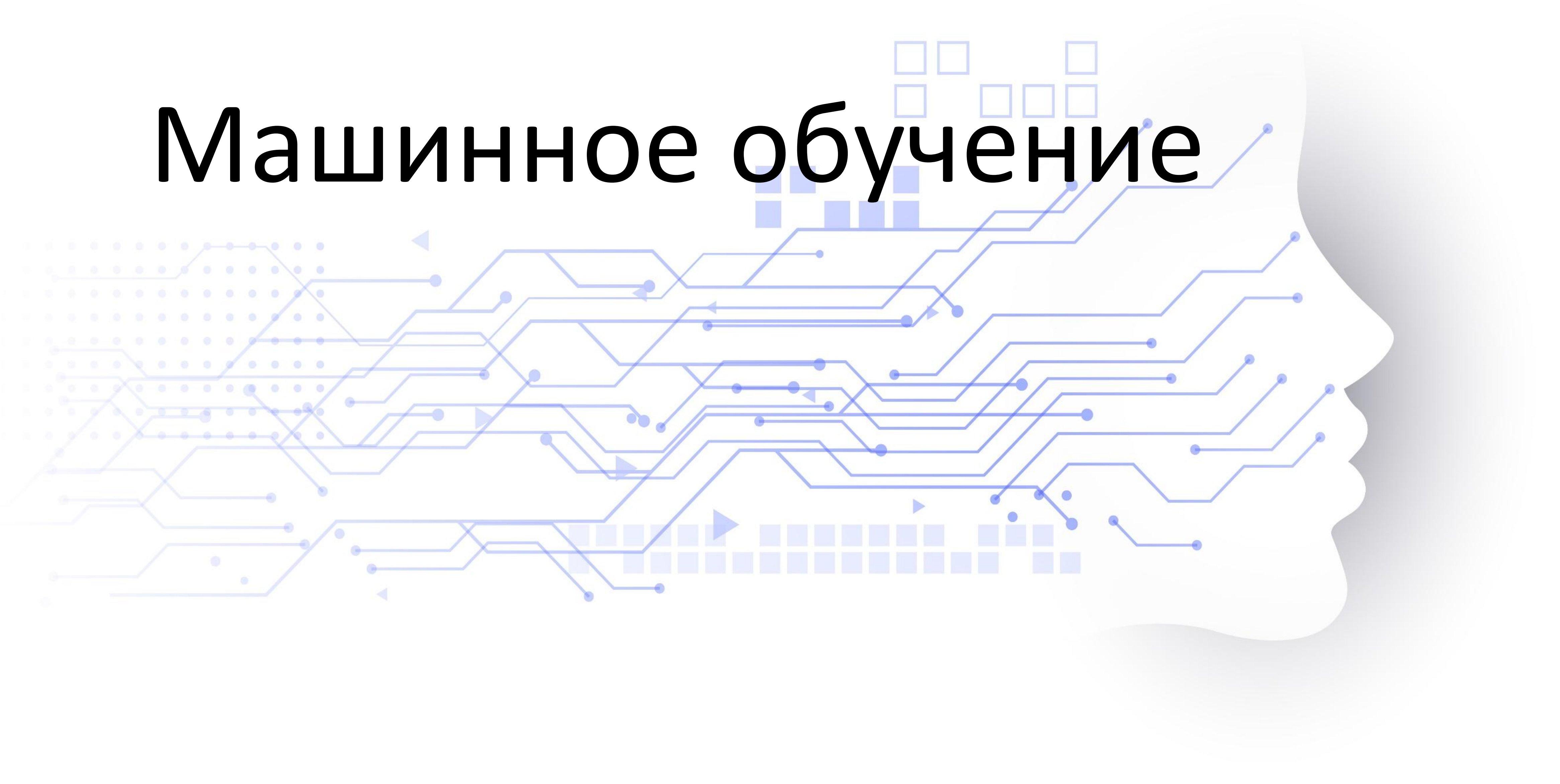
Машинное обучение



Общие понятия градиентного спуска

Градиентный спуск (Gradient descent)

Модель:

$$\boldsymbol{a}(\boldsymbol{x}) = \boldsymbol{\theta_0 x_0} + \boldsymbol{\theta_1 x_1} + \ldots + \boldsymbol{\theta_d x_d} = \langle \boldsymbol{\theta}, \boldsymbol{x} \rangle$$

Функционал ошибки

$$\mathrm{Q}(\mathrm{a}, \mathbb{X}) = \frac{\mathbf{1}}{\ell} \|\mathbf{X}\boldsymbol{\theta} - \boldsymbol{y}\|^2 \to \min_{\boldsymbol{\theta}}$$

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

Минимизация функции

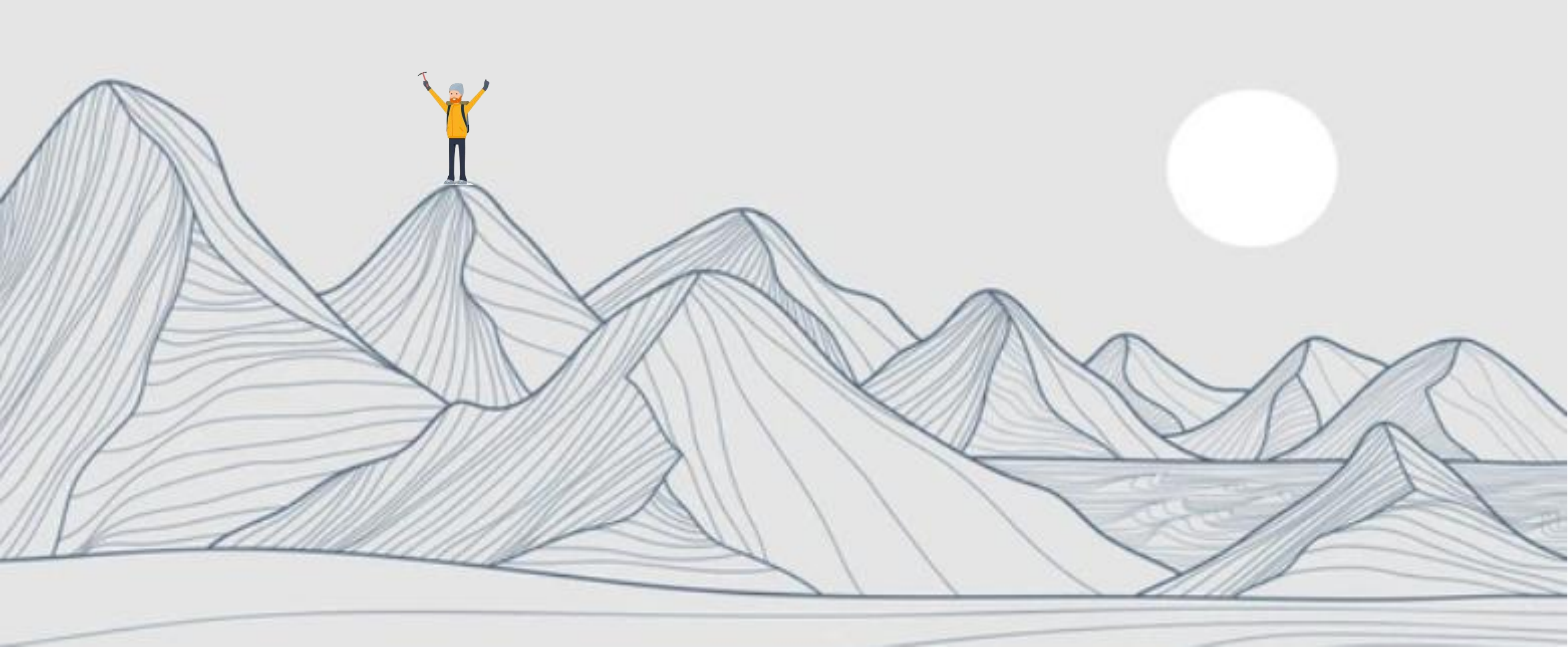
Задачи минимизации можно решать тремя способами:

1- подбором параметров вручную

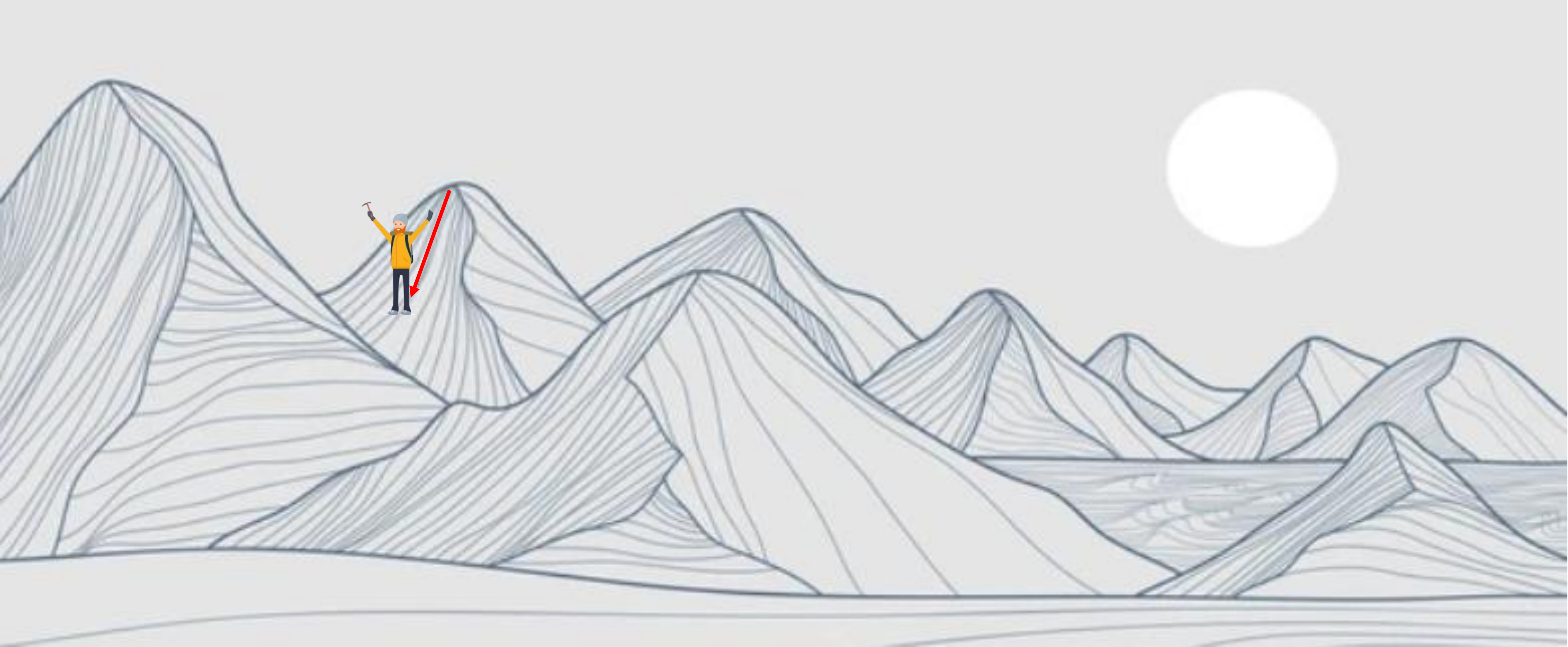
2- аналитически

3- численными методами

Градиентный спуск (Gradient descent)



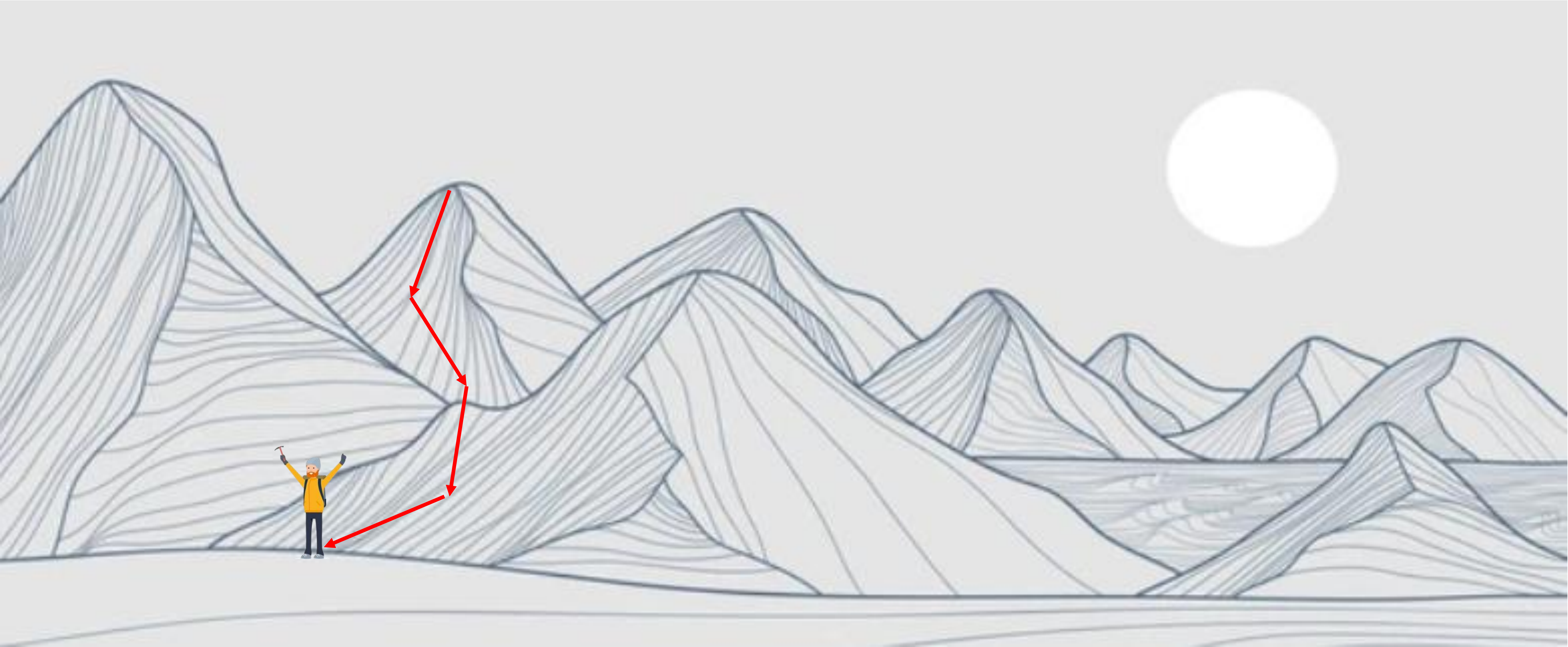
Градиентный спуск (Gradient descent)



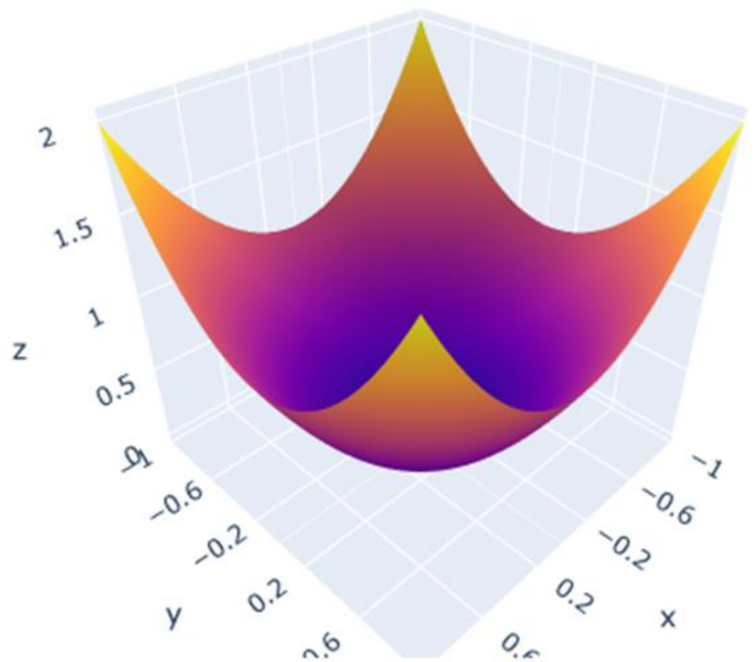
Градиентный спуск (Gradient descent)

