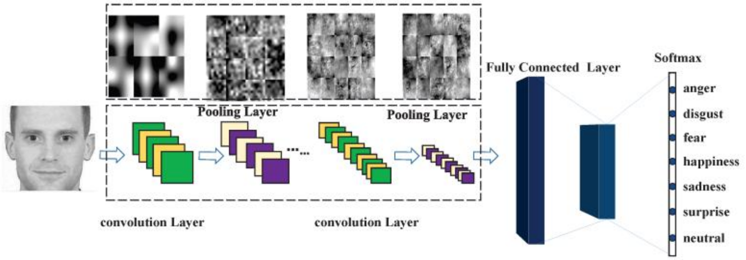
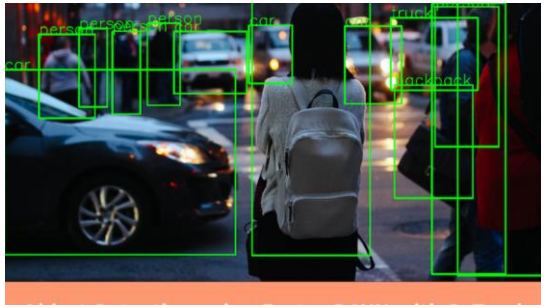
Основы искусственного интеллекта





Лекция

Обучение с учителем: Классическая регрессия

к.ф.-м.н., доцент кафедры ИСиЦТ
Корнаева Е.П.

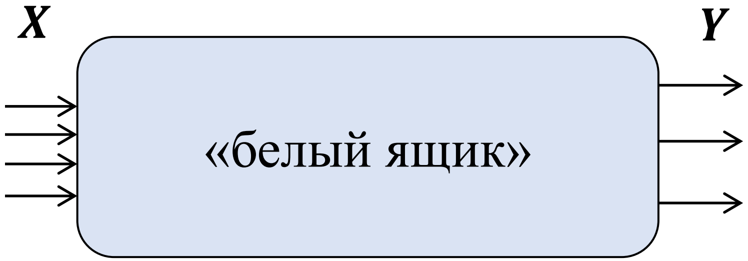
Обучение с учителем: Классическая регрессия

Постановка задачи классической регрессии

Признаки / Features (X): свойства объекта исследования, выделяемые при обучении

Виды зависимости:

- Функциональная;
- Статистическая (в частности - корреляционная).



Известен закон(ы):
$F_i(X_j, Y_k) = 0.$

Известны статистические характеристики:
- законы распределения, параметры законов распределения

Обучение с учителем: Классическая регрессия

Постановка задачи классической регрессии

Регрессионный (от лат. regressio – обратное движение, отход) **анализ:**
набор статистических действий для оценки связи зависимой переменной Y и одной или нескольких независимых переменных X.
Основная идея регрессии: поиск наилучшей функциональной зависимости среднего значения зависимой переменной $\boldsymbol{Y}$ от независимых переменных $\boldsymbol{X}$ по критерию минимума некоторой функции качества.

