

Производная и градиентный спуск



Функции. Производная. Экстремумы функции. Выпуклость функции.
Правила дифференцирования. Правила дифференцирования сложной
функции. Chain-rule. Функция нескольких аргументов. Градиент.
Градиентный спуск как метод оптимизации.

Юстина Иванова

Специалист по Анализу Данных



Юстина Иванова
студент-аспирант
University of Bolzano

Инженер-программист МГТУ им. Баумана

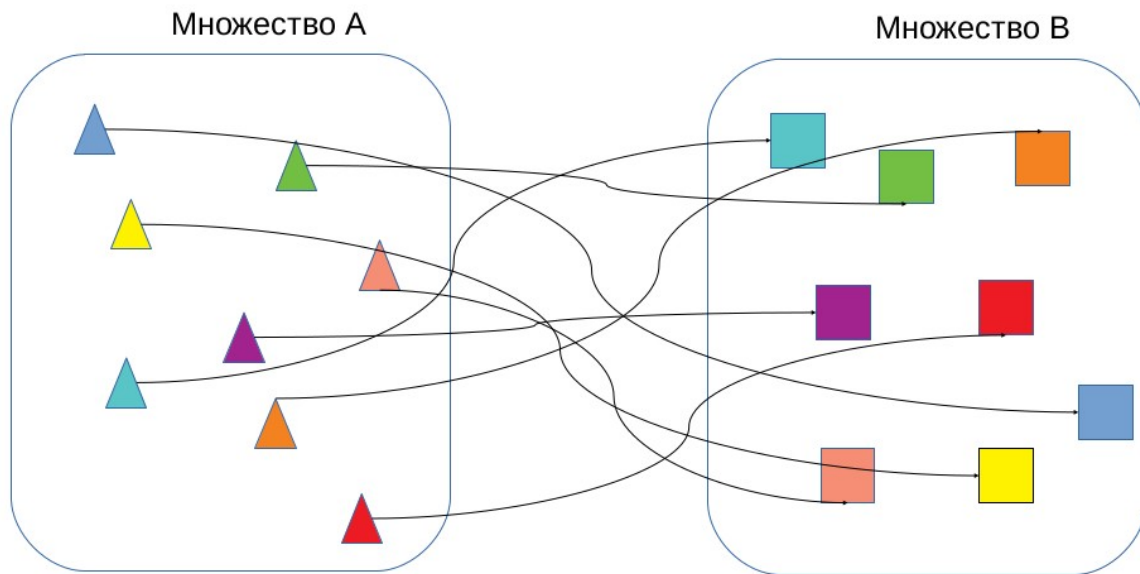
Master of Science in Artificial Intelligence
University of Southampton

Специалист по компьютерному зрению
в компании Dataplex.

Специалист по анализу данных
в компании ОЦРВ.

Функции

Функция - это некоторое соответствие $x \rightarrow f(x)$, причём для каждого x определено единственное значение $f(x)$.

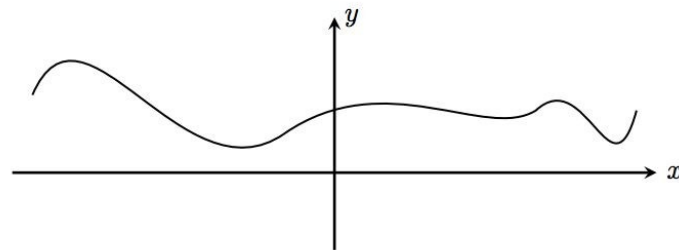


Функции и их свойства

$D(f)$ - область определения функции

$E(f)$ - область значений функции

Будем работать только с функциями, у которых $D(f)$ и $E(f)$ - подмножество действительных чисел \mathbb{R} .



Функции и их свойства

Каковы область определения и область значений следующих функций?

1)
$$f(x) = \frac{1}{x-1}$$

2)
$$f(x) = 2^x$$

Функции и их свойства

Каковы область определения и область значений следующих функций?