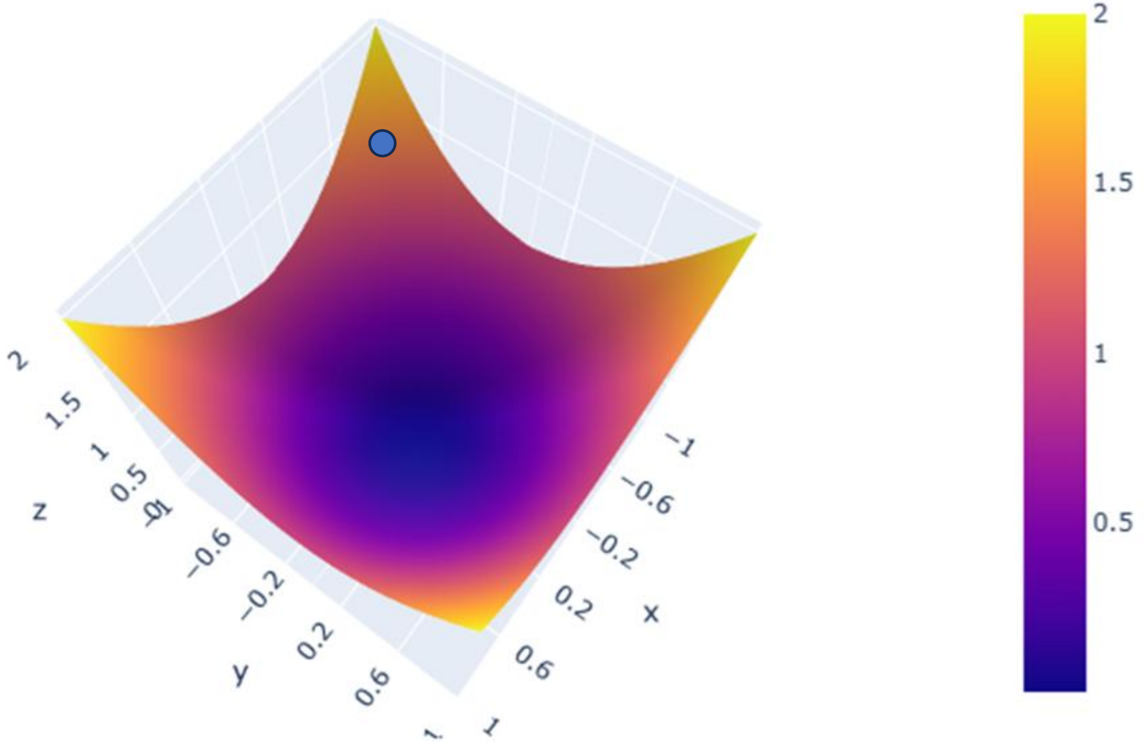
Градиентный спуск (Gradient descent)

Градиентный спуск (Gradient descent)

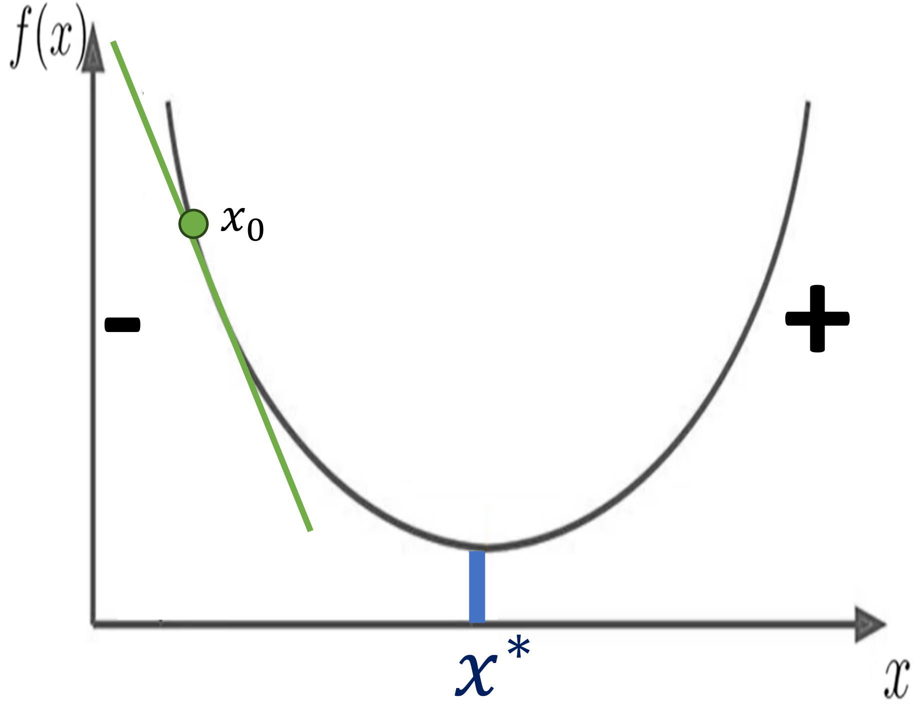


Градиентный спуск (Gradient descent)



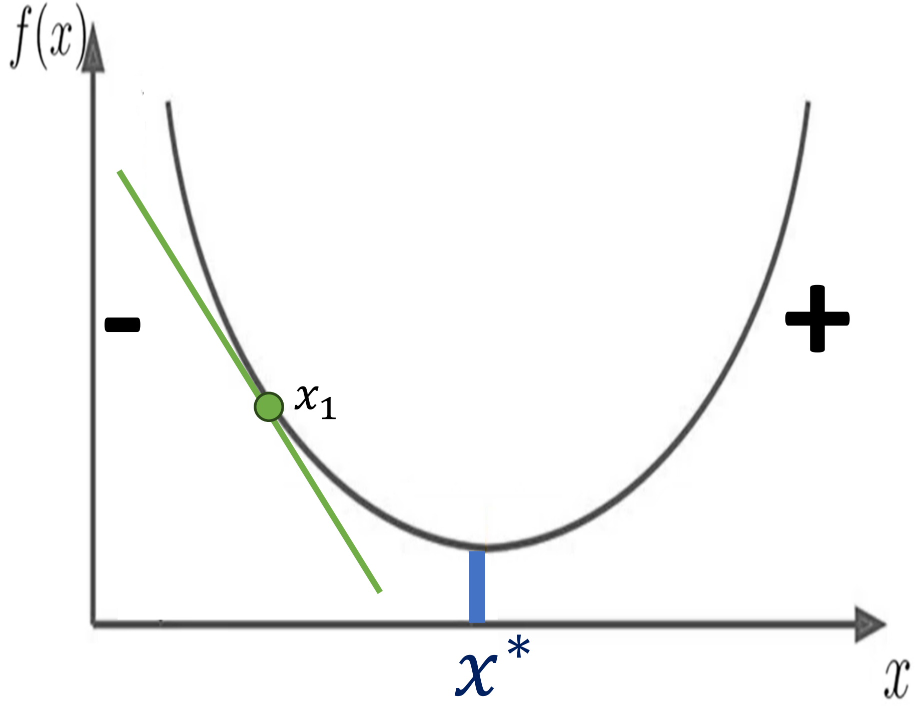


Эвристика градиентного спуска



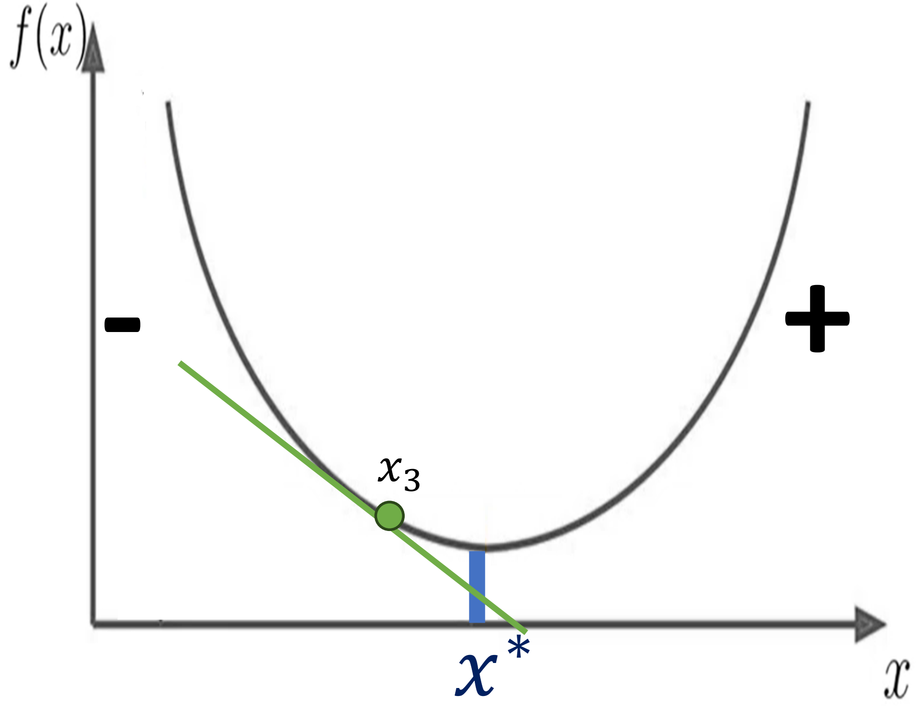
$$x_{n+1} = x_n - \alpha \frac{\partial f(x)}{dx} \quad , n = 0,1,2,3, \dots$$

Эвристика градиентного спуска



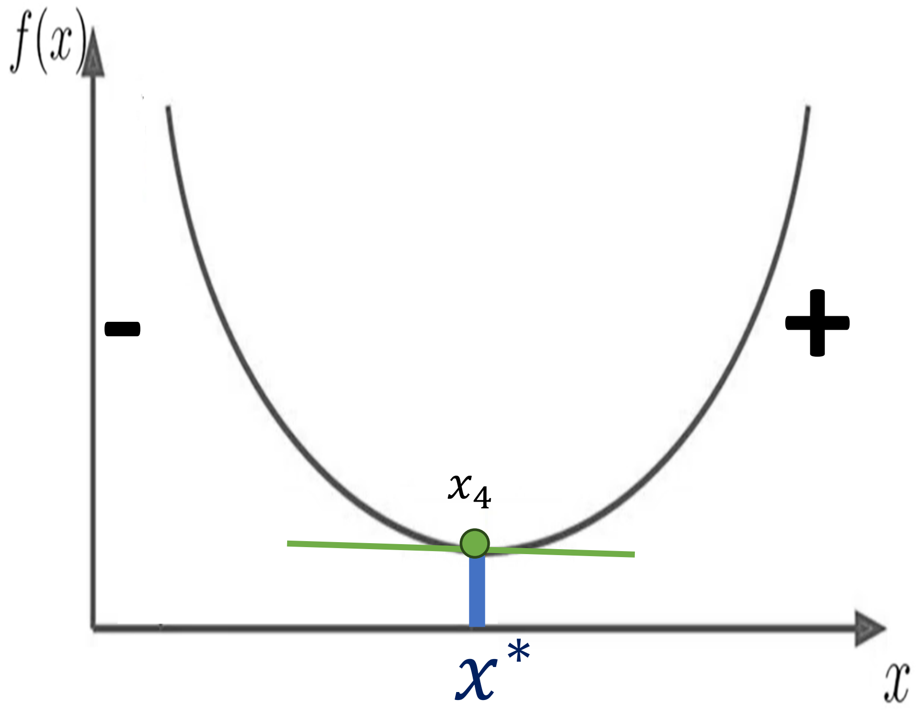
$$x_{n+1} = x_n - \alpha \frac{\partial f(x)}{dx} \quad , n = 0,1,2,3, \ldots$$

Эвристика градиентного спуска



$$x_{n+1} = x_n - \alpha \frac{\partial f(x)}{dx} \quad , n = 0,1,2,3, \ldots$$

Эвристика градиентного спуска



$$x_{n+1} = x_n - \alpha \frac{\partial f(x)}{dx} \quad , n = 0,1,2,3, \ldots$$

Градиентный спуск (Gradient descent)

Стартуем из случайной точки

Сдвигаемся по антиградиенту

Повторяем, пока не окажемся в точке минимума

$$\theta^t = \theta^{t-1} - \alpha \nabla Q(\theta^{t-1})$$

Новая точка

Размер шага

Градиент в предыдущей точке

Применение градиентного спуска в линейных задачах

Парная регрессия

Модель:

$$\boldsymbol{a}(\boldsymbol{x}) \ = \boldsymbol{\theta_1} \boldsymbol{x_1} + \boldsymbol{\theta_0}$$

Функционал ошибки

$$\mathrm{Q}(\boldsymbol{\theta_0}, \boldsymbol{\theta_1}) = \frac{\mathbf{1}}{\ell} \sum_{\boldsymbol{i}=\mathbf{1}}^{\ell} (\boldsymbol{a}(\boldsymbol{x_i}) - \boldsymbol{y_i})^{\mathbf{2}}$$

Парная регрессия

Модель:

$$\boldsymbol{a}(\boldsymbol{x}) = \boldsymbol{\theta_1} \boldsymbol{x_1} + \boldsymbol{\theta_0}$$

Функционал ошибки

$$\mathrm{Q}(\boldsymbol{\theta_0}, \boldsymbol{\theta_1}) = \frac{\mathbf{1}}{\ell} \sum_{\boldsymbol{i}=\mathbf{1}}^{\ell} (\boldsymbol{a}(\boldsymbol{x_i}) - \boldsymbol{y_i})^{\mathbf{2}}$$

Парная регрессия

Модель:

$$\boldsymbol{a}(\boldsymbol{x}) = \boldsymbol{\theta_1} \boldsymbol{x_1} + \boldsymbol{\theta_0}$$

Функционал ошибки

$$\mathrm{Q}(\boldsymbol{\theta_0}, \boldsymbol{\theta_1}) = \frac{\mathbf{1}}{\ell} \sum_{\boldsymbol{i}=\mathbf{1}}^{\ell} (\boldsymbol{a}(\boldsymbol{x_i}) - \boldsymbol{y_i})^{\mathbf{2}}$$

Парная регрессия

Модель:

$$\boldsymbol{a}(\boldsymbol{x}) \ = \boldsymbol{\theta}_{\mathbf{1}}\boldsymbol{x}_{\mathbf{1}} + \boldsymbol{\theta}_{\mathbf{0}}$$

Функционал ошибки

$$\mathrm{Q}(\boldsymbol{\theta}_{\mathbf{0}}, \boldsymbol{\theta}_{\mathbf{1}}) = \frac{\mathbf{1}}{\ell}\sum_{\boldsymbol{i}=\mathbf{1}}^{\ell}(\boldsymbol{a}(\boldsymbol{x}_{\boldsymbol{i}}) - \boldsymbol{y}_{\boldsymbol{i}})^{\mathbf{2}}$$

$$\boldsymbol{\nabla Q}(\boldsymbol{\theta}) = ?$$

Парная регрессия

Модель:

$$\boldsymbol{a(x)} = \boldsymbol{\theta_1 x_1} + \boldsymbol{\theta_0}$$

Функционал ошибки

$$\mathrm{Q}(\boldsymbol{\theta_0}, \boldsymbol{\theta_1}) = \frac{\mathbf{1}}{\ell}\sum_{\boldsymbol{i=1}}^{\ell}(\boldsymbol{a}(\boldsymbol{x_i}) - \boldsymbol{y_i})^{\mathbf{2}}$$

Градиент

$$\frac{\partial Q}{\partial \theta_0} = ?$$

$$\frac{\partial Q}{\partial \theta_1} = ?$$

Цепное правило в математике

это способ посчитать производную сложной функции, которая состоит из нескольких вложенных функций.

