Модуль 4
Оптимизация

ЗАДАЧА ОПТИМИЗАЦИИ

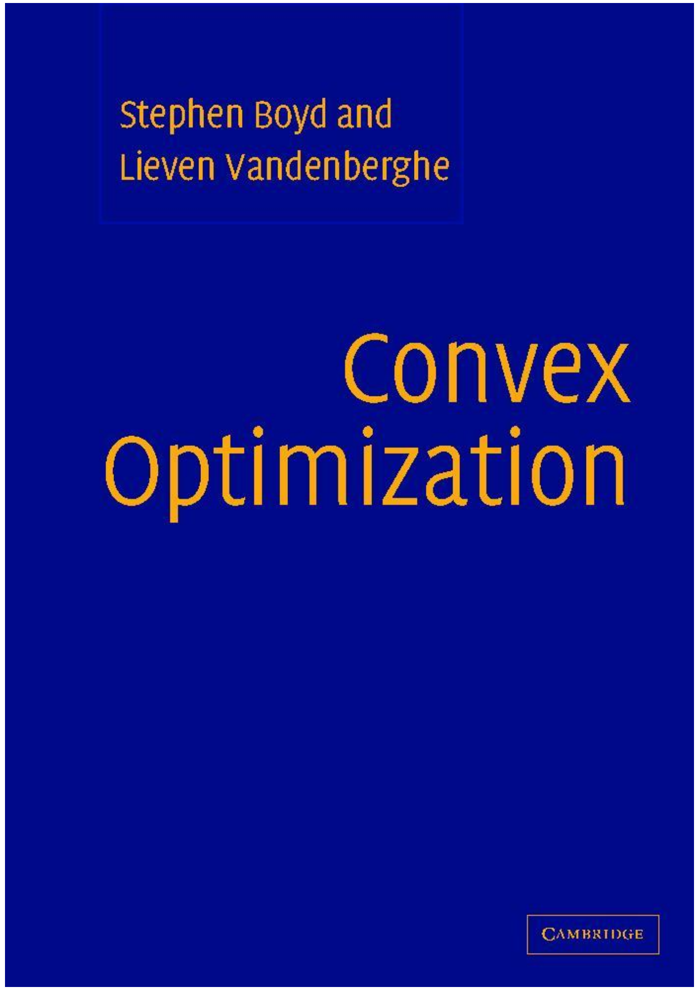
МЕТОД ЛАГРАНЖА

ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

План вебинара

- Общая информация о задачах оптимизации: понятия, постановка задачи
- Метод Лагранжа
- Линейное программирование
- Градиентный спуск

Литература для дополнительного изучения



Оптимизация – сердце data science

Оптимизация лежит в основе каждого крупного бизнеса, социального, экономического и - внезапно!-личного решения, которое принимается отдельным человеком, коллективным представлением людей или интеллектуальными машинами и программными агентами.

Мы решаем задачи оптимизации ежедневно:

- Планируем распорядок своих дел, чтобы затратить минимум времени
- Выбор оптимального маршрута домой, чтобы добраться как можно быстрее (минимизируем время!)
- Делаем список покупок, чтобы потратить минимум денег, купив все необходимое

С какими задачами оптимизации мы сталкиваемся при работе с данными

- Эффективное использование данных
- Минимизирование вычислительной нагрузки при обработке большого набора данных
- Поиск минимумов и лучшего решения из сложного многомерного пространства.

Общий вид задачи для нелинейной оптимизации

Целевая функция

$$\min f_0(x)$$

$$\text{при } f_j(x) \,\&\, 0, \quad j = 1, \ldots, m,$$

$$x \in S,$$

Вектор функциональных ограничений

Типы задач минимизации

- Условные задачи
- Безусловные задачи
- Гладкие задачи
- Негладкие задачи
- Задачи с линейными ограничениями

Общая итеративная схема

Общая итеративная схема

Вводные данные: начальная точка x_0 и требуемая точность $\varepsilon > 0$.

Настройка. Полагаем $k = 0$ и $I_{-1} = \varnothing$. Здесь k — это счетчик итераций, а I_k — это накапливаемая *информационная модель* решаемой задачи.