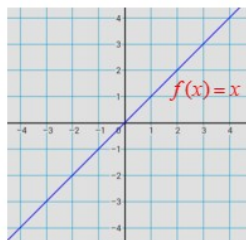
1)
$$f(x) = \frac{1}{x-1}$$

2)
$$f(x) = 2^x$$

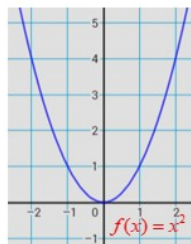
1) $D(f) = \mathbf{R \setminus \{1\}}, E(f) = \mathbf{R \setminus \{0\}}$

2) $D(f) = \mathbf{R}, E(f) = \mathbf{(0, +inf)}$

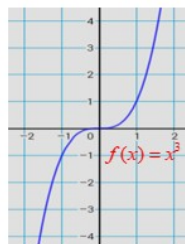
Функции



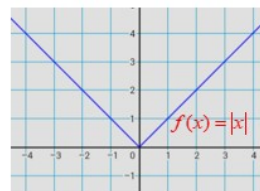
Linear



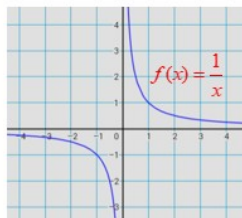
Quadratic



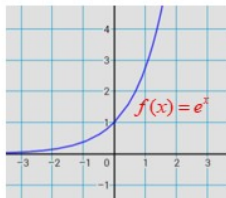
Cubic



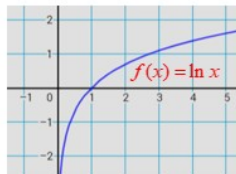
Absolute



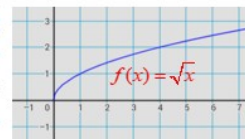
Reciprocal



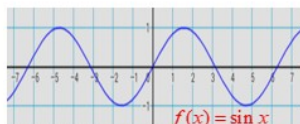
Exponential



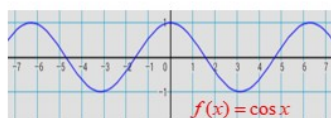
Logarithmic



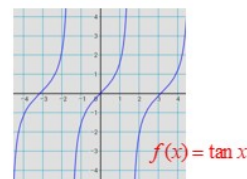
Square Root



Sine



Cosine



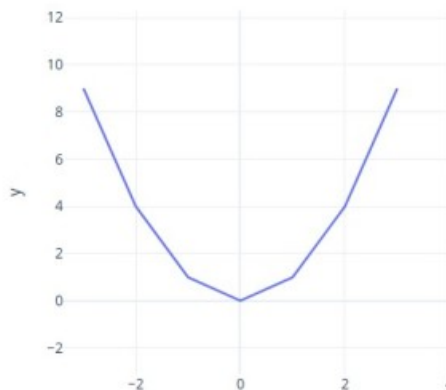
Tangent

Способы представления функций

Табличный вид

x	y
-2	4
-1	1
0	0
1	1
2	4

Графический вид



В виде формулы