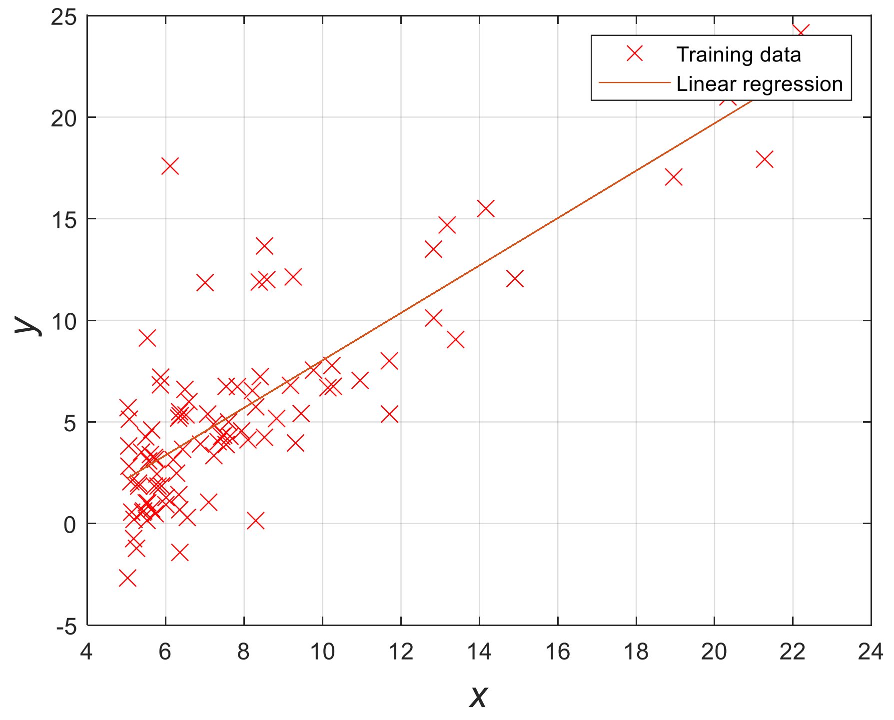


Рис. 4.1. Линейная регрессия

По выборке можно найти только оценку уравнения регрессии:

$$y(\boldsymbol{X}) = \bar{y}(\boldsymbol{X}) + \varepsilon$$

Пример линейной регрессии для данных на рис. 4.1:

$$\bar{y}(x) = \theta_0 + \theta_1 x$$

Скалярная форма записи:

$h(x) = \bar{y}(x)$ – оценка регрессии (функция гипотезы h)

Обучение с учителем: Классическая регрессия

Постановка задачи классической регрессии

Данные выборки представлены множеством пар: $\{(\boldsymbol{X}_i, \boldsymbol{y}_i)\}, \boldsymbol{i} = \overline{\boldsymbol{1}, \boldsymbol{n}}$

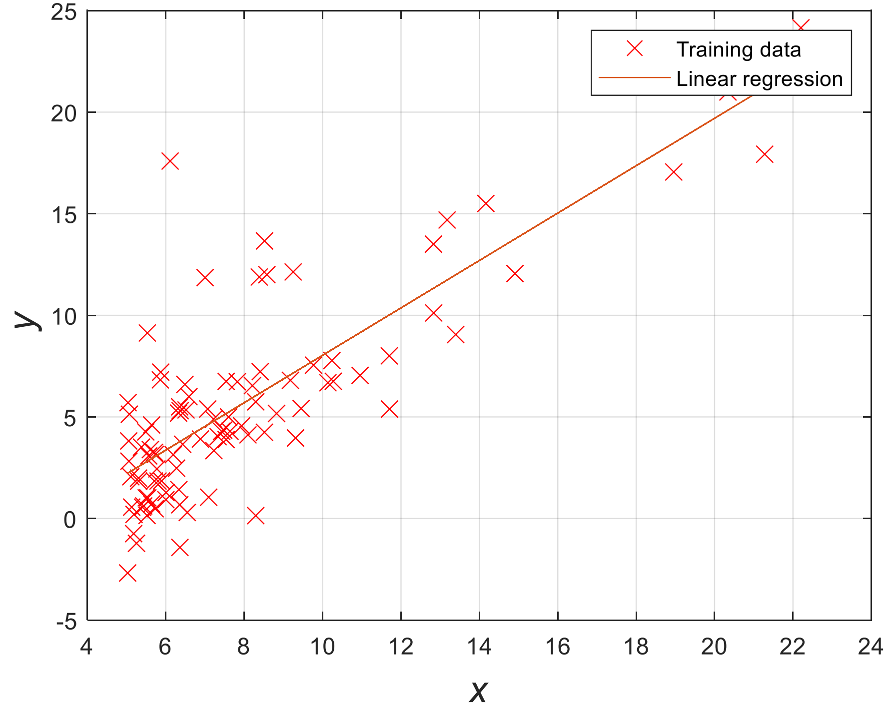


Рис. 4.1. Линейная регрессия

<u>Пример</u> линейной однофакторной регрессии для данных на рис. 4.1:

$$\mathrm{h}(x) = \theta_0 x_0 + \theta_1 x_1 = \theta_j x_j, \; j = \overline{0,1}$$

Матричная форма записи:

$$\boldsymbol{H}(\boldsymbol{\theta}, \boldsymbol{X}) = \boldsymbol{X}\boldsymbol{\theta}$$

$\boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}$ - весовые коэф-ты; $\boldsymbol{Y} = \begin{pmatrix} y_0 \\ \dots \\ y_n \end{pmatrix}$ - матрица значений отклика;

$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} \\ & \dots \\ 1 & x_{n1} \end{pmatrix}$ - матрица значений признаков.

