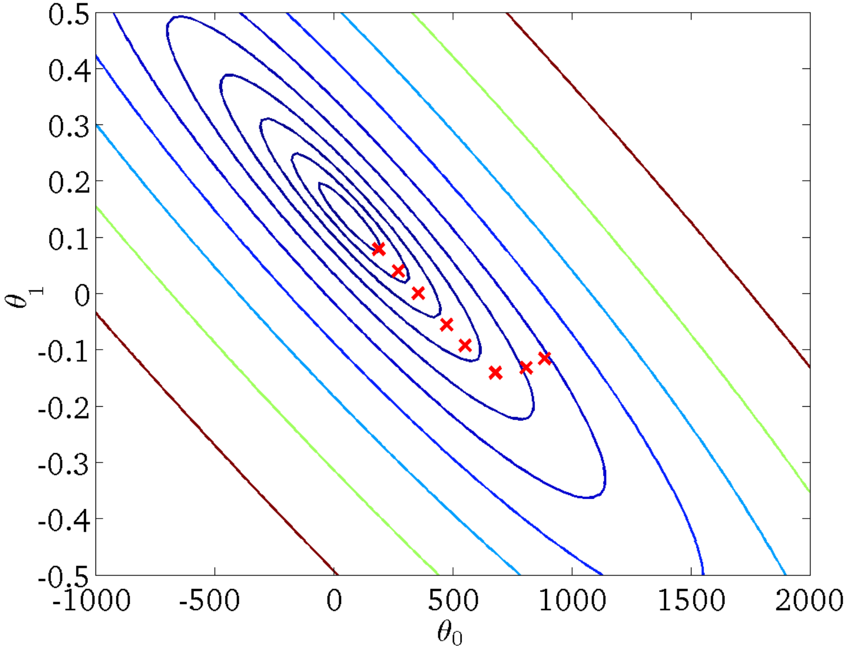
$$h_\theta(x)$$

(для фиксированных θ_0, θ_1 - это функция x)



$$J(\theta_0, \theta_1)$$

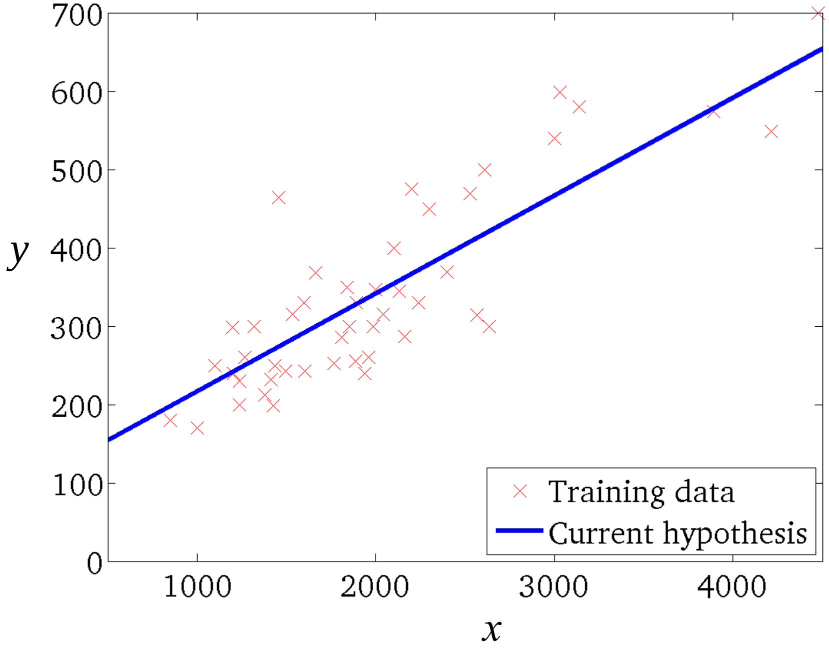
(функция, зависящая от θ_0, θ_1)



Итерационный процесс (9/9)

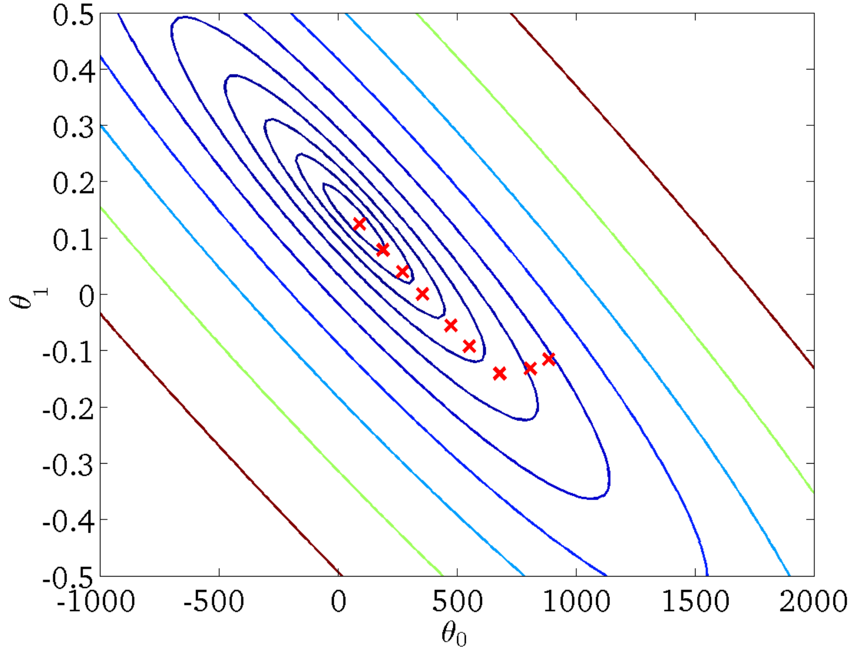
$h_\theta(x)$

(для фиксированных θ_0, θ_1 - это функция x)



$J(\theta_0, \theta_1)$

(функция, зависящая от θ_0, θ_1)



Материал лекции составлен на основании аналогичного курса «Машинное обучение» на портале он-лайн обучения Coursera.org (профессор Эндрю Ын, Стэнфордский университет - https://ru.wikipedia.org/wiki/Ын,_Эндрю)