$$y = x * x$$

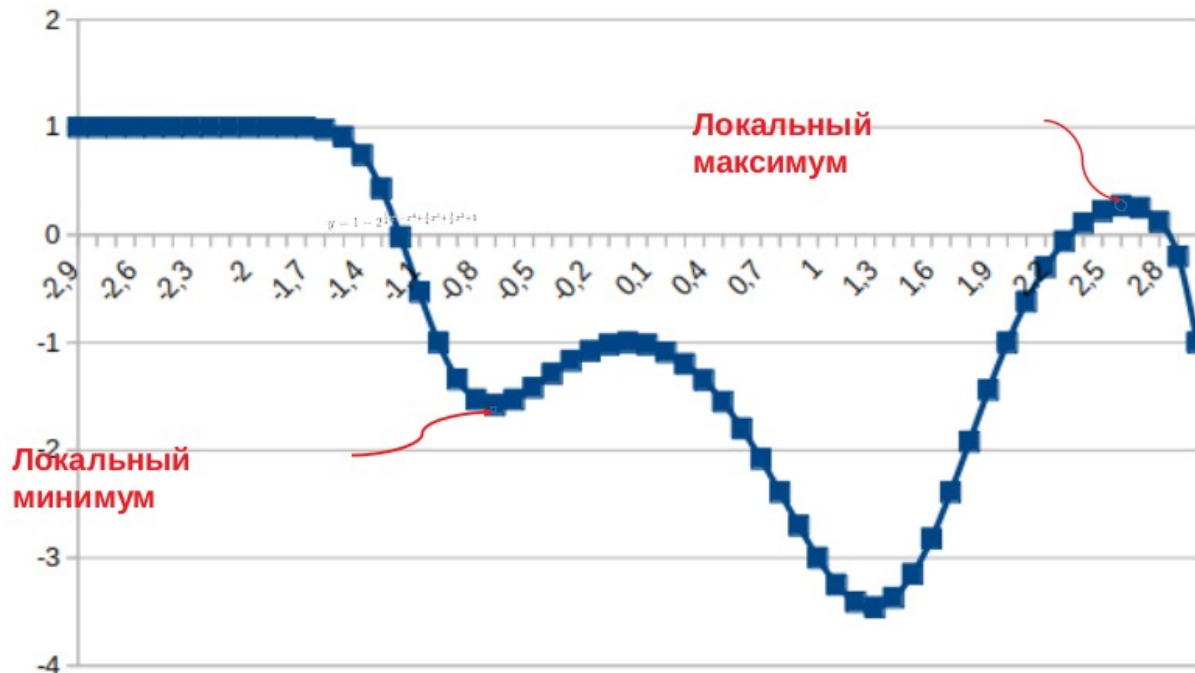
Программно

```
def func(x):  
    y = x*x  
    return y
```


Графический способ представления функции

$$y = 1 - 2\frac{1}{4}x^5 - x^4 + \frac{1}{4}x^3 + \frac{3}{2}x^2 + 1$$

Допустим, есть
некая функция. Мы
можем
проанализировать
поведение функции
графически и
математически.



Программный способ представления функции

$$y = 1 - 2\frac{1}{4}x^5 - x^4 + \frac{1}{4}x^3 + \frac{3}{2}x^2 + 1$$

```
def function(x):  
    stepen = 1/4 * x ** 5 - x ** 4 + 1/4 * x ** 3 + 3/2 * x ** 2 + 1  
    y = 1 - x**stepen  
    return y
```

Задание

Для каждой функции найти пару.

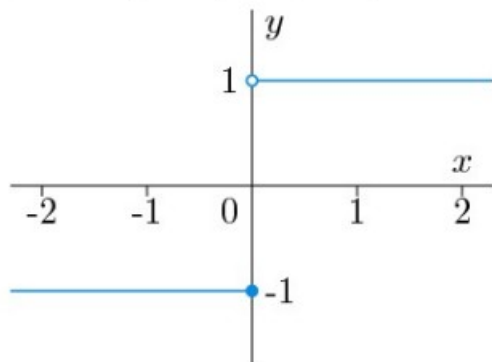
1

$$f(x) = x^2$$

2

x	0	1	4	9
$f(x)$	0	1	2	3

3



4

```
def func(x):  
    return abs(x)
```

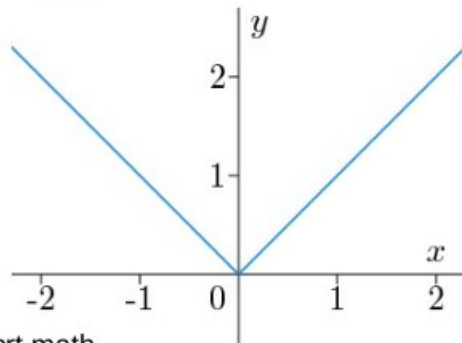
5

$$f(x) = \begin{cases} -1 & \text{for } x \leq 0 \\ 1 & \text{for } x > 0 \end{cases}$$