Функция качества (средняя кв. ошибка): МНК

$$J(\theta_0, \theta_1) = \frac{1}{n} \sum_{i=1}^{n} (h(\boldsymbol{\theta}, \boldsymbol{X}_i) - y_i)^2 \Rightarrow \min$$

Обучение с учителем: Классическая регрессия

Какие должны быть параметры модели (т.е. весовые коэффициенты)? Пример.

Наглядный выбор параметров модели (рис. 4.2а)

Формализованный выбор параметров модели (рис. 4.2б):

$$J(\theta_1) = \frac{1}{n}\sum_{i=1}^{n}(\theta_1 x_i - y_{эi})^2 \rightarrow \min$$

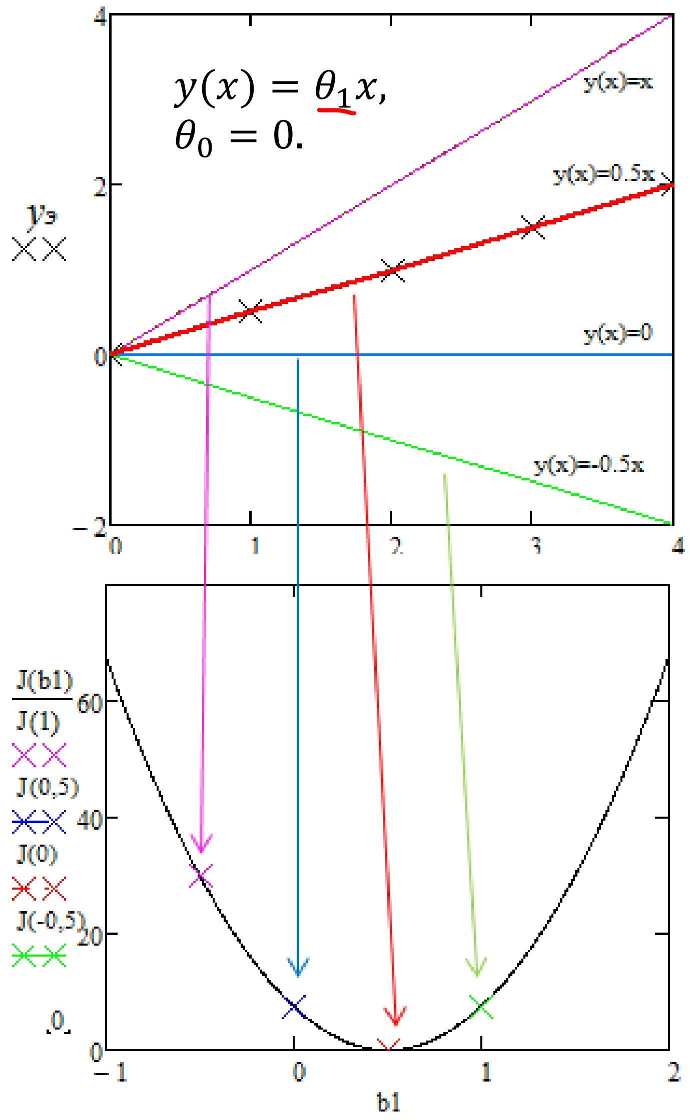


Рис. 4.2

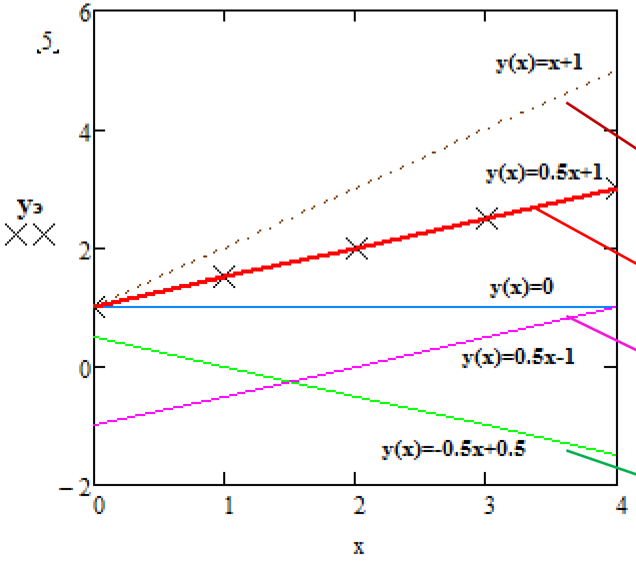
x	y
0	0
1	0,5
2	1
3	1,5
4	2

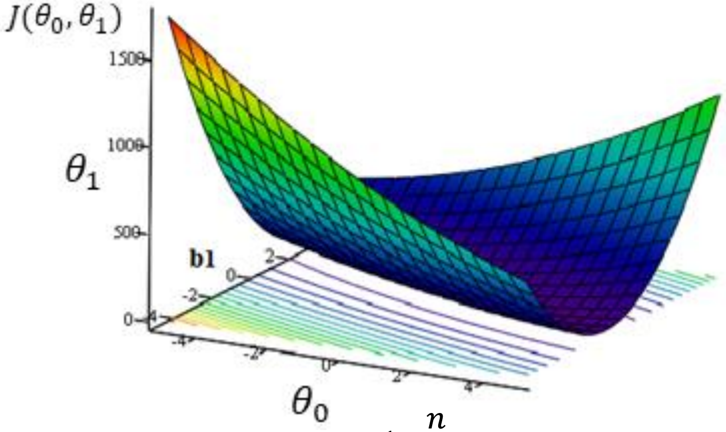