Формула:

$$\frac{\partial}{\partial x} f\big(g(x)\big) = f^{'}\big(g(x)\big).\, g^{'}(x)$$

Функция стоимости Q не зависит напрямую от параметров $\boldsymbol{\theta_0}, \boldsymbol{\theta_1}$

$\boldsymbol{\theta_0}, \boldsymbol{\theta_1} \;\rightarrow\; \boldsymbol{a_\theta}(\boldsymbol{x}) = \boldsymbol{\theta_0} + \boldsymbol{\theta_1}\boldsymbol{x}$ $\qquad$ $\boldsymbol{a_\theta}(\boldsymbol{x}) \rightarrow$ u = $\boldsymbol{a_\theta}(\boldsymbol{x}) - \mathrm{y}$ $\qquad$ u $\rightarrow$ Q $= \frac{1}{\ell}\sum u^2$

Это **цепочка функций**. Поэтому, чтобы найти $\frac{\partial J}{\partial \theta_0}$, мы применяем цепное правило:

- сначала берём производную Q по u,
- потом u по $\boldsymbol{a_\theta}$
- потом $\boldsymbol{a_\theta}$ по $\boldsymbol{\theta_0}$

$$\frac{\partial Q}{\partial \theta_0} = \frac{\partial Q}{\partial u} \cdot \frac{\partial u}{\partial a_\theta} \cdot \frac{\partial a_\theta}{\partial \theta_0}$$

$$\begin{cases} Q(\theta_0,\theta_1) = \frac{1}{\ell}\sum_{i=1}^{n}\left(a_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 \\ a_\theta(x) = \theta_0 + \theta_1 x_i \end{cases} \to \min_{\theta_0,\theta_1} Q$$

$$Q(\theta_0,\theta_1) = \frac{1}{\ell}\sum_{i=1}^{n} u^2 \to \begin{cases} \frac{\partial J}{\partial \theta_0} = ? \\ \frac{\partial J}{\partial \theta_1} = ? \end{cases}$$

$$\frac{\partial Q}{\partial u} = \frac{1}{\ell}\sum_{i=1}^{n} 2u = \frac{2}{\ell}\sum_{i=1}^{n} u = \frac{2}{\ell}\sum_{i=1}^{n}\left(a_\theta\left(x^{(i)}\right) - y^{(i)}\right)$$

$$\frac{\partial u}{\partial a_\theta} = \frac{\partial a_\theta\left(x^{(i)}\right) - y^{(i)}}{\partial a_\theta} = 1$$

$$\frac{\partial a_\theta}{\partial \theta_0} = \frac{\partial \theta_0 + \theta_1 x_i}{\partial \theta_0} = 1$$

$$\frac{\partial a_\theta}{\partial \theta_1} = \frac{\partial \theta_0 + \theta_1 x_i}{\partial \theta_1} = x_i$$

$$\frac{\partial Q}{\partial \theta_0} = \frac{\partial Q}{\partial u} \cdot \frac{\partial u}{\partial a_\theta} \cdot \frac{\partial a_\theta}{\partial \theta_0} = \frac{2}{\ell}\sum_{i=1}^{n}\left(a_\theta\left(x^{(i)}\right) - y^{(i)}\right) . 1 . 1$$

$$\frac{\partial Q}{\partial \theta_1} = \frac{\partial Q}{\partial u} \cdot \frac{\partial u}{\partial a_\theta} \cdot \frac{\partial a_\theta}{\partial \theta_1} = \frac{2}{\ell}\sum_{i=1}^{n}\left(a_\theta\left(x^{(i)}\right) - y^{(i)}\right) . 1 . x_i$$

Парная регрессия

Модель:

$$\boldsymbol{a}(\boldsymbol{x}) = \boldsymbol{\theta_1 x_1} + \boldsymbol{\theta_0}$$

Функционал ошибки

$$\mathrm{Q}(\boldsymbol{\theta_0}, \boldsymbol{\theta_1}) = \frac{\mathbf{1}}{\ell} \sum_{\boldsymbol{i}=\mathbf{1}}^{\ell} (\boldsymbol{a}(\boldsymbol{x_i}) - \boldsymbol{y_i})^{\mathbf{2}}$$

Градиент

$$\frac{\partial Q}{\partial \theta_0} = \frac{2}{\ell} \sum_{i=1}^{\ell} (\boldsymbol{\theta_1 x_1} + \boldsymbol{\theta_0} - \boldsymbol{y_i})$$

$$\frac{\partial Q}{\partial \theta_1} = ?$$

Парная регрессия

Модель:

$$\boldsymbol{a}(\boldsymbol{x}) = \boldsymbol{\theta_1 x_1} + \boldsymbol{\theta_0}$$

Функционал ошибки

$$\mathrm{Q}(\boldsymbol{\theta_0}, \boldsymbol{\theta_1}) = \frac{\mathbf{1}}{\ell} \sum_{\boldsymbol{i}=\mathbf{1}}^{\ell} (\boldsymbol{a}(\boldsymbol{x_i}) - \boldsymbol{y_i})^{\mathbf{2}}$$

Градиент

$$\frac{\partial Q}{\partial \theta_0} = \frac{2}{\ell} \sum_{i=1}^{\ell} (\boldsymbol{\theta_1 x_1} + \boldsymbol{\theta_0} - \boldsymbol{y_i})$$

$$\frac{\partial Q}{\partial \theta_1} = \frac{2}{\ell} \sum_{i=1}^{\ell} x_i(\boldsymbol{\theta_1 x_1} + \boldsymbol{\theta_0} - \boldsymbol{y_i})$$

Парная регрессия

Модель:

$$\boldsymbol{a}(\boldsymbol{x}) \ = \boldsymbol{\theta_1 x_1} + \boldsymbol{\theta_0}$$

Функционал ошибки

$$\mathrm{Q}(\boldsymbol{\theta_0}, \boldsymbol{\theta_1}) = \frac{\mathbf{1}}{\ell}\sum_{\boldsymbol{i=1}}^{\ell}(\boldsymbol{a}(\boldsymbol{x_i}) - \boldsymbol{y_i})^{\mathbf{2}}$$

Градиент

$$\frac{\partial Q}{\partial \theta_0} = \frac{2}{\ell}\sum_{i=1}^{\ell}(\boldsymbol{\theta_1 x_1} + \boldsymbol{\theta_0} - \boldsymbol{y_i})$$

$$\frac{\partial Q}{\partial \theta_1} = \frac{2}{\ell}\sum_{i=1}^{\ell}x_i(\boldsymbol{\theta_1 x_1} + \boldsymbol{\theta_0} - \boldsymbol{y_i})$$

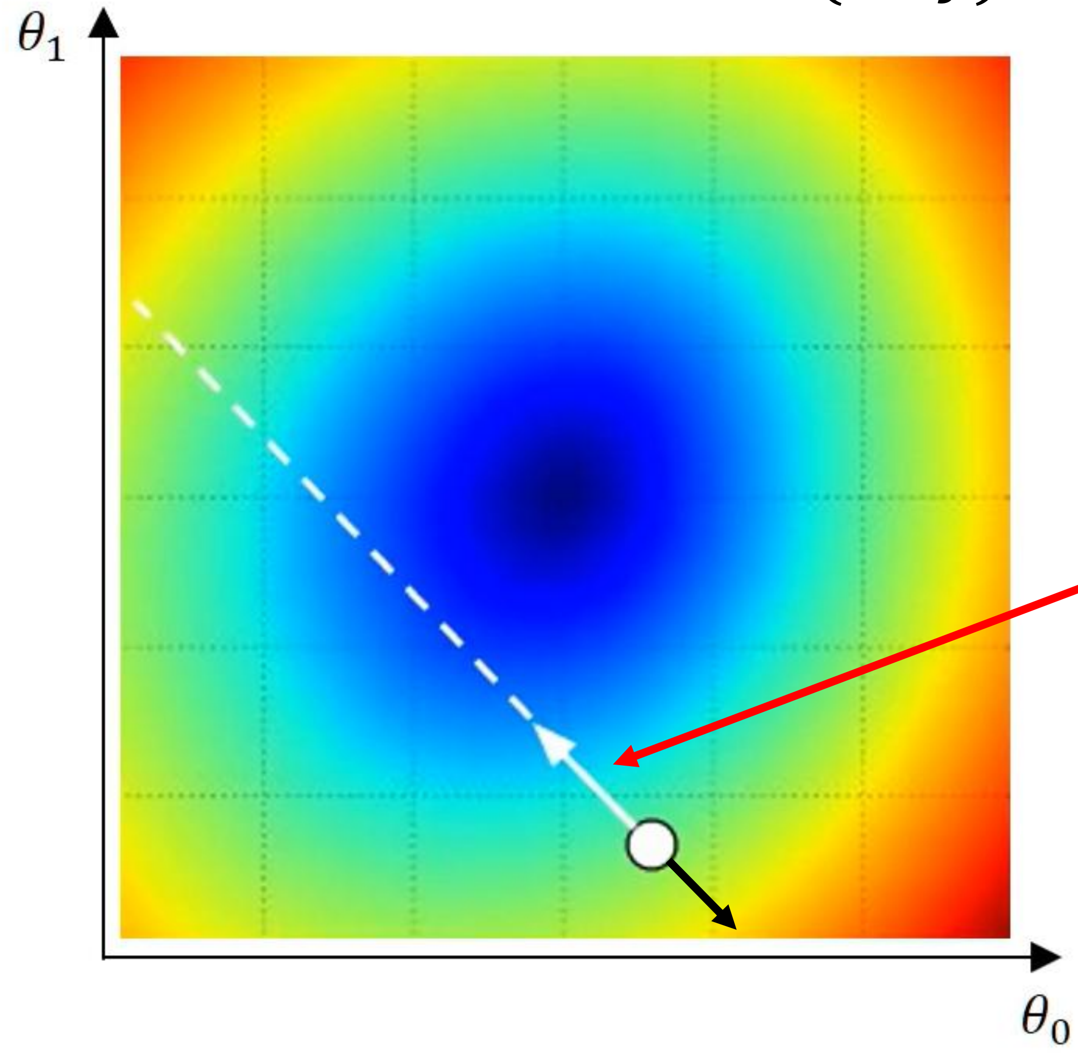
$$\boldsymbol{\nabla Q}(\boldsymbol{\theta}) = \left(\frac{\mathbf{2}}{\ell}\sum_{\boldsymbol{i=1}}^{\ell}(\boldsymbol{\theta_1 x_1} + \boldsymbol{\theta_0} - \boldsymbol{y_i}), \frac{\mathbf{2}}{\ell}\sum_{\boldsymbol{i=1}}^{\ell}\boldsymbol{x_i}(\boldsymbol{\theta_1 x_1} + \boldsymbol{\theta_0} - \boldsymbol{y_i})\right)$$

Градиентный спуск (Gradient descent)

$$\theta^{(new)} = \theta^{(old)} + \alpha(-\nabla J)$$

$$\mathbf{Q}(\boldsymbol{\theta_0}, \boldsymbol{\theta_1}) = \mathrm{J}(\theta)$$

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta_0} \\ \boldsymbol{\theta_1} \end{bmatrix}$$



$$-\boldsymbol{\nabla}\mathrm{Q} = \begin{bmatrix} -\dfrac{\partial Q}{\partial \theta_0} \\ -\dfrac{\partial Q}{\partial \theta_1} \end{bmatrix}$$

$$\boldsymbol{\nabla}\mathrm{Q} = \begin{bmatrix} \dfrac{\partial Q}{\partial \theta_0} \\ \dfrac{\partial Q}{\partial \theta_1} \end{bmatrix}$$

Градиентный спуск (Gradient descent)