6

x	0	1	2	3
$f(x)$	0	1	4	9

7



import math

8

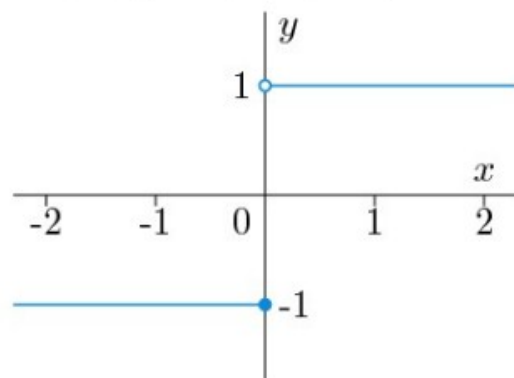
```
def func(x):  
    return math.sqrt(x)
```

Ответ

Для каждой функции найти пару.

$$f(x) = x^2$$

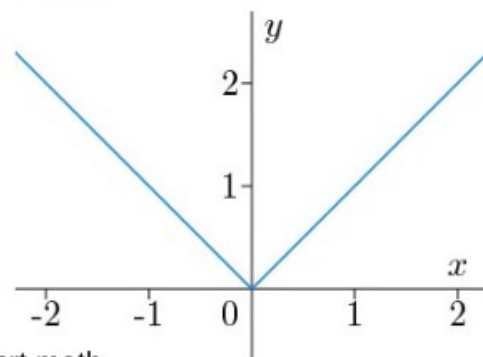
x	0	1	4	9
$f(x)$	0	1	2	3



```
def func(x):  
    return abs(x)
```

$$f(x) = \begin{cases} -1 & \text{for } x \leq 0 \\ 1 & \text{for } x > 0 \end{cases}$$

x	0	1	2	3
$f(x)$	0	1	4	9



```
import math
```

```
def func(x):  
    return math.sqrt(x)
```

Полиномиальная функция

Целая рациональная функция (также полиномиальная функция) — числовая функция одного действительного переменного вида:

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_2 x^2 + a_1 x + a_0 = \sum_{i=0}^n a_i x^i.$$

На практике любую функцию можно представить в виде суммы нескольких полиномов.

Полиномиальная функция:

состоит из:

- Чисел
- Умножения
- Сложения
- Переменной x

Степень полиномиальной функции

Одним из основных показателей полиномиальной функции является её **степень**. Степень полиномиальной функции - натуральное число n - наибольший показатель степени переменной x .

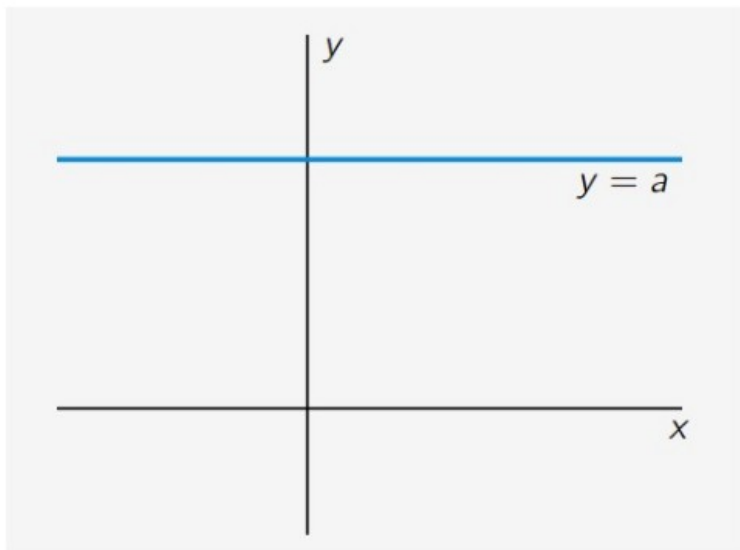
$$f(x) = 2x^6 + 3x^5 - \frac{3}{2}x - 3$$



Степень полинома = 6.