Обучение с учителем: Классическая регрессия

- МНК. Линейная однофакторная регрессия

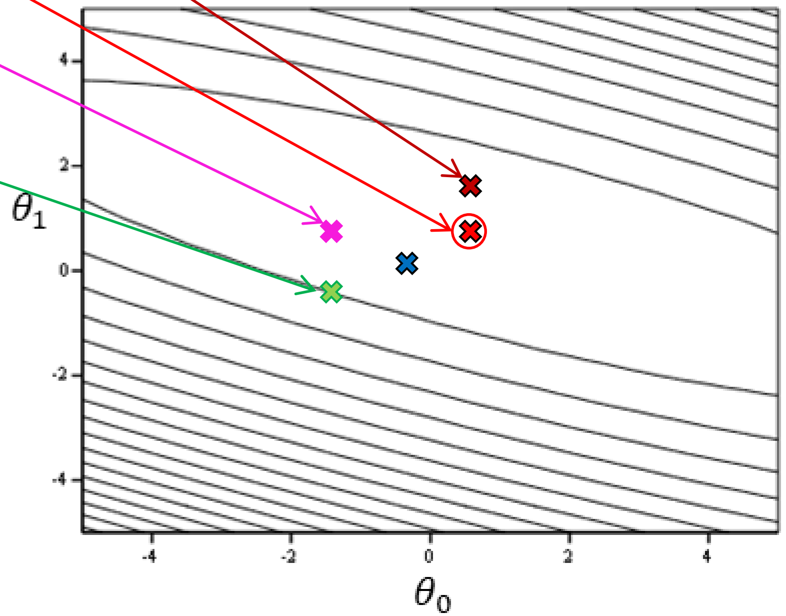
$$y(x) = \theta_0 + \theta_1 x$$

x	y
0	1
1	1,5
2	2
3	2,5
4	3





$$J(\theta_0, \theta_1) = \frac{1}{n}\sum_{i=1}^{n}(\theta_0 + \theta_1 x_i - y_{эi})^2 \rightarrow \min$$



Обучение с учителем: Классическая регрессия

Линейная регрессия в общем виде:

$$y(x_1, \dots x_m) = \theta_0 + \theta_j x_j = \theta_j x_j{}^{*}, \qquad \mathrm{j} = \overline{0, \mathrm{m}}$$

m=2:

$$y(x_1, x_2) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

	X_0	X_1	X_2	Y
1	1	x_{11}	x_{12}	y_1
2	1	x_{21}	x_{22}	y_2
	…			
n	1	x_{n1}	x_{n2}	y_n