1. Стартуем из случайной точки

$$\theta \;\; - \text{Инициализация весов}$$

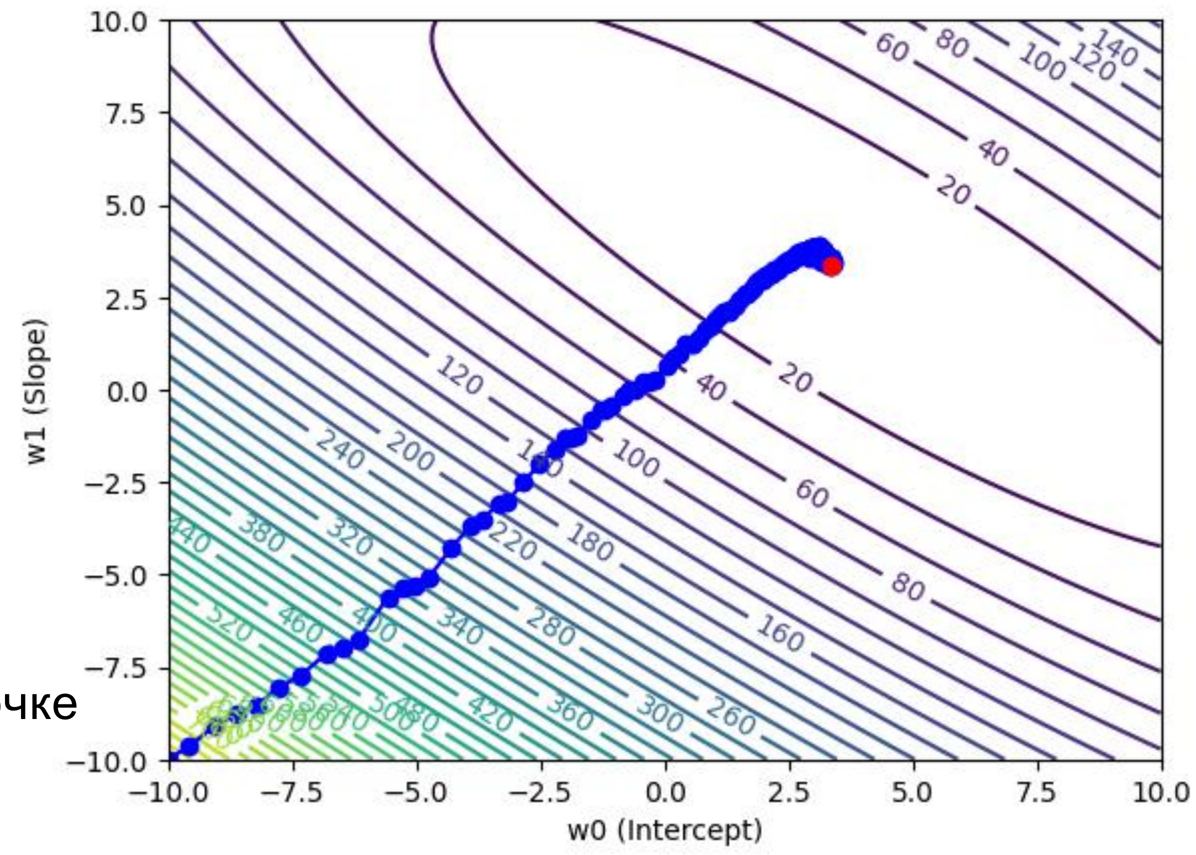
2. Сдвигаемся по антиградиенту

$$\theta^t = \theta^{t-1} - \alpha \nabla Q(\theta^{t-1})$$

Новая точка

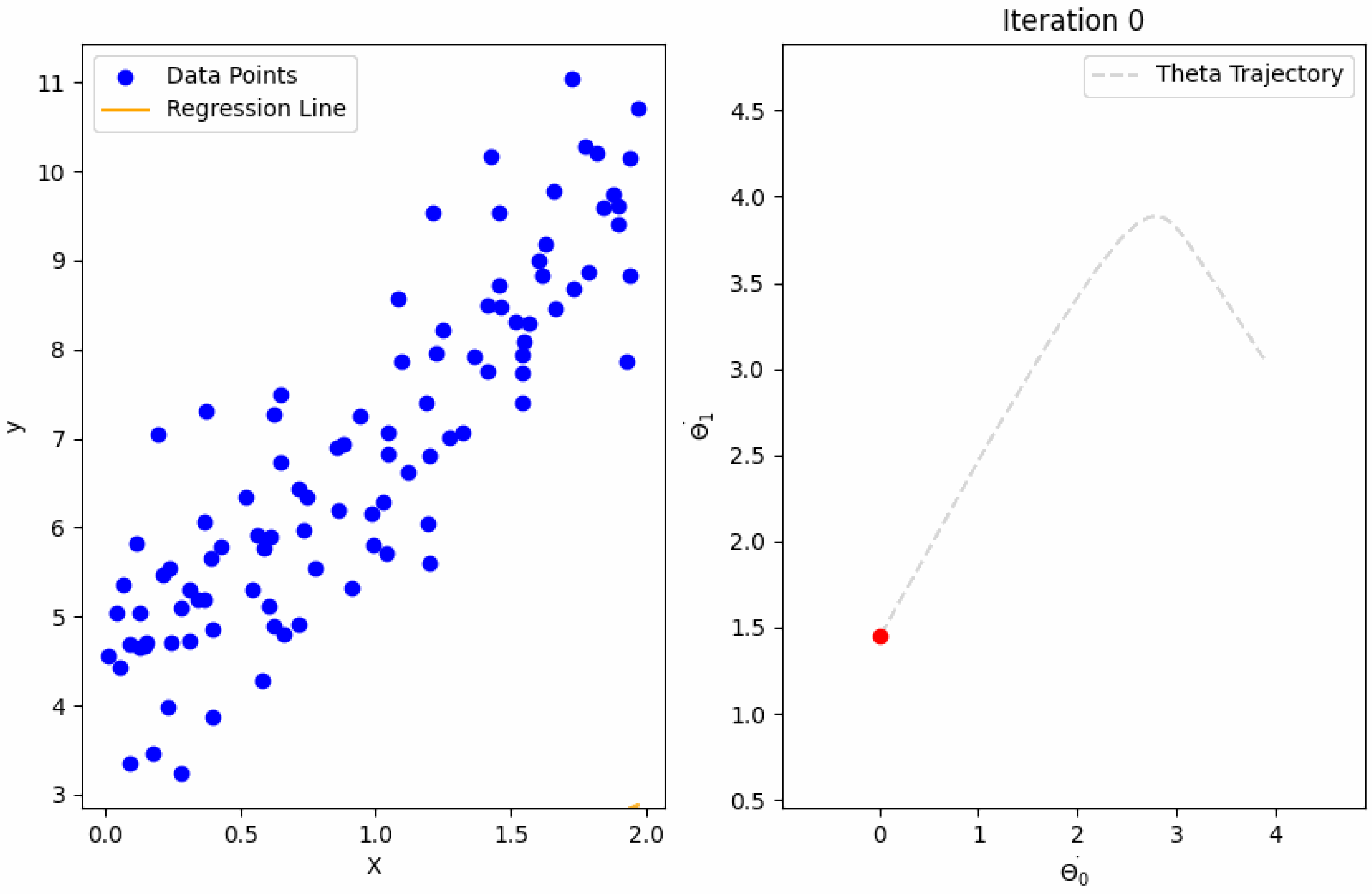
Размер шага

Градиент в предыдущей точке



3. Повторяем до сходимости

Градиентный спуск



Градиентный спуск для поиска минимума функции многих переменных

$$\mathrm{Q}(\theta_1, \theta_1, \dots, \theta_d) = \frac{\mathbf{1}}{\ell}\sum_{\boldsymbol{i=1}}^{\ell}(\boldsymbol{a}(\boldsymbol{x_i}) - \boldsymbol{y_i})^{\mathbf{2}}$$

Чтобы применять метод градиентного спуска, необходимо вычислять градиент функции в точке:

$$\nabla f(\theta_1, \theta_1, \dots, \theta_d) = (\frac{\partial f}{\theta_1}, \frac{\partial f}{\theta_2}, \dots, \frac{\partial f}{\theta_d})$$

На каждом шаге будем менять все переменные, от которых зависит функция:

$$\theta_1 = \theta_1 - \boldsymbol{\alpha}\frac{\partial f}{d\theta_1}$$

$$\dots$$

$$\theta_d = \theta_d - \boldsymbol{\alpha}\frac{\partial f}{d\theta_n}$$

Повторяем пока изменение не будет достаточно маленьким, пройдет много итераций или до сходимости

Градиентный спуск в вектором виде

Модель:

$$\boldsymbol{a}(\boldsymbol{x}) = \boldsymbol{\theta_0 x_0} + \boldsymbol{\theta_1 x_1} + \ldots + \boldsymbol{\theta_d}\, \boldsymbol{x_d} = \langle \boldsymbol{\theta}, \boldsymbol{x} \rangle$$

Градиентный спуск в векторном виде

Модель:

$$\boldsymbol{a}(\boldsymbol{x}) = \boldsymbol{\theta_0 x_0} + \boldsymbol{\theta_1 x_1} + \ldots + \boldsymbol{\theta_d\, x_d} = \langle\boldsymbol{\theta}, \boldsymbol{x}\rangle$$

Функционал ошибки

$$\mathrm{Q}(\boldsymbol{\theta}) = \frac{\mathbf{1}}{\boldsymbol{\ell}}\|\mathbf{X}\boldsymbol{\theta} - \boldsymbol{y}\|^2$$

Градиент

$$\nabla_{\boldsymbol{\theta}}\boldsymbol{Q}(\boldsymbol{\theta}) = \frac{\mathbf{2}}{\boldsymbol{\ell}}\boldsymbol{X}^{T}(\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y})$$

Правило обновления

$$\boldsymbol{\theta} : \boldsymbol{\theta} - \boldsymbol{\alpha} \, . \, \boldsymbol{\nabla_{\theta} Q}(\boldsymbol{\theta})$$

Градиентный спуск в вектором виде

Градиент

$$\boldsymbol{\nabla Q(\theta) = \frac{2}{\ell} X^T (\mathbf{X}\theta - y)}$$

1. θ $-$ Инициализация весов

2. $\theta^t = \theta^{t-1} - \alpha \nabla Q(\theta^{t-1})$

3. Повторяем до сходимости

Критерии останова

1. Количество итераций

2. Останавливаем процесс, если

$$\|\theta^t - \theta^{t-1}\| < \varepsilon$$

3. Другой вариант

$$\|\nabla Q(\theta^t)\| < \varepsilon$$

4. Другой вариант

Размер шага или скорость обучения

Градиентный шаг (learning rate)

Модель:

$$\boldsymbol{a(x) = \theta_0 x_0 + \theta_1 x_1 + \ldots + \theta_d\, x_d = \langle \theta, x \rangle}$$

Функционал ошибки

$$\mathrm{Q}(\boldsymbol{\theta}) = \frac{\mathbf{1}}{\boldsymbol{\ell}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{\ell}} (\langle \boldsymbol{\theta}, \boldsymbol{x} \rangle - \boldsymbol{y_i})^{\mathbf{2}}$$

Парная регрессия

Градиент

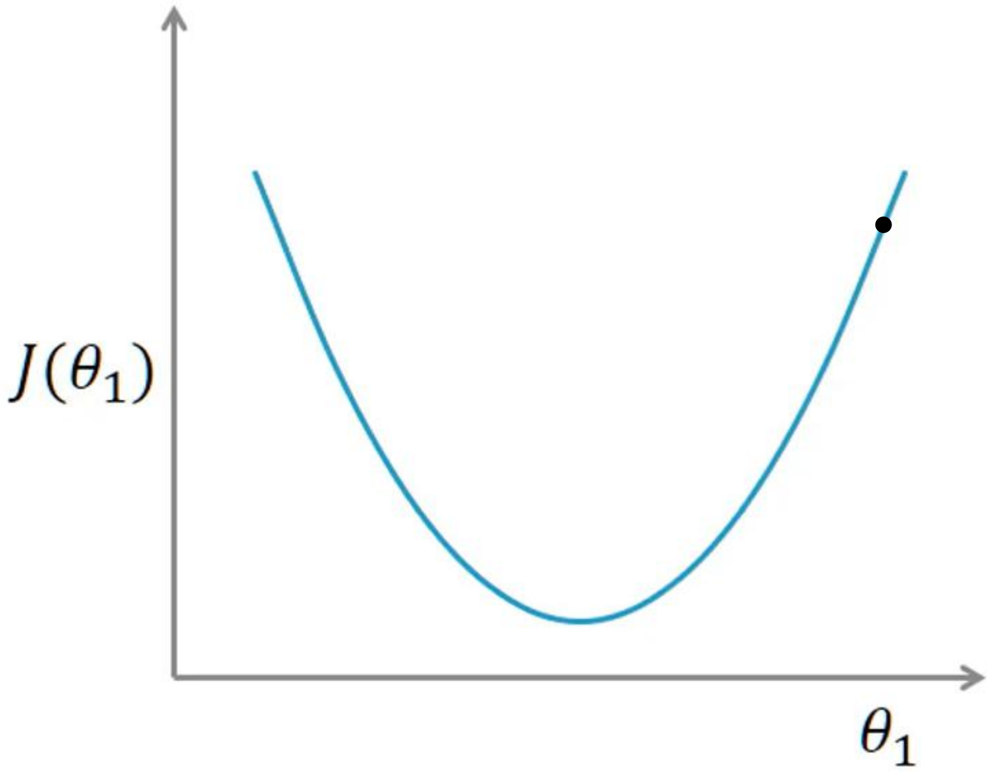
$$\boldsymbol{\nabla Q(\theta) = \frac{2}{\ell} X^T (\mathrm{X}\theta - y)}$$

$$\theta^t = \theta^{t-1} - \alpha \nabla Q(\theta^{t-1})$$

Размер шага

Размер шага (learning rate)

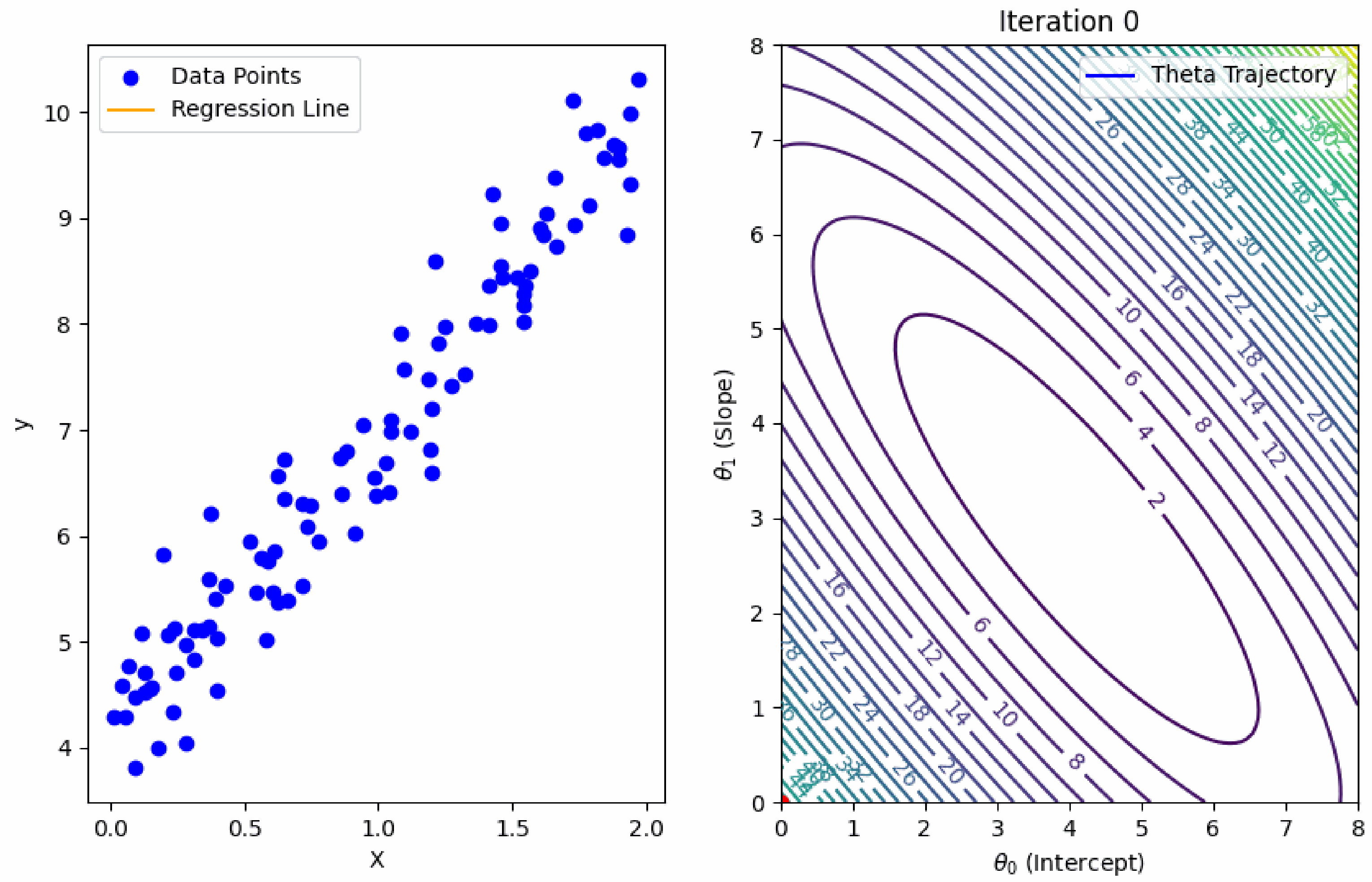
Если размер шага (learning rate) слишком мал, градиентный спуск сходится слишком медленно или не доходит до точки минимума



$$\theta^t = \theta^{t-1} - \alpha \nabla Q(\theta^{t-1})$$

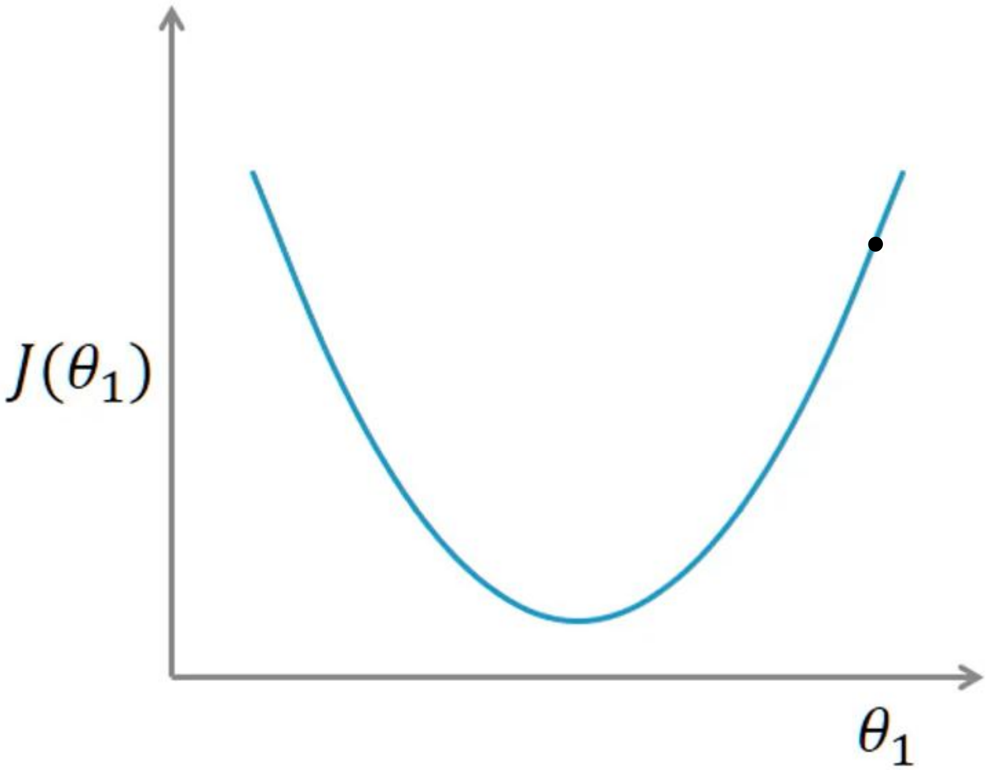
Размер шага (learning rate)