Степень определяет поведение функции при x стремящемся к бесконечности.

Степень = 0: функция-константа



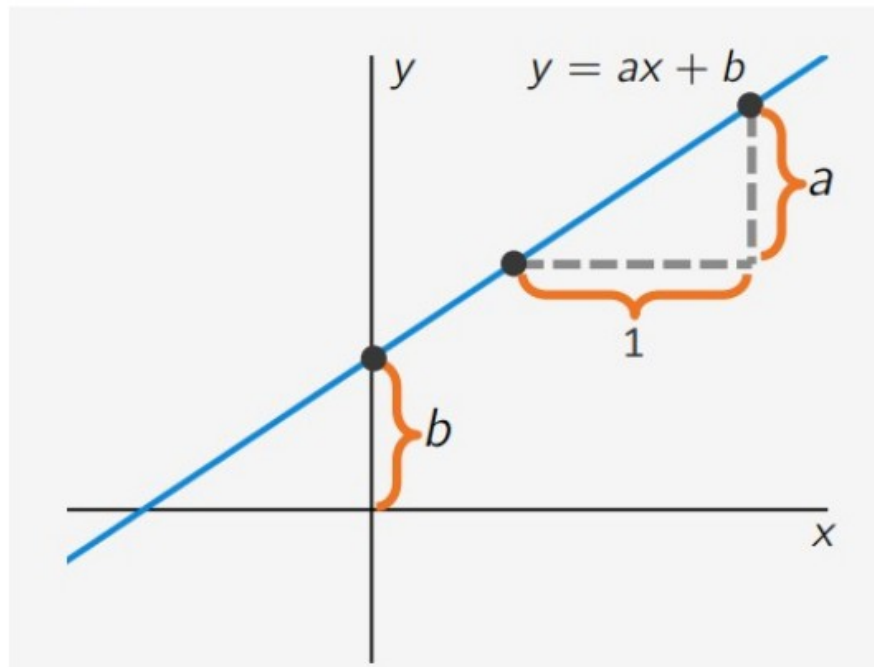
**Стандартная
формула:**

- $f(x) = a$

График:

- Горизонтальная
линия

Степень = 1: функция линейна



Стандартная формула:

- $f(x) = ax + b$

График:

- Прямая линия
- a — коэффициент пропорциональности
- b — смещение вдоль оси OY

Степень = 2: квадратичная функция

Пример: движение тела, подброшенное вверх.*



Формула изменения высоты во времени:

$$h(t) = -4,9t^2 + 6,1t + 1,4$$

Формула изменения траектории от

положения x :
$$h(x) = 3,3 - 0,4x^2$$

Стандартная формула:

$$f(x) = ax^2 + bx + c$$

График:

- парабола
- a — старший коэффициент
- b — младший коэффициент
- c — свободный коэффициент

*<https://images.app.goo.gl/h9hsAeWjxgHaHTvs8>

Степень = 2: пример

Golden Gate в Сан-Франсиско.*

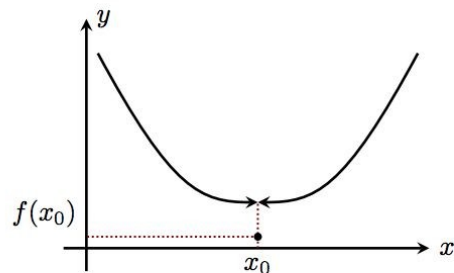


$$h(x) \approx 0,00037x^2 - 0,475x + 230$$

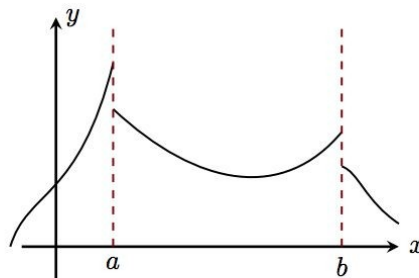
Высота над уровнем моря
X — в метрах от левого столба