X

Функция качества (функция ошибки):

$$J(\theta_0, \theta_1, \theta_2) = \frac{1}{n} \sum_{i=1}^{n} (\theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2} - y_i)^2 \to \min$$

$$\frac{\partial J}{\partial \theta_j} = 0, j = 0..2$$

* Используется правило Эйнштейна

Обучение с учителем: Классическая регрессия

Постановка задачи классической регрессии

$g(f(x,y)) =$

Обобщение для линейной регрессии на m факторов:

$$g(x,y) = (\underbrace{5x+y}^{f})^2 = f^2(x,y)$$

Например, m=2:

$$y(x_1, x_2) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

$$g'_x = 2f \cdot 5 = 10(5x+y)$$

Функция качества (функция ошибки):

$$J(\theta_0, \theta_1, \theta_2) = \frac{1}{n}\sum_{i=1}^{n}(\theta_0 x_{i0} + \theta_1 x_{i1} + \theta_2 x_{i2} - y_i)^2 \to \min$$

$$\frac{2}{n}\sum_{i=1}^{n} x_{i0}(\theta_0 x_{i0} + \theta_1 x_{i1} + \theta_2 x_{i2} - y_{эi}) = 0$$

$$\frac{\partial J}{\partial \theta_j} = 0, j = 0..2$$

$$\frac{2}{n}\sum_{i=1}^{n} x_{i1}(\theta_0 x_{i0} + \theta_1 x_{i1} + \theta_2 x_{i2} - y_{эi}) = 0$$

$$\frac{2}{n}\sum x_{i2}(\theta_0 x_{i0} + \theta_1 x_{i1} + \theta_2 x_{i2} - y_{эi}) = 0$$

$$\frac{\partial y}{\partial \theta_j} = 0: \quad \sum_i x_{ij}(\theta_0 x_{i0} + \theta_1 x_{i1} + \theta_2 x_{i2} - y_{эi}) = 0$$

Обучение с учителем: Классическая регрессия

Постановка задачи классической регрессии

Обобщение для линейной регрессии на m факторов:

Например, m=2:

$$y(x_1, x_2) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

$$\frac{\partial J}{\partial \theta_j} = 0, j = 0..2$$

$$A = X^T X \qquad \theta \qquad F = X^T y$$

$$\begin{bmatrix} \sum_i (x_{i0}^2) & \sum_i x_{i0} x_{i1} & \sum x_{i0} x_{i2} \\ \sum x_{i1} x_{i0} & \sum_i (x_{i1}^2) & \sum_i x_{i1} x_{i2} \\ \sum (x_{i2} x_{i0}) & \sum (x_{i2} x_{i1}) & \sum x_{i2}^2 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} \sum x_{i0} y_i \\ \sum x_{i1} y_i \\ \sum x_{i2} y_i \end{bmatrix}$$

$$A\theta = F \Rightarrow \theta = A^{-1} F$$

Обучение с учителем: Классическая регрессия

Постановка задачи классической регрессии

Обобщение для линейной регрессии на m факторов. Аналитическое решение

$$\frac{\partial J}{\partial \theta_j} = 0, j = 0..2$$

$$\mathbf{X^T X}\ \boldsymbol{\theta} = \mathbf{X^T Y}$$

	X_0	X_1	X_2	**Y**
1	1	x_{11}	x_{12}	y_1
2	1	x_{21}	x_{22}	y_2
…				
n	1	x_{n1}	x_{n2}	y_n

X

для относительно небольшого кол-ва признаков и объема выборки можно использовать аналитическое решение:

$$\boxed{\boldsymbol{\Theta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}}$$

$\boldsymbol{X}_{[n\times(m+1)]}$, $\boldsymbol{Y}_{[n\times 1]}$, $\boldsymbol{\Theta}_{[m+1,1]}$,

Обучение с учителем: Классическая регрессия

Постановка задачи классической регрессии

Полиномиальная регрессия. Сведение к линейной

$$y(x_1, x_2, \dots, x_k) = \theta_0 + \sum_{k=1}^{p} \sum_{i=1}^{k} \theta_{(p-1)k+i} x_i^p + \sum_{\substack{i,j=1 \\ i \neq j}}^{k} \theta_{pk+i} x_i x_j$$