Если размер шага (learning rate) слишком мал, градиентный спуск сходится слишком медленно или не доходит до точки минимума



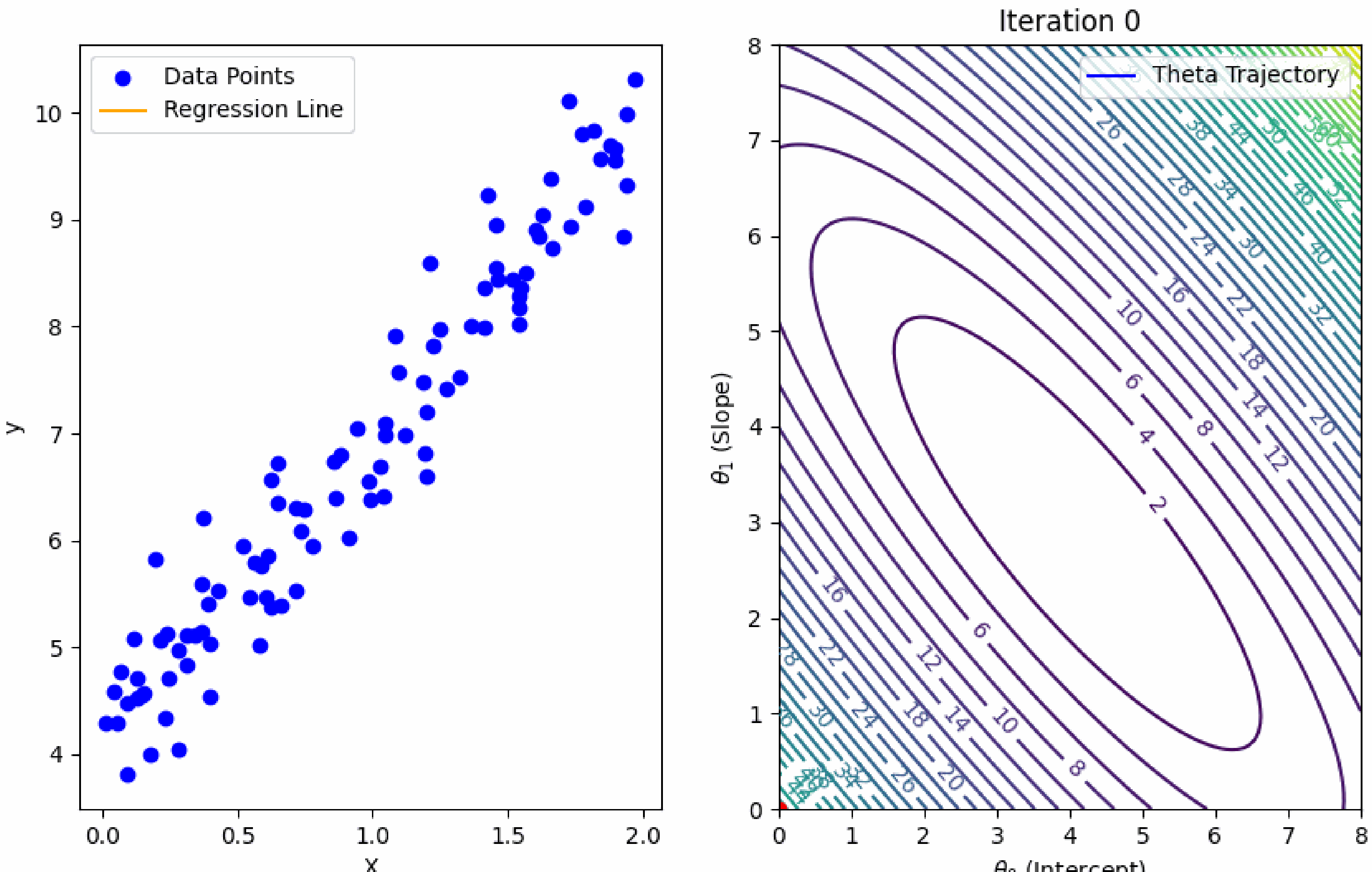
Градиентный спуск (Gradient descent)

Если размер шага (learning rate) большой, градиентный спуск сходится слишком медленно или вообще не сходится



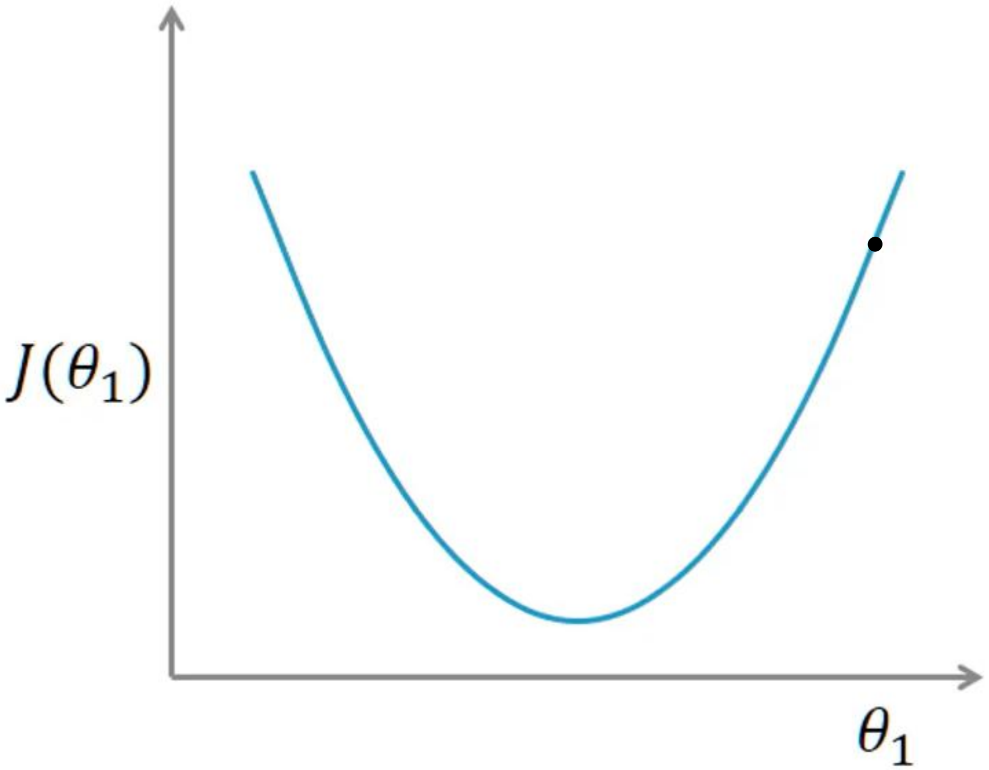
Градиентный спуск (Gradient descent)

Если размер шага (learning rate) большой, градиентный спуск сходится слишком медленно



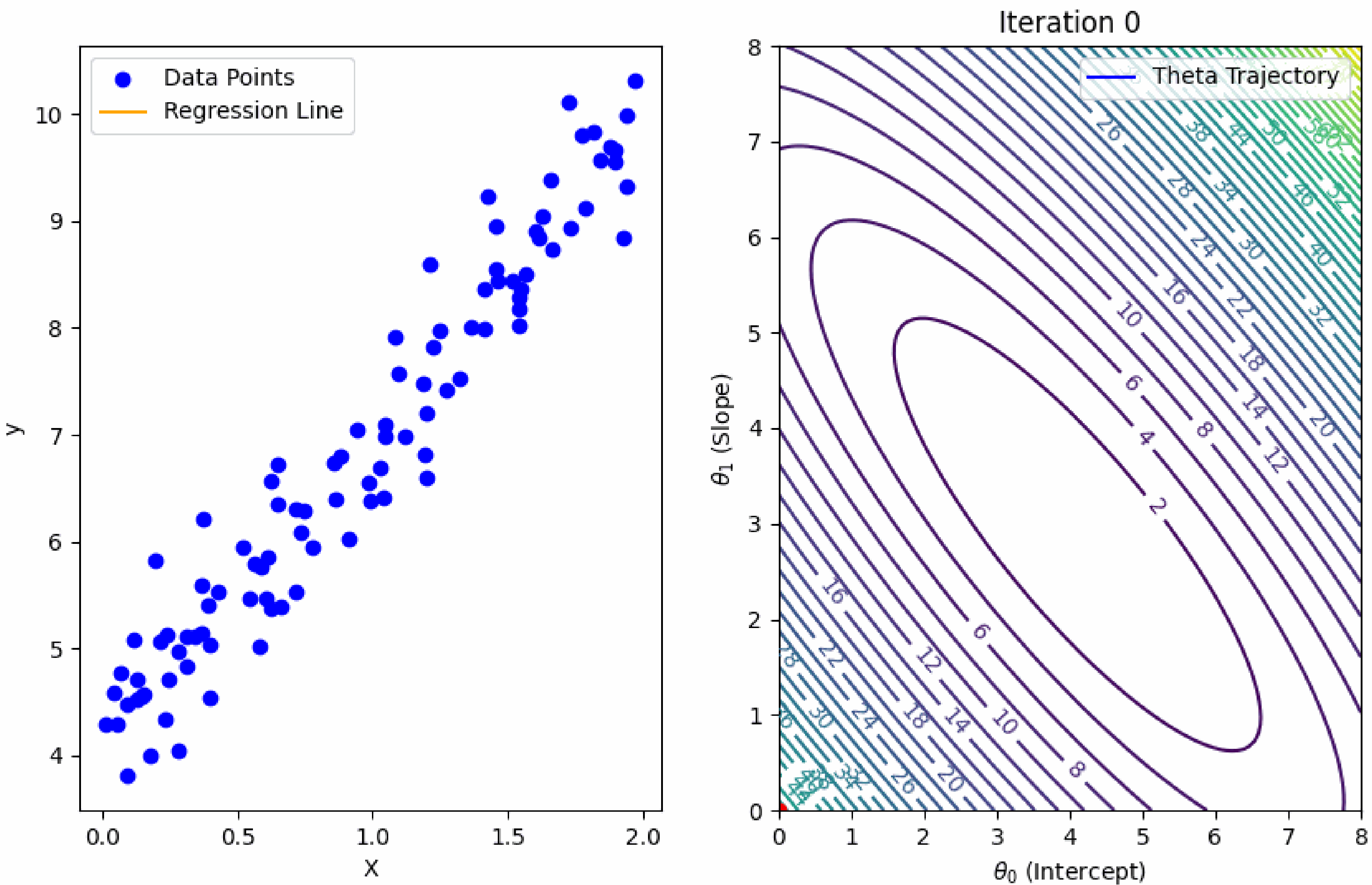
Градиентный спуск (Gradient descent)

Если размер шага (learning rate) слишком большой, градиентный спуск вообще не сходится



Градиентный спуск (Gradient descent)

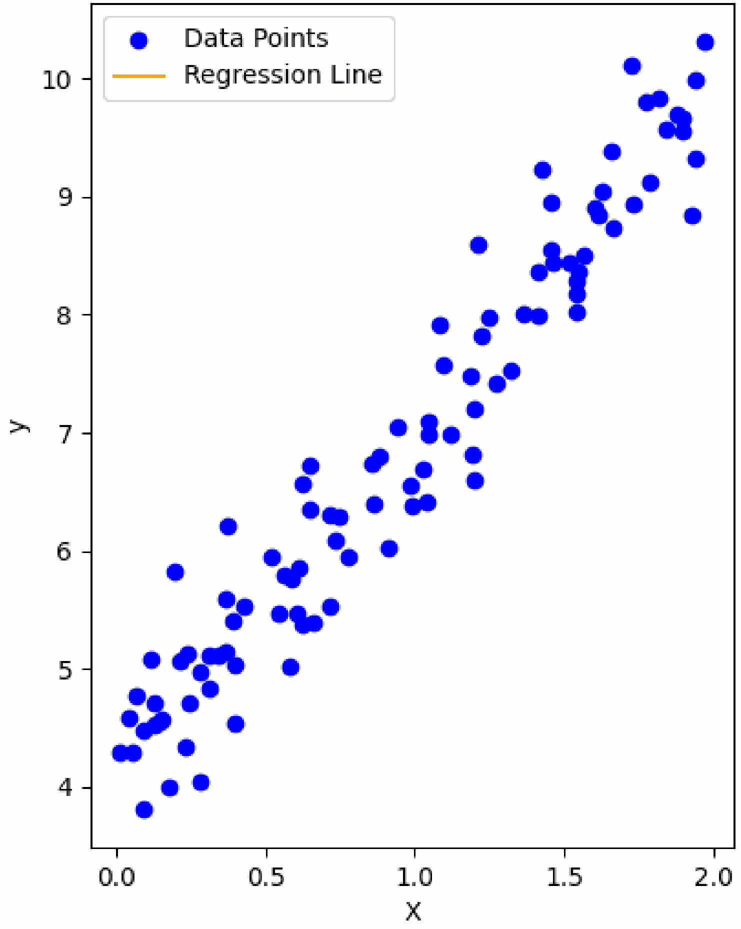
Если размер шага (learning rate) слишком большой, градиентный спуск вообще не сходится.