Основной цикл

1. Задаем вопрос оракулу $\mathcal{O}$ в точке x_k.
2. Пересчитываем информационную модель:
$$I_k = I_{k-1} \cup \big(x_k, \mathcal{O}(x_k)\big).$$
3. Применяем правила метода $\mathcal{M}$ для анализа модели I_k и формируем точку x_{k+1}.
4. Проверяем критерий остановки $\mathcal{T}_\varepsilon$. Если ответ положительный, то генерируем ответ $\bar{x}$. В противном случае полагаем $k := k + 1$ и переходим на шаг 1.

Оракулы

- Оракул *нулевого порядка*: возвращает значение функции $f(x)$.
- Оракул *первого порядка*: возвращает значение функции $f(x)$ и ее градиент $f'(x)$.
- Оракул *второго порядка*: возвращает $f(x)$, $f'(x)$ и матрицу гессиана $f''(x)$.

Важно:
Мы ВСЕГДА решаем задачу на нахождение минимума

$$\max f(x) \Leftrightarrow \min(-f(x))$$

Поиск экстремумов

Ищем экстремумы:

- Локальные и глобальные.
- Условные и безусловные.

Поиск экстремумов

- Если $f'(x)>0$, то функция локально возрастает.
- Если $f'(x)<0$, то функция локально убывает.
- Чем больше производная, тем быстрее растет функция.
- Нулевая производная означает отсутствие изменения функции.
- Может существовать единственный глобальный экстремум
- В точке перегиба производная обнуляется, но не меняет знак.

Функция многих переменных

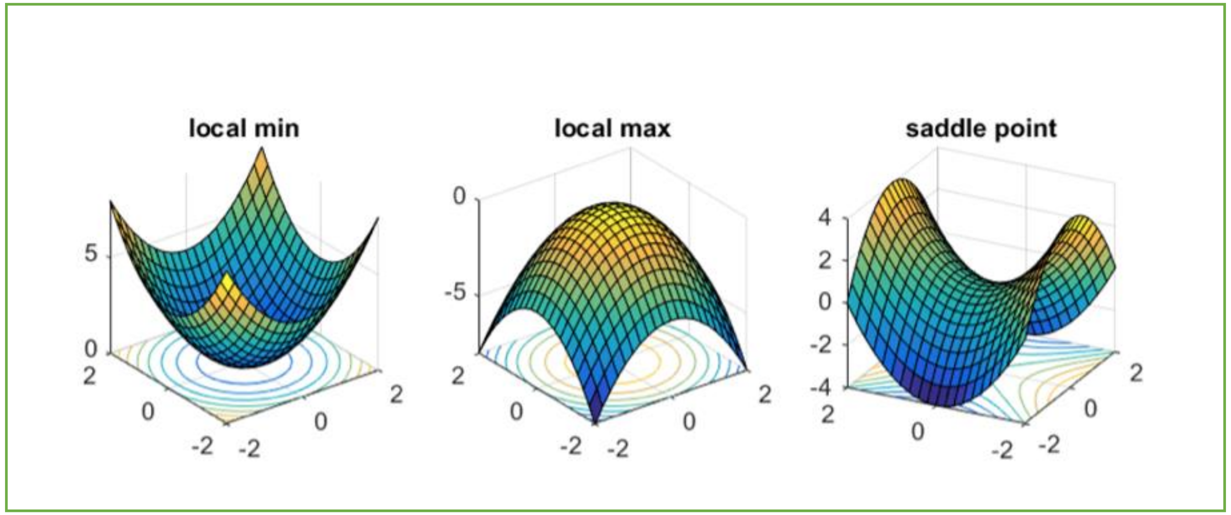
1. Найти $\overrightarrow{grad f}$.

2. Решить $\overrightarrow{grad f} = \vec{0}$.

3. Если матрица Гессе положительно или неотрицательно определена, то имеем строгий или нестрогий минимум.

4. Если матрица Гессе отрицательно (неположительно) определена, то имеем строгий (нестрогий) максимум.

5. В противном случае имеем точку перегиба.