Например, $p = 2, k = 2$:

$$y(x_1, x_2) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2,$$

	X_0	X_1	X_2	X_3	X_4	X_5	Y
1	1	x_{11}	x_{12}	x^2_{11}	x^2_{12}	$x_{11}x_{12}$	y_1
2	1	x_{21}	x_{22}	x^2_{21}	x^2_{22}	$x_{21}x_{22}$	y_2
…							…
n	1	x_{n1}	x_{n2}	x^2_{n1}	x^2_{n2}	$x_{n1}x_{n2}$	y_n

$\boldsymbol{X}$

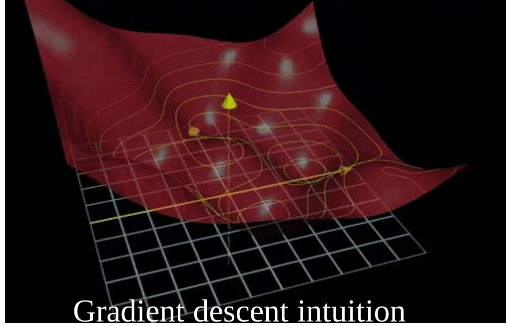
пусть $\quad x_3 = x_1^2, x_4 = x_2^2, x_5 = x_1 x_2.$

$$y(x_1, \dots, x_5) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4 + \theta_5 x_5, \qquad \boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \end{pmatrix}$$

$$J(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} (\theta_0 + \theta_1 x_{i1} + \dots + \theta_5 x_{i5} - y_i)^2 \to \min$$

$$\frac{\partial J}{\partial \theta_j} = 0, j = 0..5$$

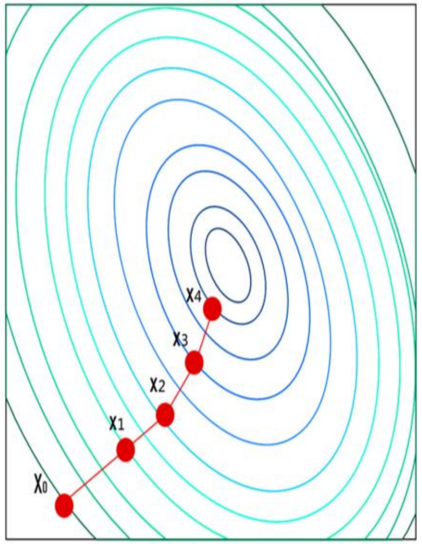
Обучение с учителем: Классическая регрессия



Метод градиентного спуска

$f = f(\boldsymbol{\Theta})$ - скалярная функция, $\boldsymbol{\Theta} = (\Theta_j),\ j = 0, \ldots m$

$f_{min} = f(\boldsymbol{\Theta}^*)$ - минимум (экстремум) функции, $\boldsymbol{\Theta}^*$ - точка минимума (экстремума)



<u>Инициализация:</u>

- α - шаг градиентного спуска (скорость обучения);
- $\boldsymbol{\Theta}^{\mathbf{0}} = (\boldsymbol{\Theta}_j^{\mathbf{0}})$ – начальное приближение точки минимума;
- $\boldsymbol{J}^0 = J(\boldsymbol{\Theta}^0)$ – начальное значение экстремума;
- $\boldsymbol{\nabla J}^0 = \left[\!\left[\frac{\partial J(\boldsymbol{X})}{\partial \Theta_j}\right]\!\right]^0$ - значение градиента в начальной точке.