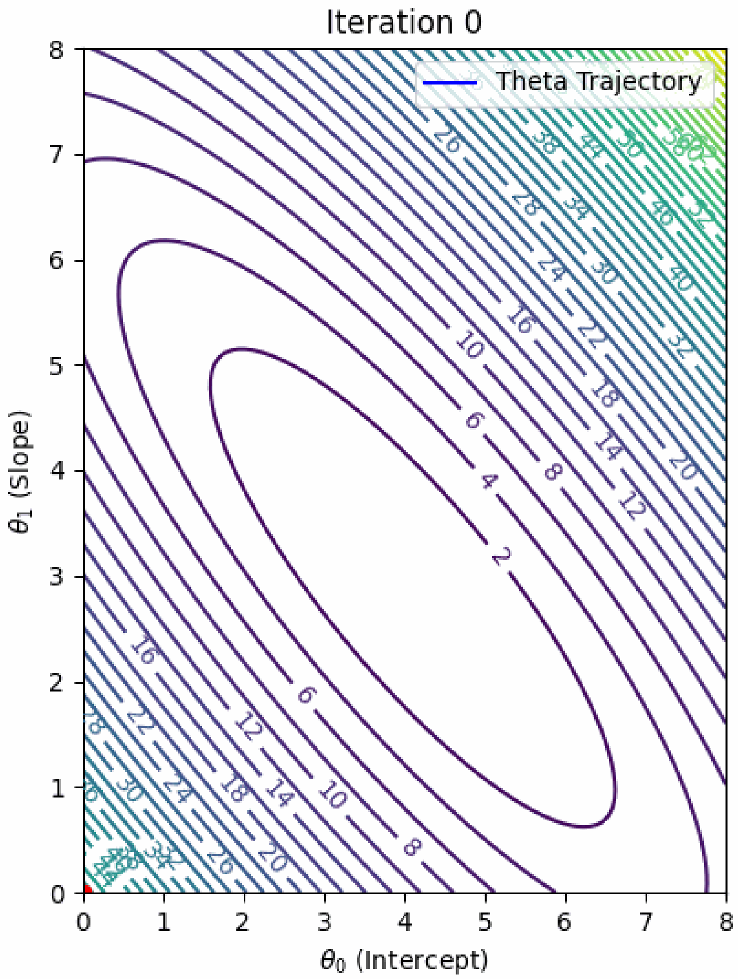


Градиентный спуск (Gradient descent)

Длину шага можно менять в зависимости от шага

$$\theta^t = \theta^{t-1} - \alpha \nabla J(\theta^{t-1}) \qquad \alpha_t = \frac{1}{t}$$



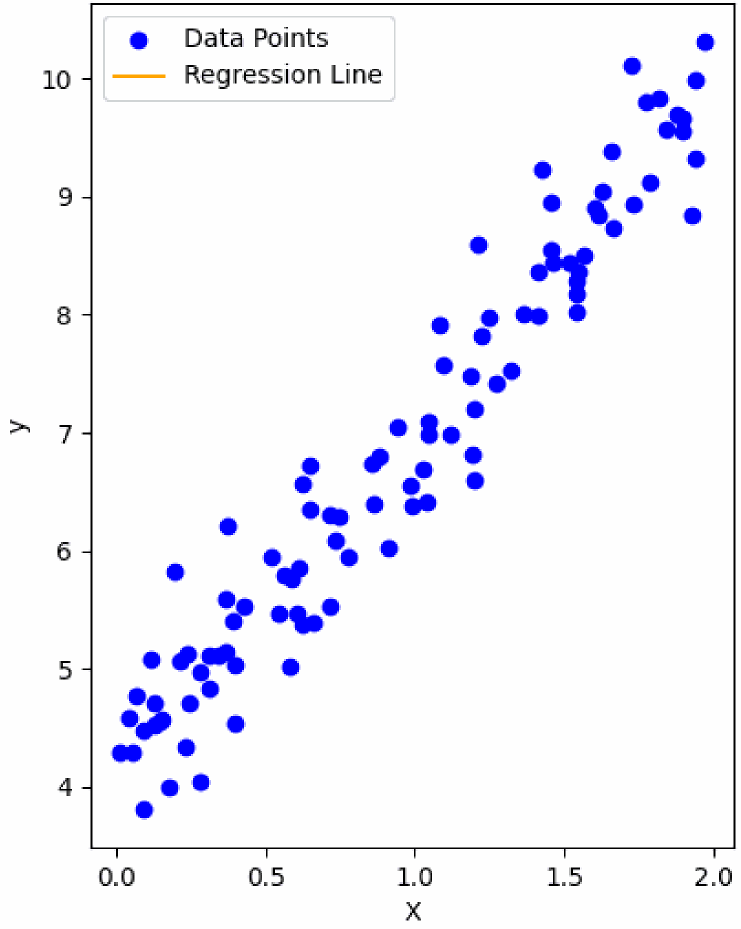


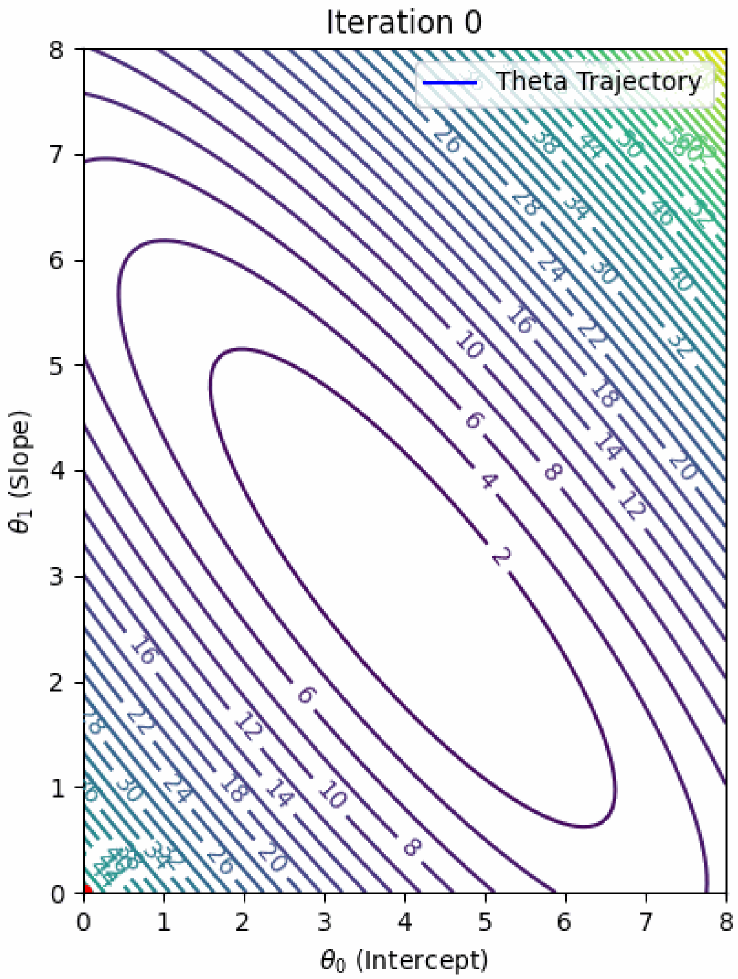
Градиентный спуск (Gradient descent)

Еще вариант

$\theta^t = \theta^{t-1} - \alpha \nabla J(\theta^{t-1})$

$$\alpha_t = \frac{0.1}{t^{\beta}}$$





Стохастический градиентный спуск

Линейная регрессия

Модель:

$$\boldsymbol{a}(\boldsymbol{x}) = \boldsymbol{\theta_1 x_1} + \boldsymbol{\theta_0}$$

Функционал ошибки

$$\mathrm{Q}(\boldsymbol{\theta}) = \frac{\mathbf{1}}{\ell}\sum_{\boldsymbol{i}=\mathbf{1}}^{\ell}(\langle\boldsymbol{\theta}, \boldsymbol{x}\rangle - \boldsymbol{y_i})^{\mathbf{2}}$$

Градиент

$$\boldsymbol{\nabla Q}(\boldsymbol{\theta}) = \frac{\mathbf{2}}{\ell}\boldsymbol{X^T}(\mathbf{X}\boldsymbol{\theta} - \boldsymbol{y})$$

Стохастический градиентный спуск (Stochastic gradient descent)

Для вычисления градиента, как правило, надо просуммировать что-то по всем объектам для одного маленького шага

Градиент:

$$\boldsymbol{\nabla Q(\theta) = \frac{1}{\ell}\sum_{i=1}^{\ell} \nabla L(y_i, a(x_i))}$$

Можно найти градиент по одному слагаемому

$$\nabla J(\theta) \cong \nabla L(y_i, a(x_i))$$

Стохастический градиентный спуск (Stochastic gradient descent)

1. Стартуем из случайной точки

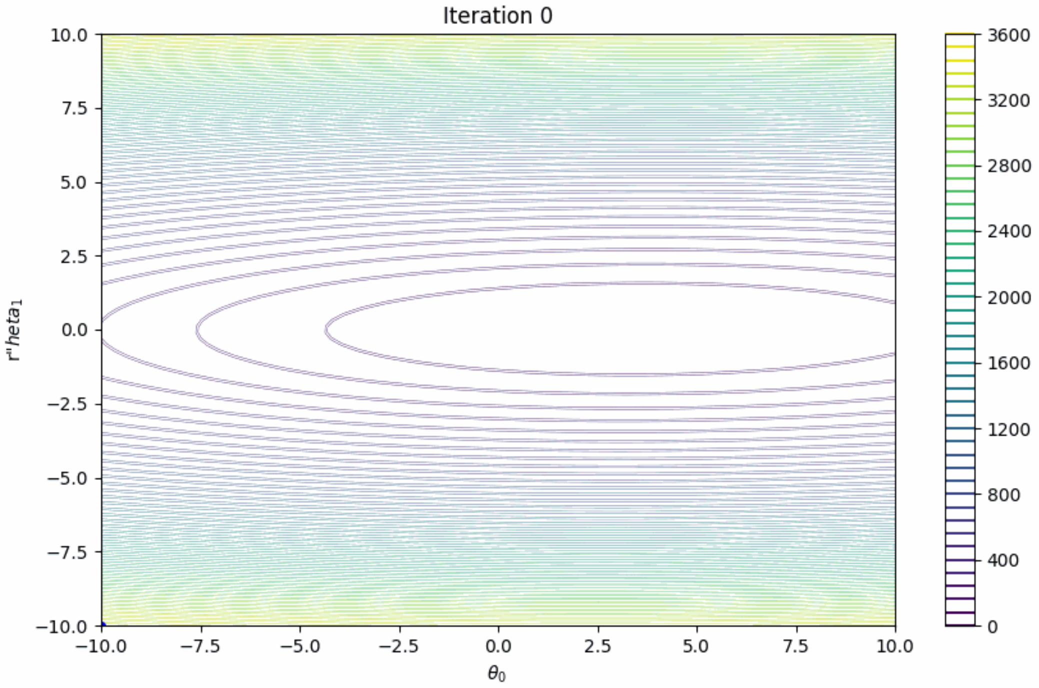
$$\theta \ - \text{Инициализация весов}$$

2. Сдвигаемся по антиградиенту, каждый раз выбираем случайный объект i_t

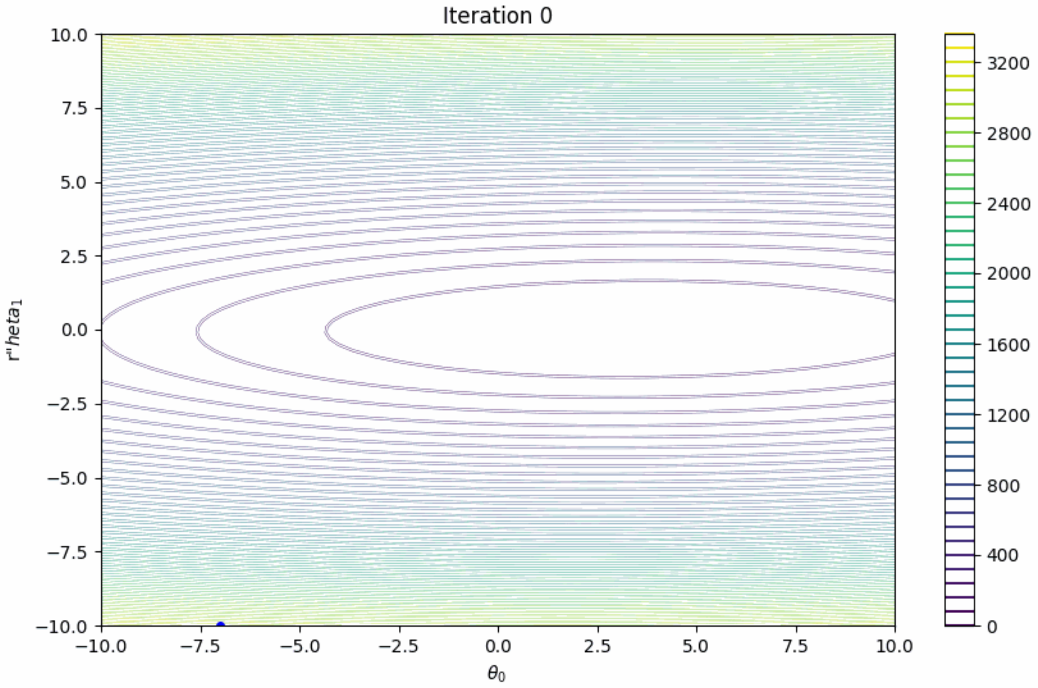
$$\theta^t = \theta^{t-1} - \alpha \nabla L(y_{i_t}, a(x_{i_t}))$$

3. Повторяем пока не окажемся в точке минимума

Стохастический градиентный спуск (Stochastic gradient descent)