Матрица будет:

- **положительно определённой**, если все её собственные значения положительны,
- **отрицательно определённой**, если все её собственные значения отрицательны,
- **положительно полуопределённой**, если все её собственные значения неотрицательны,
- **отрицательно полуопределённой**, если все её собственные значения неположительны.

Метод Лагранжа. Условие в виде равенства

- Задача оптимизации с ограничением

- Составляется **функция Лагранжа** и вычисляются частные производные построенной функции Лагранжа по ***x*** и дополнительной переменной ***λ***, далее знакомым образом находятся точки, в которых производные равны нулю, и точки экстремума.

- Задача условной оптимизации сводится к задаче безусловной оптимизации.

- Необходимое условие выполнения метода: функции ***f(x)*** и $\boldsymbol{\varphi_i(x)}$ и их производные непрерывны и дифференцируемы во всех исследуемых точках.

$$\min_{\phi_i(x)=0} f(x)$$

$$L(x,\lambda) = f(x) + \sum \lambda_i \phi_i(x)$$

Метод Лагранжа: пример вычисления

$$\min_{x_1+x_2-10=0} -x_1x_2$$

$$\downarrow$$

$$L(x_1,x_2,\lambda_1) = -x_1x_2 + \lambda_1(x_1+x_2-10) \longrightarrow \begin{aligned} \frac{dL(x_1,x_2,\lambda_1)}{dx_1} &= -x_2+\lambda_1 \\ \frac{dL(x_1,x_2,\lambda_1)}{dx_2} &= -x_1+\lambda_1 \\ \frac{dL(x_1,x_2,\lambda_1)}{d\lambda_1} &= x_1+x_2-10 \end{aligned} \longrightarrow \begin{cases} -x_2+\lambda_1 = 0 \\ -x_1+\lambda_1 = 0 \\ x_1+x_2 = 10 \end{cases}$$

$$\begin{array}{c} \begin{array}{ccc} x_1 & x_2 & \lambda_1 \end{array} \\ \left(\begin{array}{ccc|c} 0 & -1 & 1 & 0 \\ -1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 10 \end{array}\right) \end{array} \sim \left(\begin{array}{ccc|c} 1 & 1 & 0 & 10 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 1 & 0 & 10 \\ 0 & 1 & 1 & 10 \\ 0 & -1 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 1 & 0 & 10 \\ 0 & 1 & 1 & 10 \\ 0 & 0 & 2 & 10 \end{array}\right) \longrightarrow \begin{cases} x_1+x_2 = 10 \\ x_2+\lambda_1 = 10 \\ 2\lambda_1 = 10 \end{cases} \Leftrightarrow \begin{cases} x_1 = 5 \\ x_2 = 5 \\ \lambda_1 = 5 \end{cases}$$

Метод Лагранжа. Условие в виде неравенства

- Ограничения в виде неравенства можно свести к ограничениям равенства с помощью дополнительной переменной

$$\phi_i(x) \le 0 \quad \Leftrightarrow \quad \phi_i(x) + \tilde{x}_i^{\,2} = 0$$

Метод Лагранжа. Условие в виде неравенства

$$\max_{\substack{2x_1+2x_2=20 \\ x_1\geq 6}} x_1x_2$$

$$\min_{\substack{x_1+x_2-10=0 \\ -x_1+6+\tilde{x}_1^2=0}} -x_1x_2$$

$$L(x_1,x_2,\tilde{x}_1,\lambda_1,\lambda_2) = -x_1x_2+\lambda_1(x_1+x_2-10)+\lambda_2(-x_1+6+\tilde{x}_1^2)$$

$$\frac{dL}{d\lambda_1} = x_1+x_2-10$$

$$\frac{dL}{d\lambda_2} = -x_1+6+\tilde{x}_1^2$$

$$\frac{dL}{d\tilde{x}_1} = 2\lambda_2\tilde{x}_1$$

$$\frac{dL}{dx_1} = -x_2+\lambda_1-\lambda_2$$

$$\frac{dL}{dx_2} = -x_1+\lambda_1$$

$$\begin{cases} -x_2+\lambda_1-\lambda_2=0 \\ -x_1+\lambda_1=0 \\ 2\lambda_2\tilde{x}_1=0 \\ x_1+x_2-10=0 \\ -x_1+6+\tilde{x}_1^2=0 \end{cases}$$