<u>Градиентный спуск:</u>

- Шаг градиентного спуска:

$$\Theta_j^{l+1} = \Theta_j^l - \alpha \frac{\partial J}{\partial \Theta_j}\Big|_l \text{ или } \boldsymbol{\Theta}^{l+1} = \boldsymbol{\Theta}^l - \alpha\, \boldsymbol{\nabla} J|_l$$

- Компоненты градиента и значение функции на текущем шаге l:

$$\boldsymbol{\nabla} J^{l+1} = \left[\!\left[\frac{\partial J(\boldsymbol{X})}{\partial \Theta_j}\right]\!\right]^{l+1}, \qquad \boldsymbol{J}^{l+1} = J(\boldsymbol{\Theta}^{\boldsymbol{l+1}})$$

<u>Условие останова:</u>

$$opt = [\boldsymbol{\Theta}^{l+1}, J^{l+1}]\ if\ \left\|\boldsymbol{\nabla} J^{l+1}\right\| < \varepsilon$$

Обучение с учителем: Классическая регрессия

Численный расчет весовых коэффициентов методом градиентного спуска

Компоненты градиента:

$$\frac{\partial J(\boldsymbol{\theta})}{\partial \theta_j} = \frac{2}{n}\sum_{i=1}^{n}\left(h(x_{ij}) - y_i\right)x_{ij}$$

$$\nabla \boldsymbol{J} = \frac{2}{n}\boldsymbol{X}^T\,(\boldsymbol{H} - \boldsymbol{Y}),$$

Расчет коэффициентов на каждом шаге *l+1*:

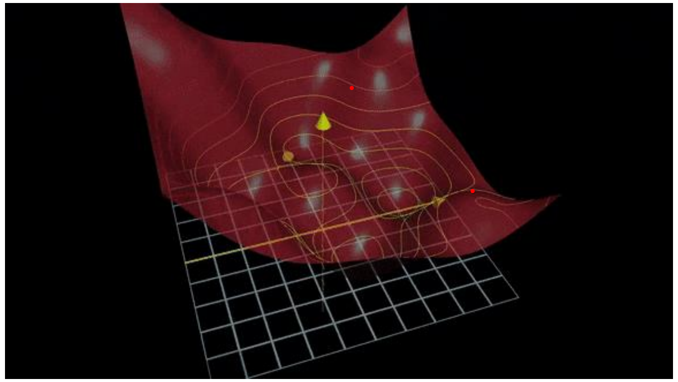
$$\theta_j^{l+1} = \theta_j^l - \alpha\frac{\partial J}{\partial \theta_j}|_l$$

$$\boldsymbol{\Theta}^{l+1} = \boldsymbol{\Theta}^l - \alpha\nabla\boldsymbol{J}|_l$$

Значение функции ошибки на каждом шаге *l+1* :

$$J\Big|_{l+1} = \frac{1}{n}\sum_{i=1}^{n}\left(h(\boldsymbol{\theta}, \boldsymbol{X}_i) - y_i\right)^2$$

$$J\Big|_{l+1} = \frac{1}{n}\left(\boldsymbol{X}\boldsymbol{\Theta}^{l+1} - \boldsymbol{Y}\right)^T\left(\boldsymbol{X}\boldsymbol{\Theta}^{l+1} - \boldsymbol{Y}\right)$$



Gradient descent intuition

$\boldsymbol{\nabla J}|_l = \left(\frac{\partial J}{\partial \theta_j}\right)|_l$ – компоненты градиента рассчитываются на каждом шаге *l;*

$\boldsymbol{\Theta}^0 = (\theta_j^0)$ – начальное значение коэффициентов;

α – шаг градиентного спуска (скорость обучения).

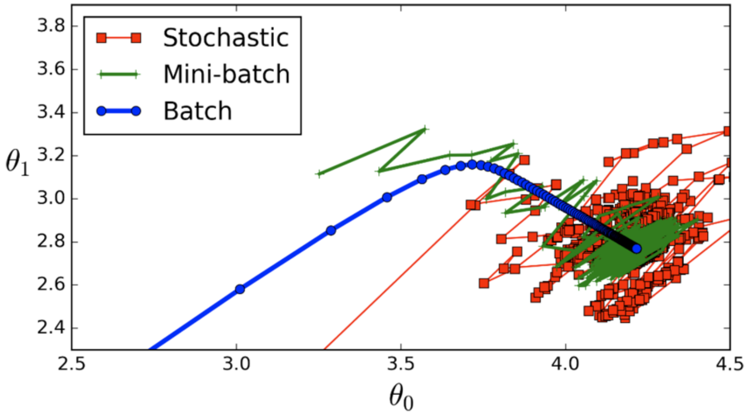
Обучение с учителем: Классическая регрессия

Метод стохастического градиентного спуска (SGD)