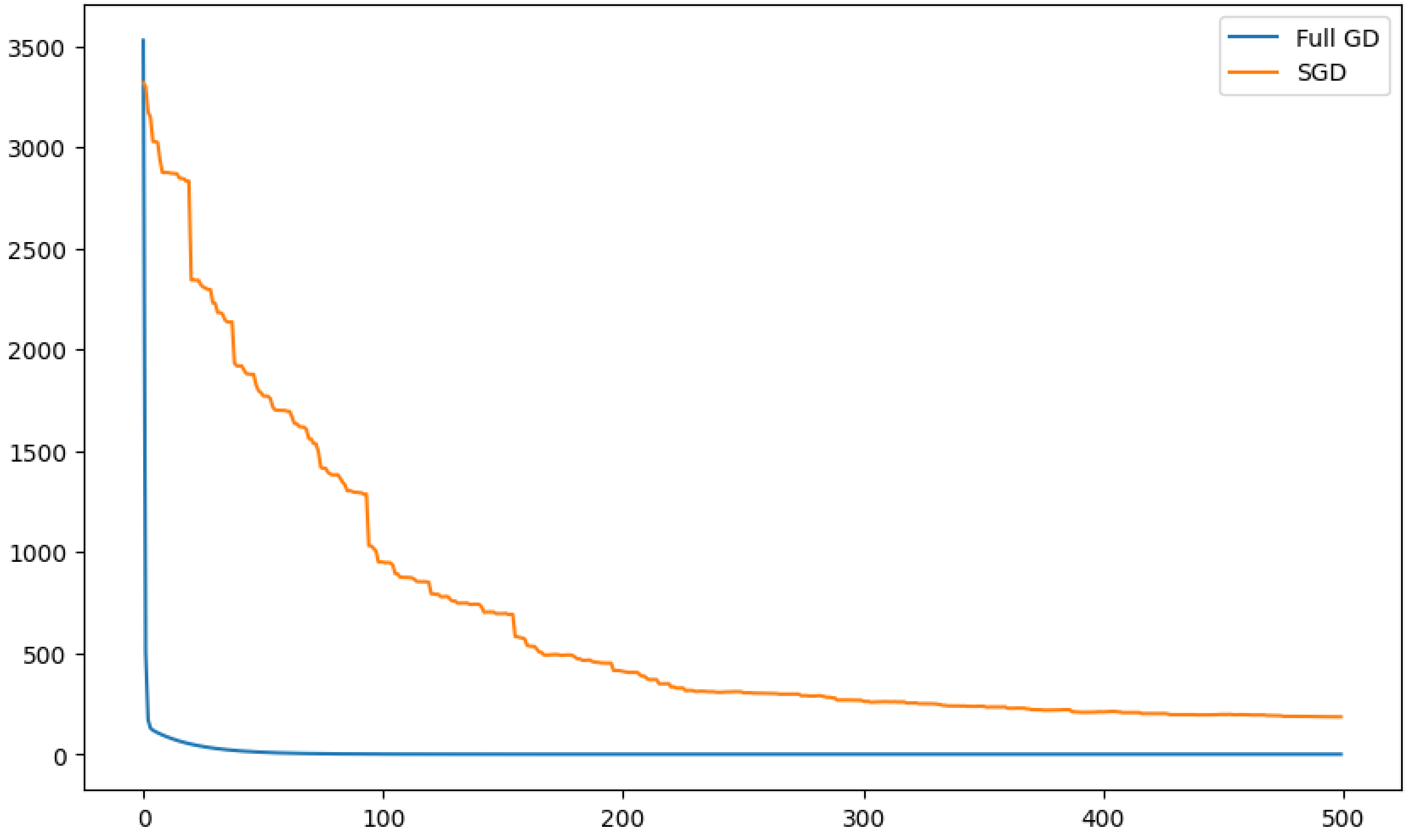


Stochastic gradient descent

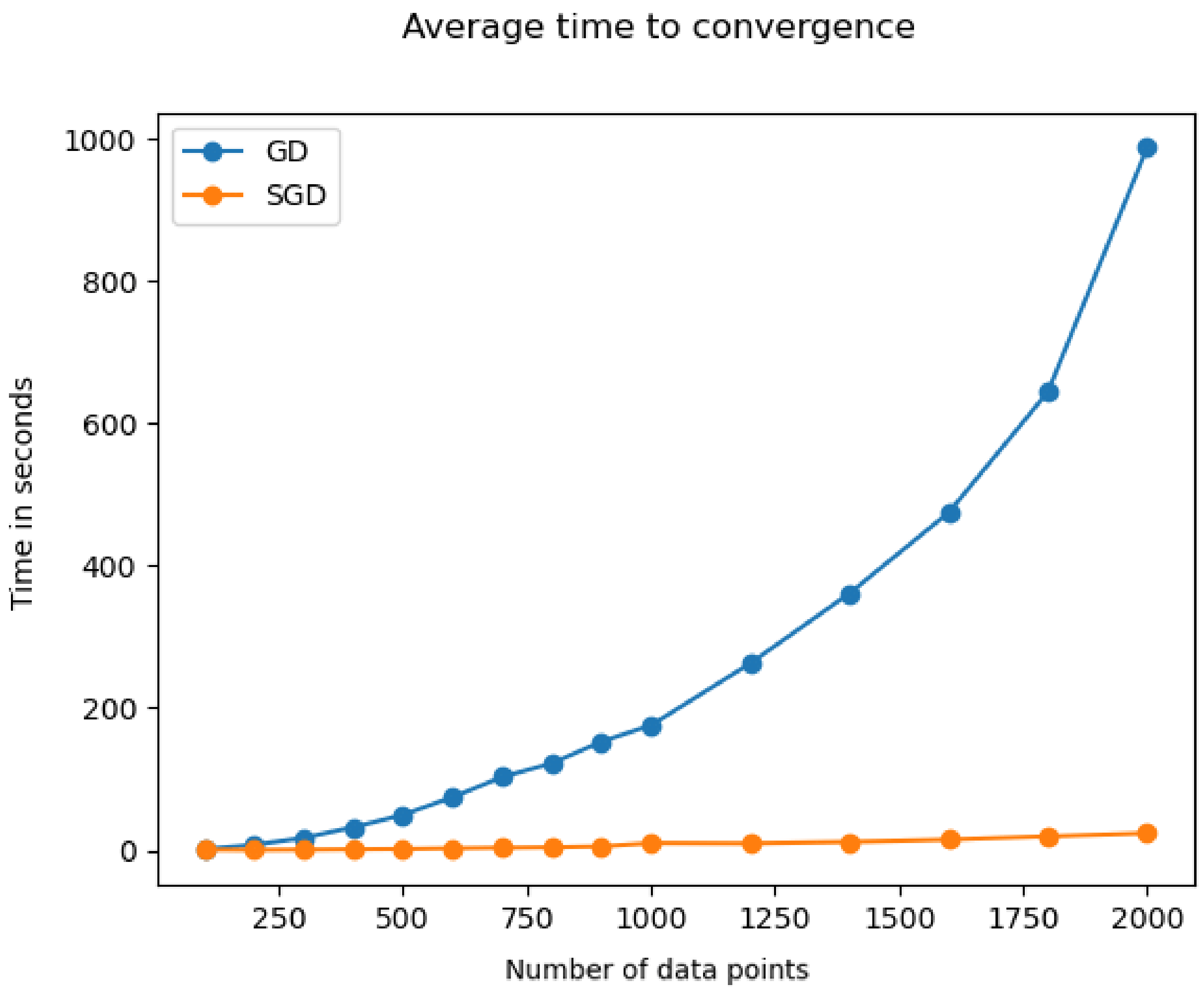


gradient descent

Стохастический градиентный спуск (Stochastic gradient descent)



Стохастический градиентный спуск (Stochastic gradient descent)



Стохастический градиентный спуск (Stochastic gradient descent)

Mini-batch

1. Стартуем из случайной точки

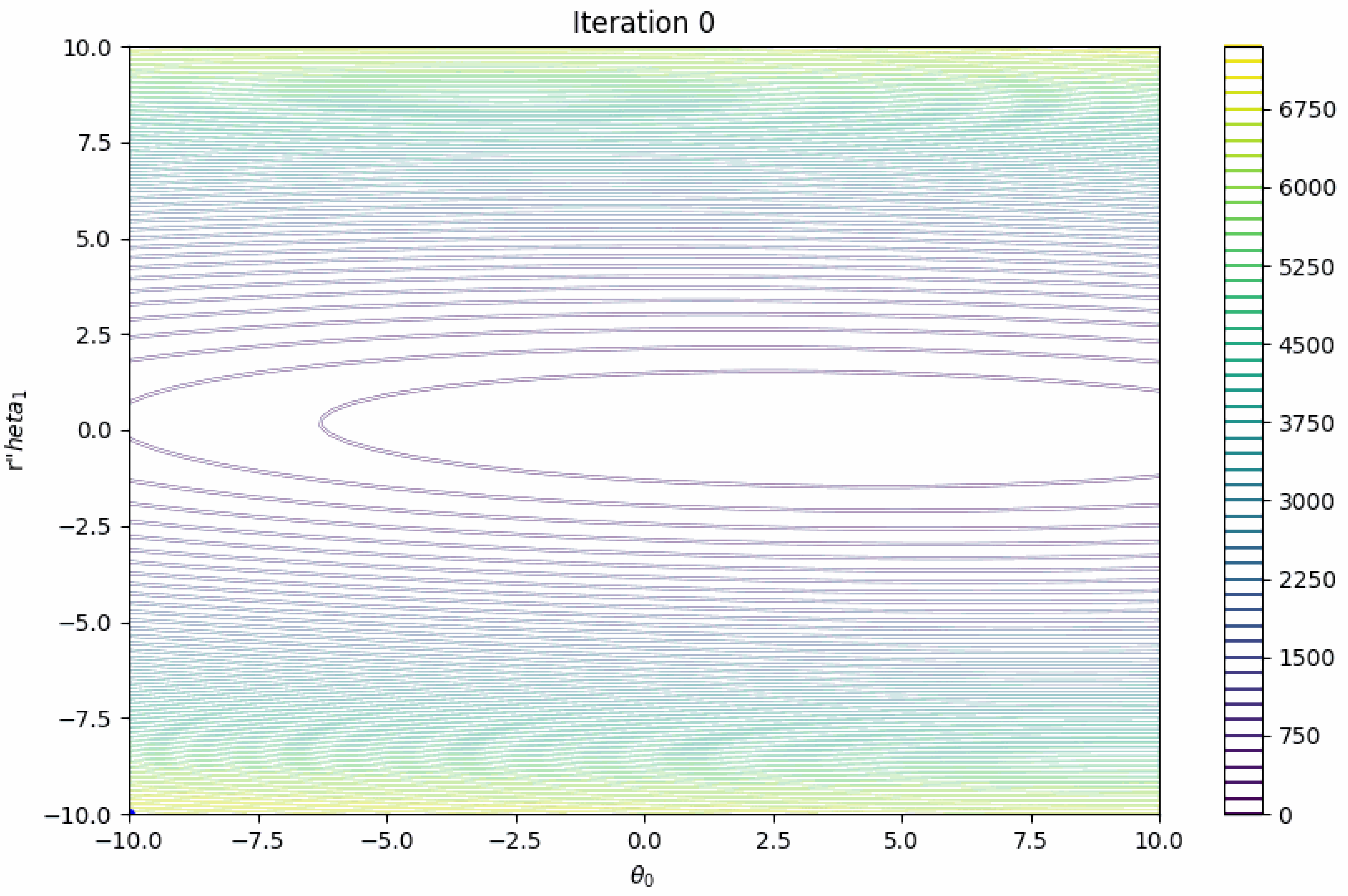
$$\boldsymbol{\theta} \; - \text{Инициализация весов}$$

2. Повторяем и сдвигаемся по антиградиенту, каждый раз выбираем m случайных объектов $i_1, \dots, i_m$

$$\boldsymbol{\theta}^{t} = \boldsymbol{\theta}^{t-1} - \boldsymbol{\alpha} \frac{\mathbf{1}}{\ell} \sum_{j=1}^{\ell} \nabla \boldsymbol{L}\left(\boldsymbol{y}_{i_j}, \boldsymbol{a}(\boldsymbol{x}_{i_j})\right)$$

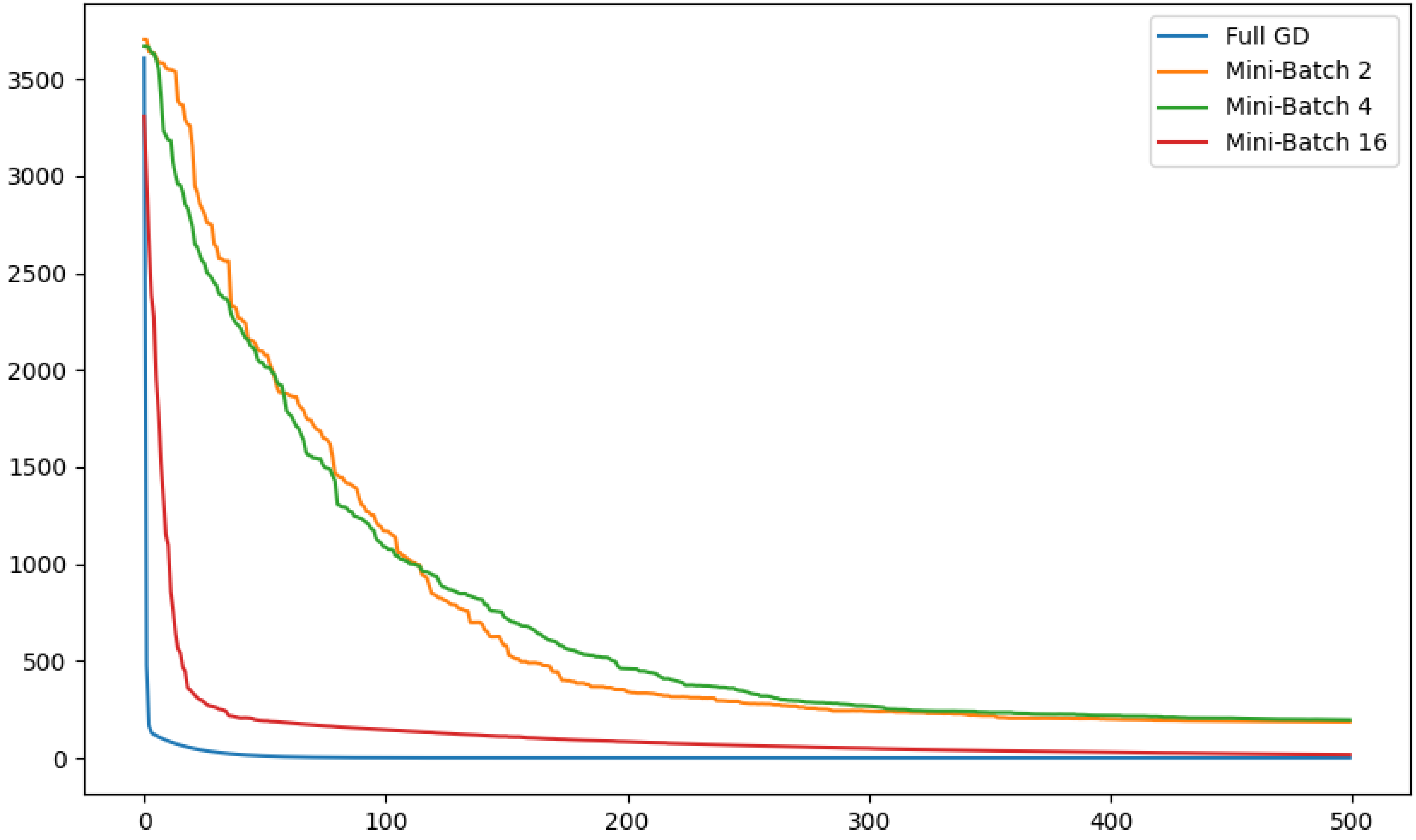
3. Повторяем пока не окажемся в точке минимума

Mini-batch



Iteration 0
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
r"heta$_1$
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
θ_0
6750
6000
5250
4500
3750
3000
2250
1500
750
0

Mini-batch



Full GD
Mini-Batch 2
Mini-Batch 4
Mini-Batch 16
3500
3000
2500
2000
1500
1000
500
0
0
100
200
300
400
500

Масштабирование данных

Масштабирование данных (Data Scaling)

Площадь квартиры (x_1)	Этаж квартиры (x_2)	Количество комнат (x_4)	Расстояние от метро (x_4)	Стоимость квартиры ($\mathbb{Y}$)
460	2	6	1800	195
230	7	4	4500	130
315	1	3	3200	140
178	3	4	2740	80
…	…	…	…	….

x_1: Площадь квартиры (1 – 2000)

x_2: Этаж квартиры (1 – 10)

x_3: Количество комнат (1 – 10)

x_4: Расстояние от метро (1 – 10000)

Не зависящие

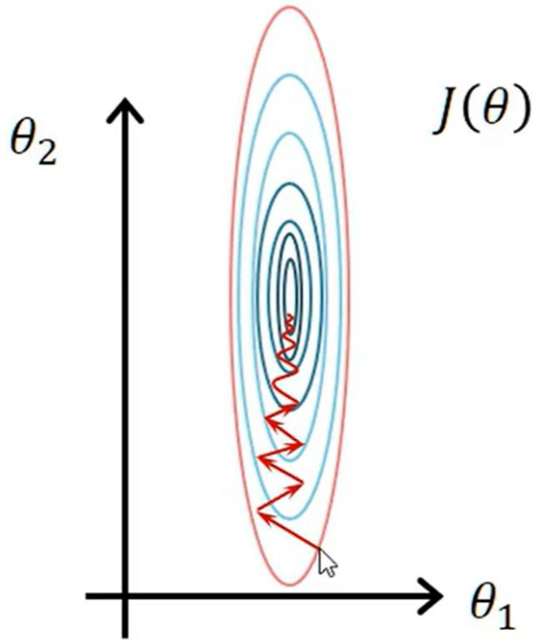
зависящие

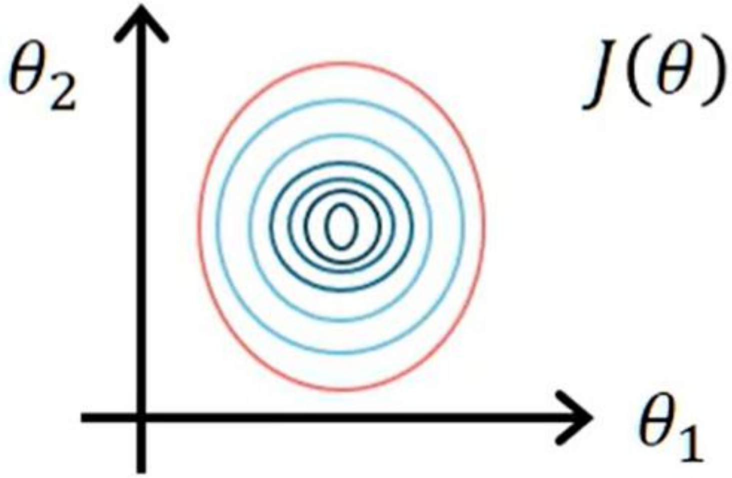
- Случайный лес
- Градиентный бустинг

- Метод ближайших соседей
- Линейная регрессия
- Логистическая регрессия
- Метод опорных векторов (SVM)
- Нейронные сети
- Некоторые алгоритмы кластеризации (K-means)
- Анализ главных компонент (Principal Component Analysis, PCA)