Линейное программирование

Задача линейного программирования — это задача оптимизации, в которой целевая функция и функции–ограничения линейны, а все переменные неотрицательны

Целочисленным линейным программированием (ЦЛП) называется вариация задачи линейного программирования, когда все переменные — целые числа.

$$\min_{\substack{Ax \leq b \\ x \geq 0}} c^T x$$

Кондитерская фабрика производит продукцию двух видов: конфеты и шоколад. Для производства продукции каждого вида требуются ресурсы двух типов: сахар и какао-бобы. Для производства одной тонны продукции каждого вида требуется по одной тонне сахара. Для производства одной тонны шоколада требуется 5 тонн какао, а для производства одной тонны конфет – 2 тонны какао. Суточные запасы ресурсов равны 4 и 10 тонн соответственно. Прибыль от реализации одной тонны шоколада и конфет составляет 5 и 3 тысячи рублей соответственно. Написать математическую модель для нахождения оптимального (т. е. максимизирующего прибыль) суточного плана производства.

Основные данные задачи можно представить в виде таблицы:

Исходные ресурсы	Расход ресурсов на 1 тонну готовой продукции		Запас ресурса
	Шоколад	Конфеты	
Сахар	1	1	4
Какао	5	2	10
Прибыль	5	3	

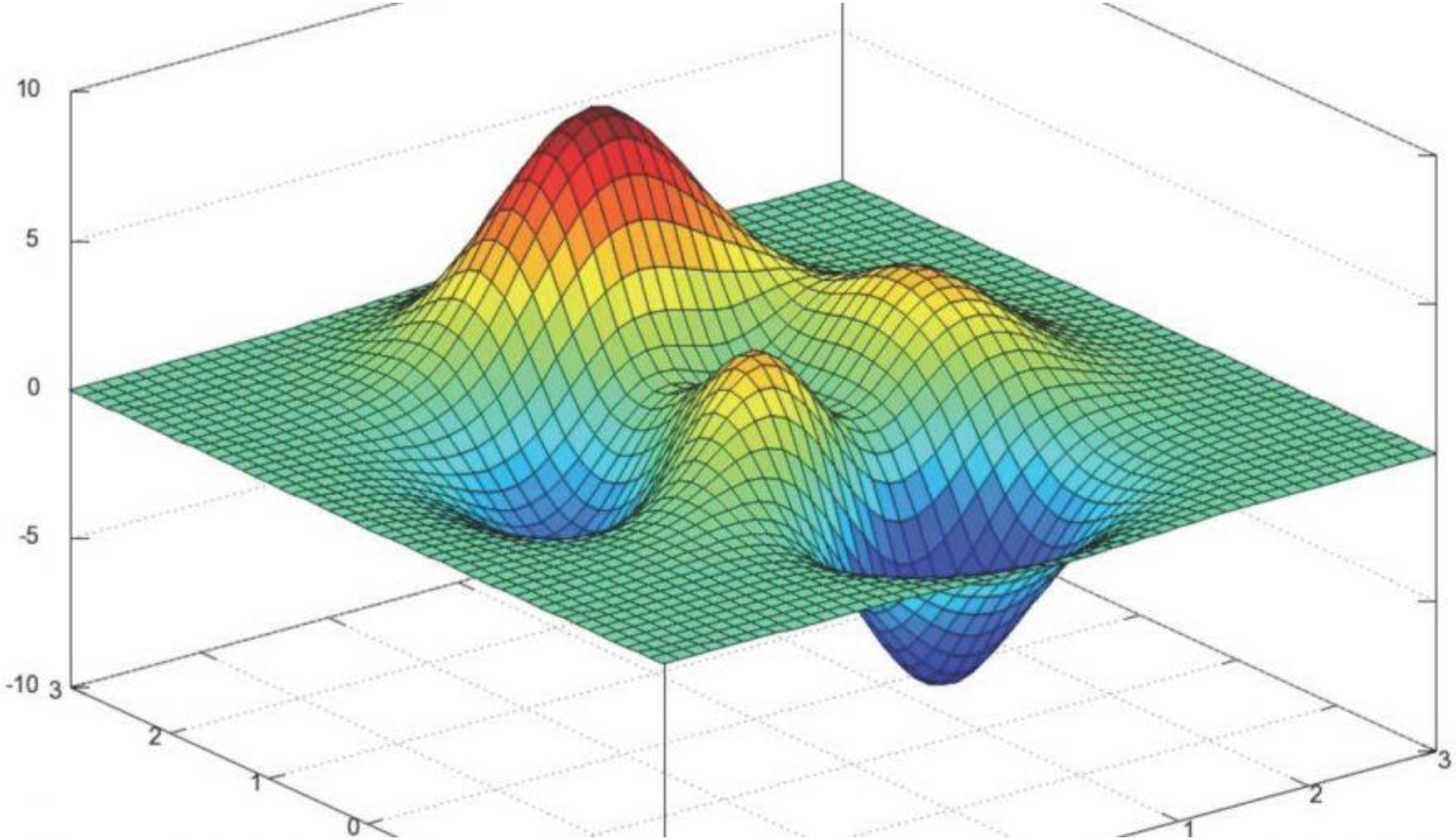
Вы сотрудник коммерческой фирмы и отвечаете за рекламу. Затраты на рекламу в месяц не должны превышать 10 000 денежных единиц (д.е). Минута радиорекламы стоит 5 д.е., а телерекламы 90 д.е. Фирма намерена использовать радиорекламу в три раза чаще чем телерекламу. Практика показывает, что 1 минута телерекламы обеспечивает объём продаж в 30 раз больший чем 1 минута радиорекламы.