<u>Проблема</u>: большой обучающий набор необходим для более качественного обобщения, но обучение на нем обходится дорого с точки зрения вычислений

$J = J(\boldsymbol{\theta})$ - скалярная функция, $\boldsymbol{\theta} = \left(\theta_j\right), i = 0, ..m$

$J_{min} = J(\boldsymbol{\theta}^*)$ - минимум (экстремум) функции, $\boldsymbol{\theta}^*$ - точка минимума (экстремума)



Основные шаги алгоритма SGD аналогичны классическому градиентному спуску:
- Инициализация начального решения $\boldsymbol{\Theta}^0$;
- Шаг градиентного спуска $\theta_j^{l+1} = \theta_j^l - \alpha \frac{\partial J}{\partial \theta_j}|_l$;
- Проверка условия остановки.

Компоненты градиента на каждом шаге l по случайно выбранной точке* $(X_s, y_s) = (1\ x_{s1}\ x_{s2}\ x_{sm}, y_s)$:

$$\frac{\partial J(\boldsymbol{\theta})}{\partial \theta_j} = \frac{2}{n}(h(X_s) - y_s)x_{sj}$$

X_s

	X_0	X_1	…	X_m	Y
1	1	x_{11}		x_{12}	y_1
…					
s	1	x_{s1}		x_{m1}	y_s
…					
n	1	x_{n1}		x_{n2}	y_n

* Случайный выбор выполняется на каждом шаге

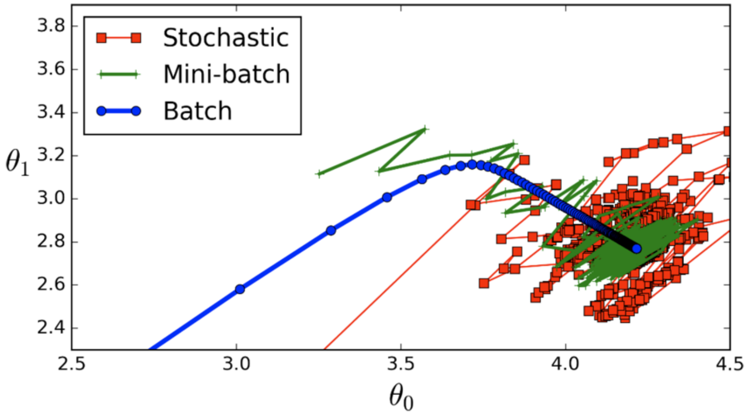
Обучение с учителем: Классическая регрессия

Метод стохастического градиентного спуска (SGD)

<u>Проблема</u>: большой обучающий набор необходим для более качественного обобщения, но обучение на нем обходится дорого с точки зрения вычислений

$J = J(\boldsymbol{\theta})$ - скалярная функция,$\boldsymbol{\theta} = (\theta_j), i = 0,..m$

$J_{min} = J(\boldsymbol{\theta}^*)$ - минимум (экстремум) функции, $\boldsymbol{\theta}^*$ - точка минимума (экстремума)



Основные шаги алгоритма Mini-batch аналогичны классическому градиентному спуску:

- Инициализация начального решения $\boldsymbol{\Theta}^0$;
- Шаг градиентного спуска $\theta_j^{l+1} = \theta_j^l - \alpha \frac{\partial J}{\partial \theta_j}|_l$;
- Проверка условия остановки.

Компоненты градиента на каждом шаге l по случайной выборке $\{\boldsymbol{X}_s, \boldsymbol{Y}_s\}$ (мини-батч) объема n' из $\{\boldsymbol{X}, \boldsymbol{Y}\}$:

$$\frac{\partial J(\boldsymbol{\theta})}{\partial \theta_j} = \frac{2}{n} \sum_{s=1}^{n'} (h(\boldsymbol{X}_s) - \boldsymbol{Y}_s)\boldsymbol{X}_{sj}$$

	X_0	X_1	…	X_m	Y
1	1	x_{11}		x_{12}	y_1
…					
n'	1	x_{n1}		x_{n2}	$y_{n'}$

$\boldsymbol{X}_s$ $\boldsymbol{Y}_s$

* Случайный выбор выполняется на каждом шаге