Масштабирование данных (Data Scaling)

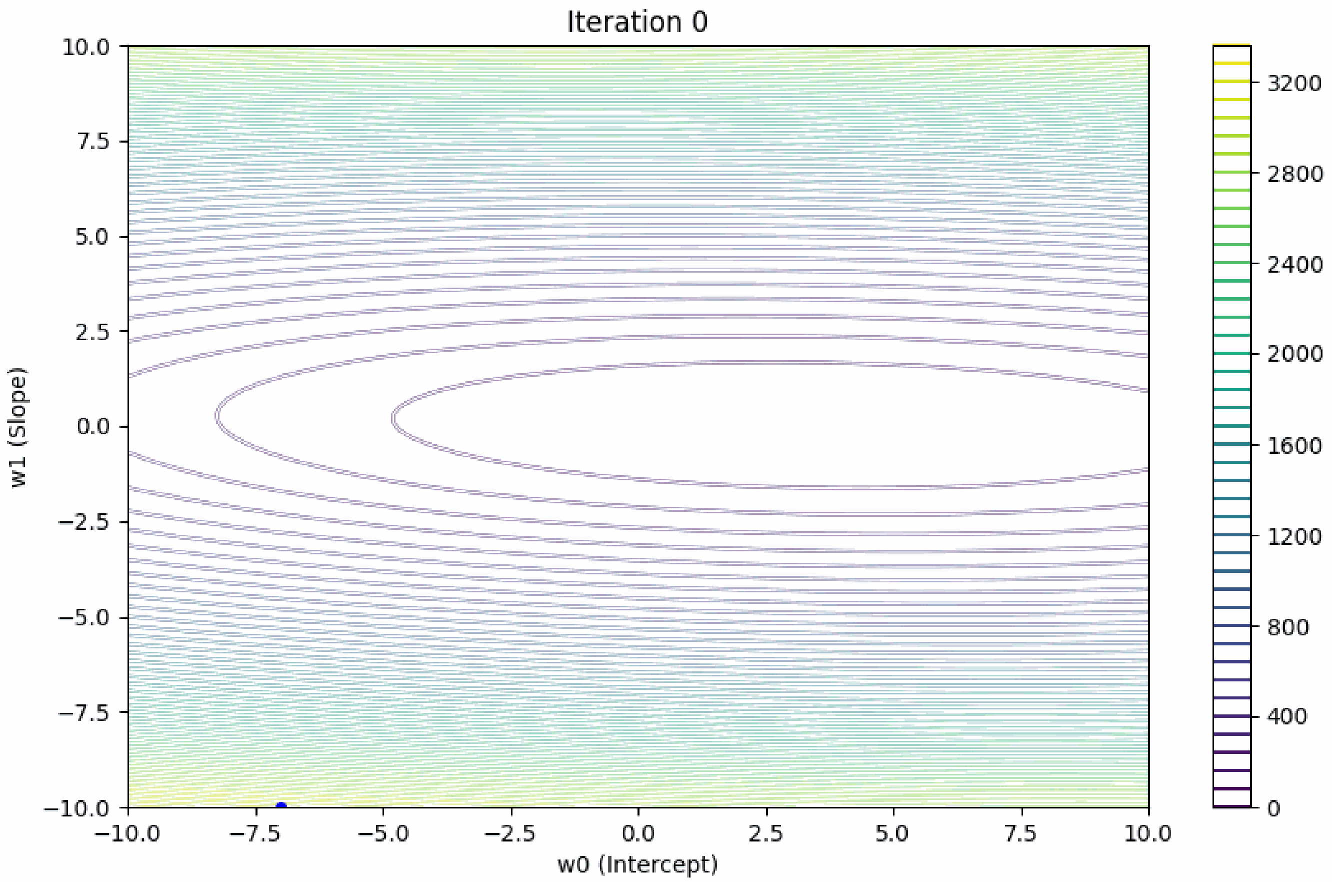
Цель: увеличение скорости сходимости в градиентном спуске.





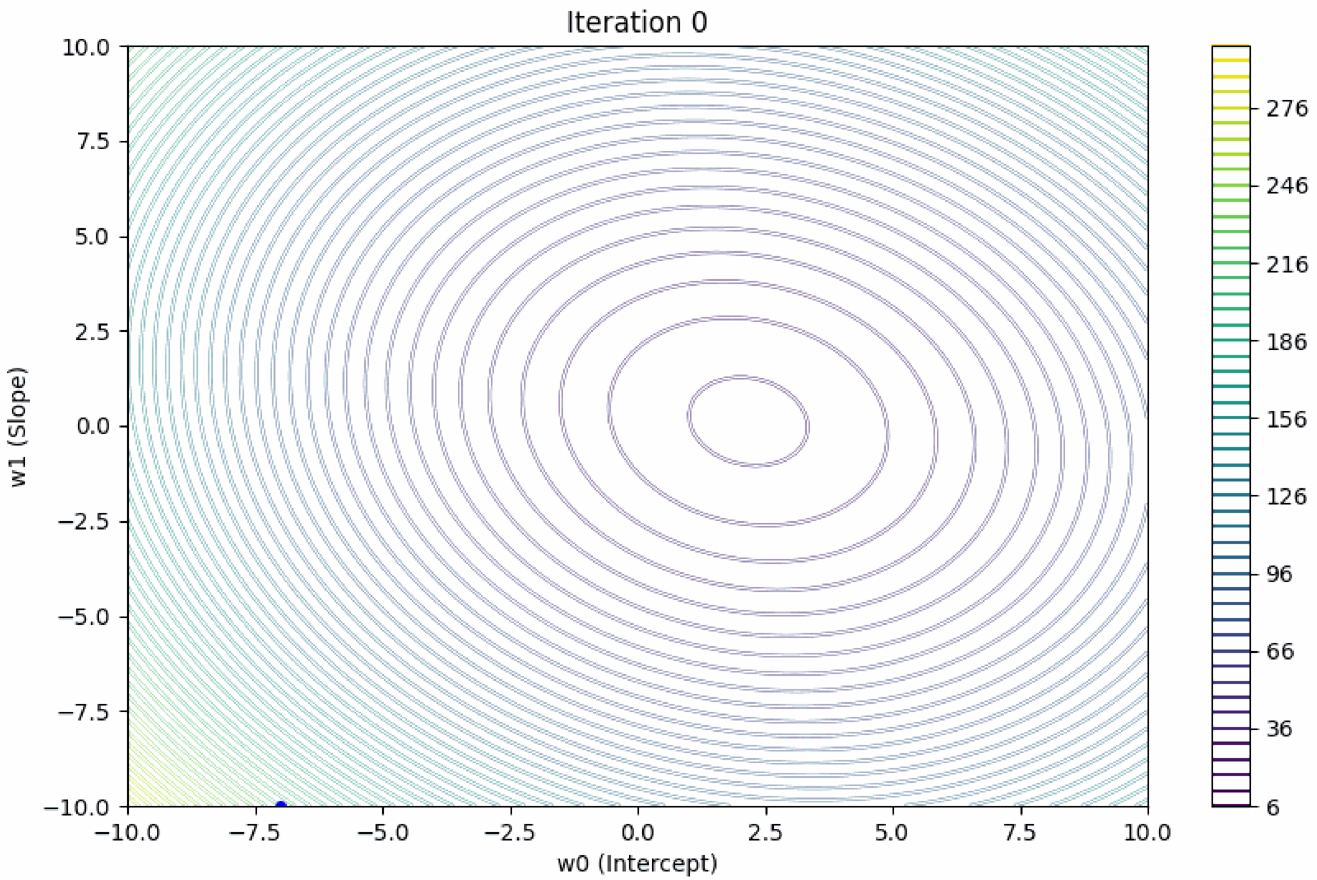
Масштабирование данных (Data Scaling)

Цель: увеличение скорости сходимости в градиентном спуске.



Масштабирование признаков

Процесс изменения данных происходит таким образом, чтобы они имели одинаковый масштаб



Масштабирование признаков на основе Z-оценки

Вычтем из каждого значения признака среднее и поделим на стандартное отклонение

$$x_i^{d_j} = \frac{x_i^{d_j} - \mu_j}{\sigma_j}$$

$x_i^{d_j}$ - признак

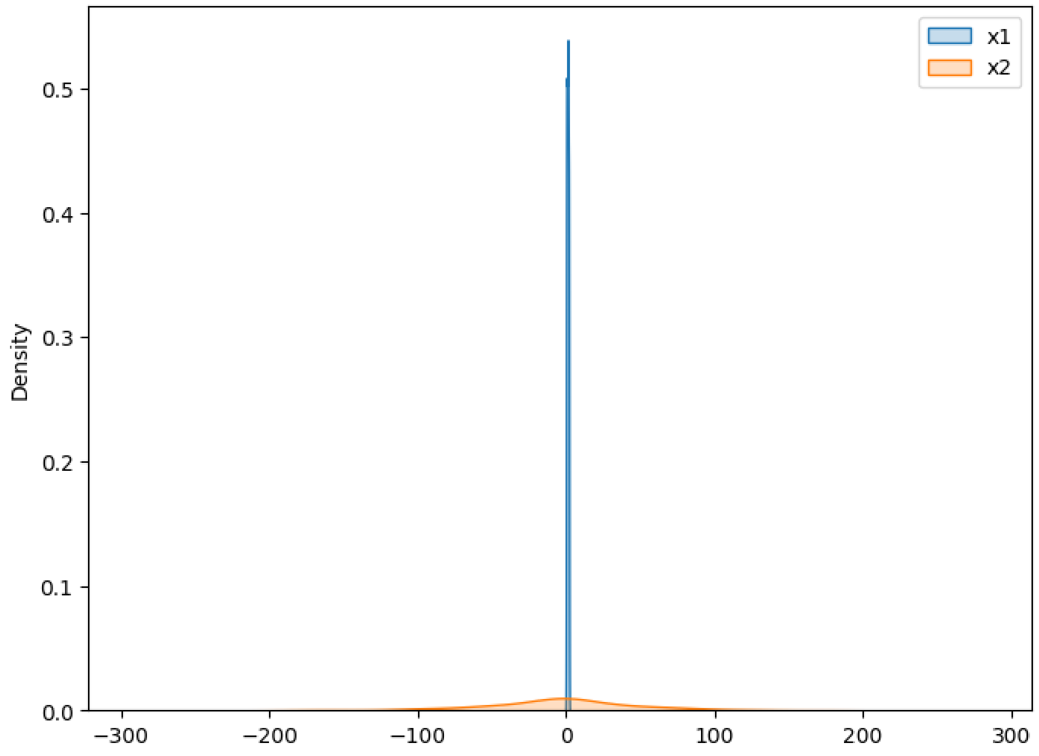
μ_j- mean(x) – среднее значение

$\sigma_{j-}\ std(x) - \text{стандартное отклонение}$

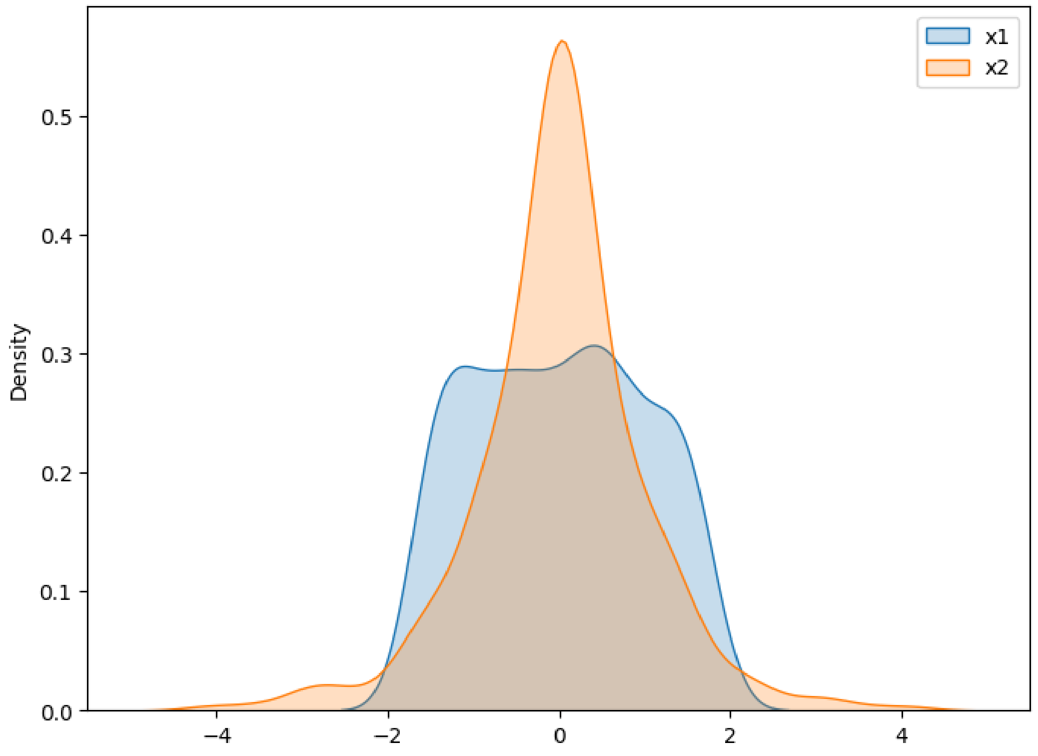
- **Среднее значение масштабированных данных становится равным 0.**
- **Стандартное отклонение масштабированных данных становится равным 1.**
- **Выбросы сохраняются**

Масштабирование признаков на основе Z-оценки

До масштабирования



После масштабирования



Mean Normalization

$$x_i^{d_j} = \frac{x_i^{d_j} - \mu_j}{\max\left(x_i^{d_j}\right) - \min(x_i^{d_j})}$$

$x_i^{d_j}$ - признак

μ_j- mean(x) – среднее значение

- **Среднее значение масштабированных данных становится равным 0.**
- **Максимальные и минимальные значения в диапазоне [-1;1]**
- **Выбросы сохраняются**

MinMax-масштабирование

$$x_i^{d_j} = \frac{x_i^{d_j} - \min(x_i^{d_j})}{\max\left(x_i^{d_j}\right) - \min(x_i^{d_j})}$$

$x_i^{d_j}$ - признак

- **Среднее значение и среднеквадратичное отклонение может варьироваться.**
- **Максимальные и минимальные значения в диапазоне [0;1]**
- **Выбросы сохраняются**

Масштабирование по максимальному значению

$$x_i^{d_j} = \frac{x_i^{d_j}}{\max\left(\left|x_i^{d_j}\right|\right)}$$

$x_i^{d_j}$ - признак

- **Среднее значение не центрируется.**
- **Максимальные и минимальные значения в диапазоне [-1;1]**
- **Среднеквадратичное отклонение не масштабируется.**