30x1+x2 –увеличение продаж от рекламы;
90x1+5x2 <=10 000 – ограничение средств;
x2=3x1 – соотношение времён радио и теле рекламы.

Вы сотрудник коммерческой фирмы, которая оказывает транспортные услуги. Есть поставщики товара со складами в разных трёх городах, причём объёмы однородной продукции на этих складах соответственно равны a1, a2, a3. Есть и потребители в других трёх городах которым нужно привести товар от поставщиков в объёмах b1, b2, b3 соответственно. Известны также стоимости доставки c1÷c9 товаров от поставщиков к потребителям, согласно таблице.

	b1=74	b2=40	b3=36
a1= 56	c1=7	c2=3	c3=6
a2= 50	c3=4	c4=8	c5=2
a3= 44	c6=1	c7=5	c8=9

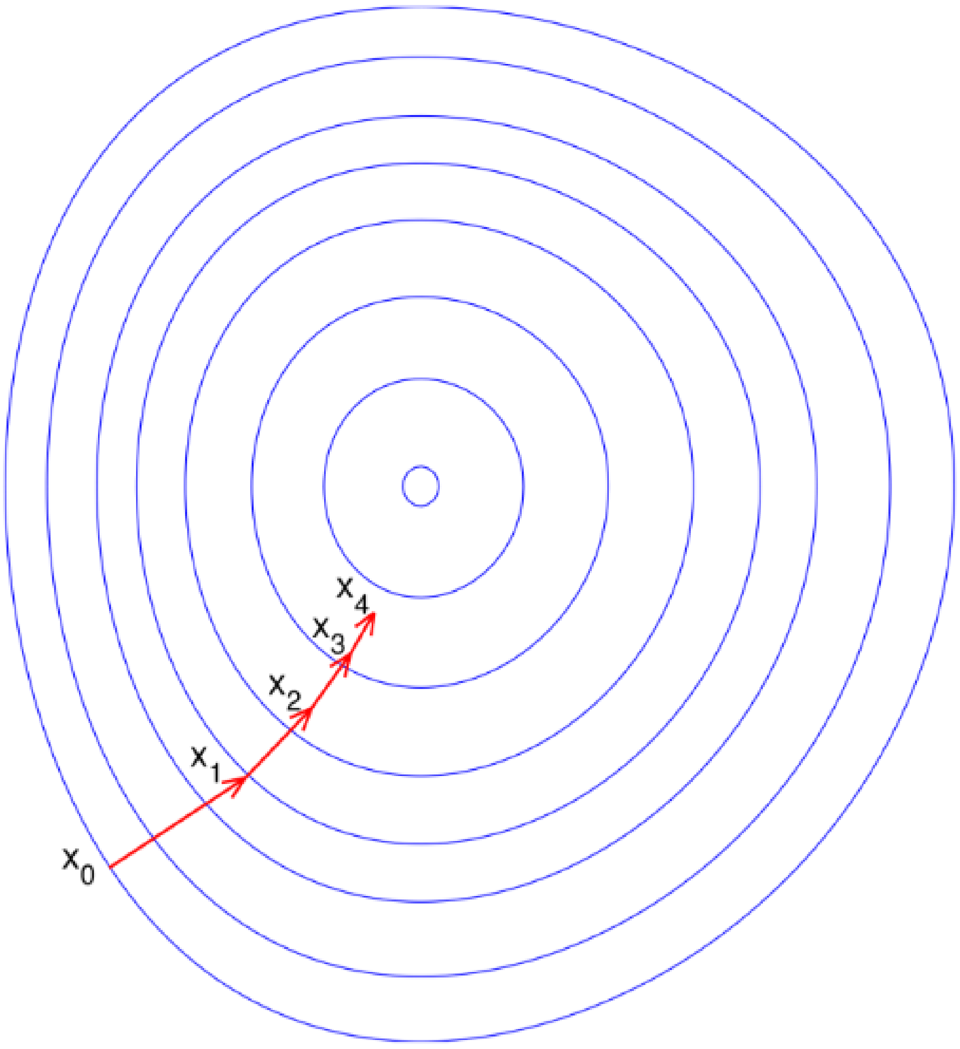
Градиентный спуск



- Градиент – направление наискорейшего роста функции
- Антиградиент – направление наискорейшего убывания функции

Градиентный спуск

- Допустим, мы выбрали начальное приближение $w^0 = (w_0^0, w_1^0)$
- Как его улучшить?
- Шагнуть в сторону наискорейшего убывания
- То есть в сторону антиградиента!



Градиентный спуск

- Повторять до сходимости:

$$w^t = w^{t-1} - \eta \, \nabla Q(w^{t-1})$$

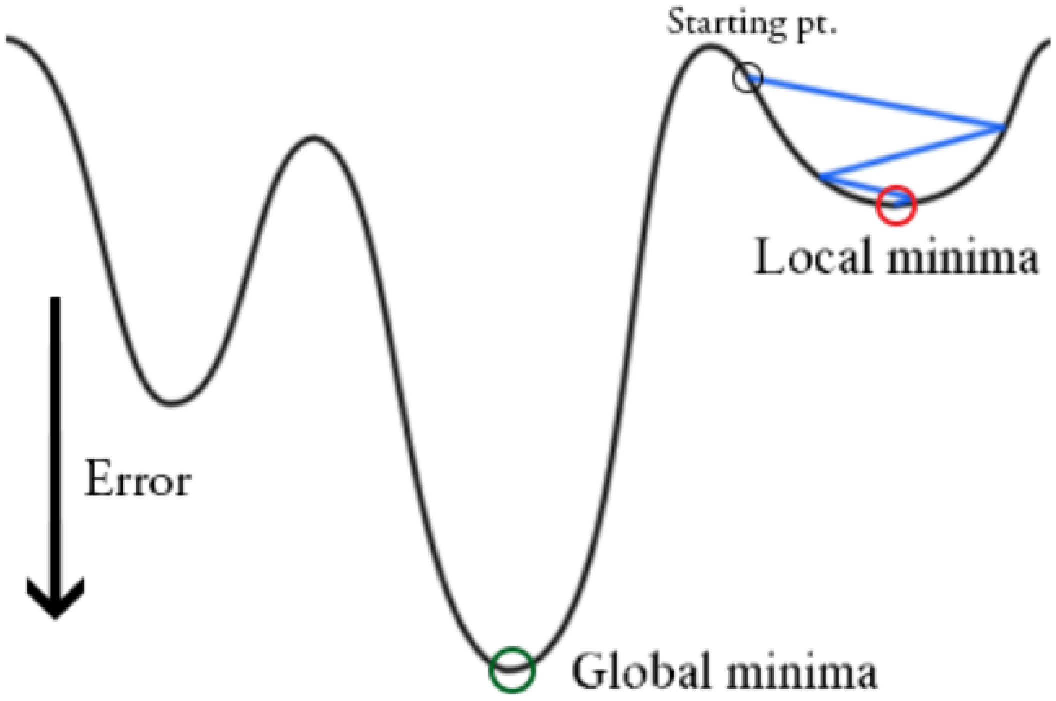
Новая точка

Размер шага

Градиент в предыдущей точке

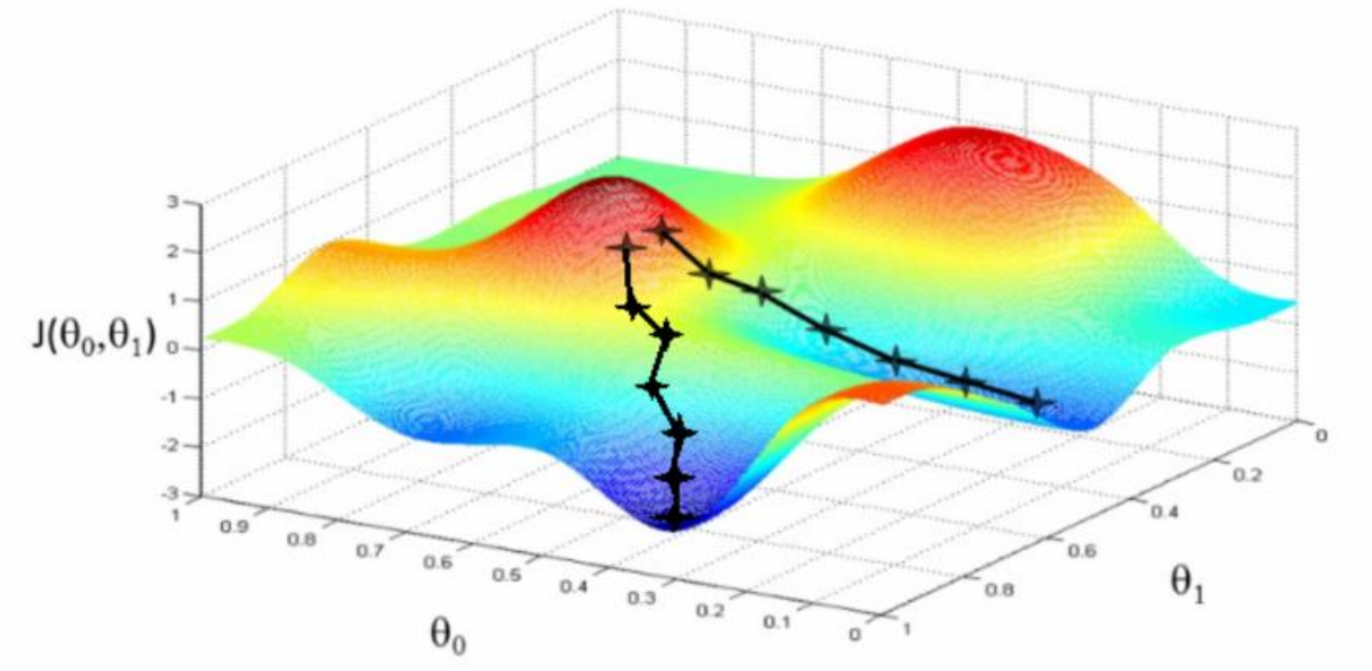
Локальные минимумы

• Градиентный спуск находит только локальные минимумы



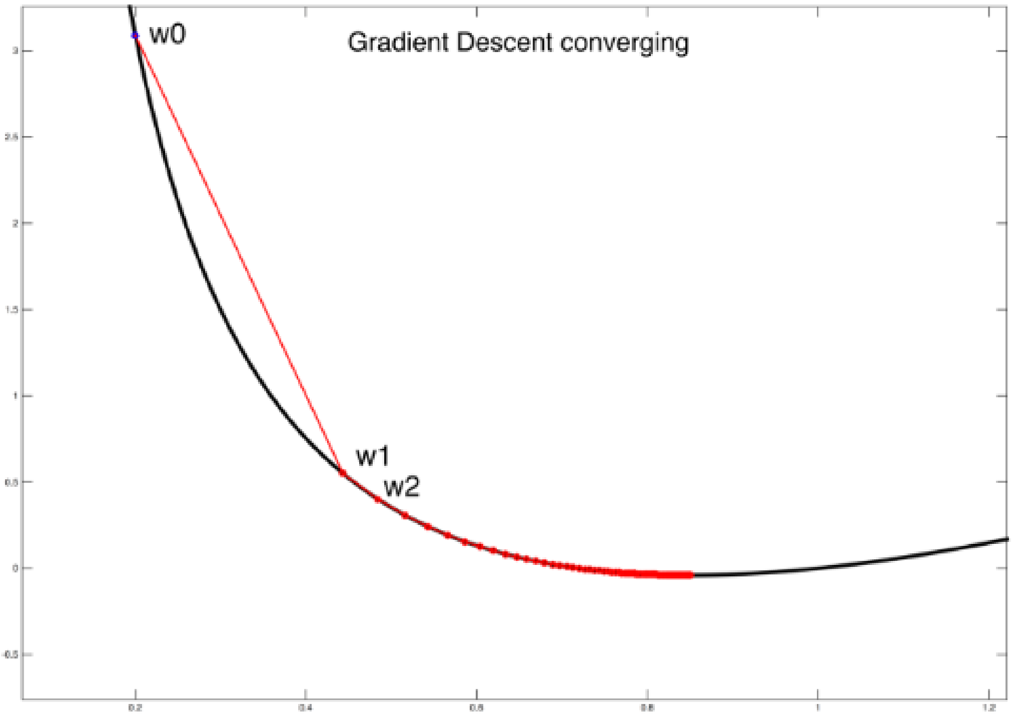
Локальные минимумы

- Результат зависит от начального приближения
- Мультистарт

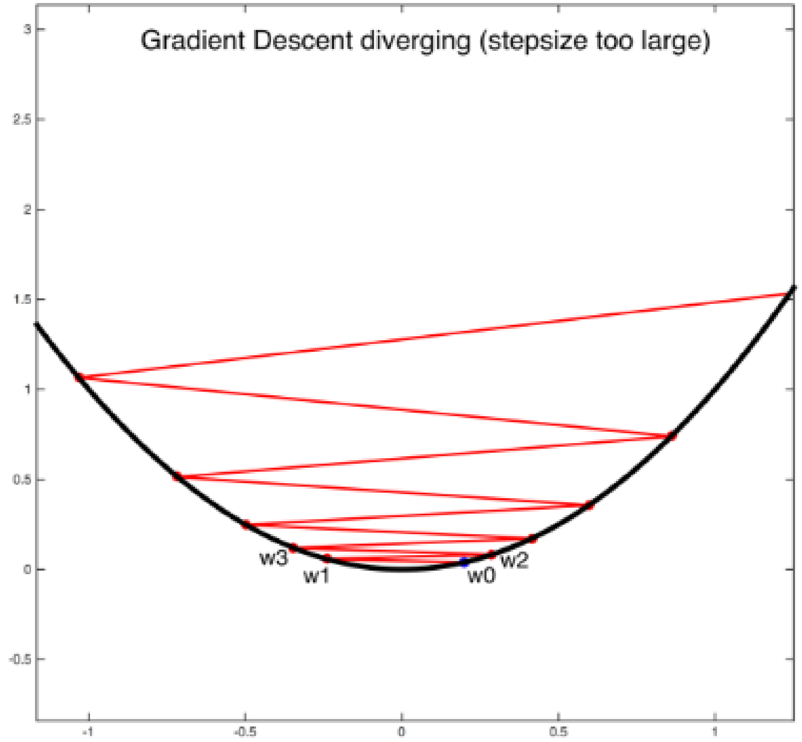


Размер шага

- Выбор размера шага η — искусство



Маленький шаг



Большой шаг