Функции и их свойства

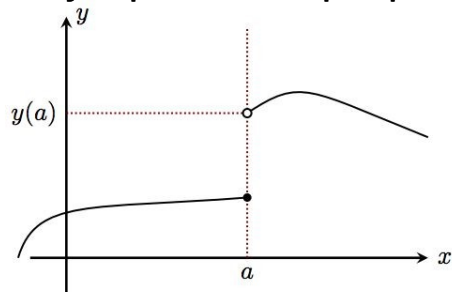
Представление о функции, её свойствах и поведении можно получить, построив ее график. Функции бывают непрерывными и разрывными.



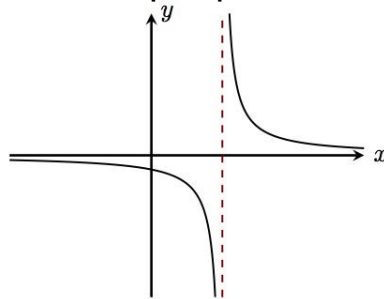
Функция с устранимым разрывом



Функция с разрывами в точках a и b



Функция с разрывом типа скачок



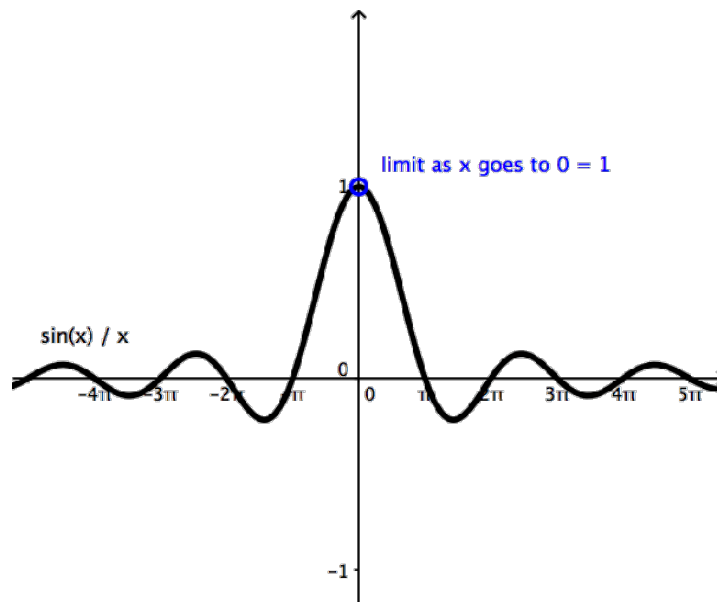
Функция с бесконечным разрывом

Предел функции

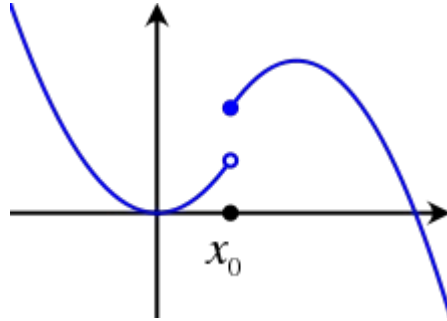
Предел (предельное значение функции) в заданной точке – такая величина, к которой стремится значение рассматриваемой функции при стремлении её аргумента к данной точке.

$$f(x) = \frac{\sin(x)}{x}$$

Хотя функция не определена при $x=0$, при стремлении $x \rightarrow 0$ её значение становится сколь угодно близко к 1



Предел функции



Предел при $x \rightarrow -x_0$ не равен пределу при $x \rightarrow +x_0$, следовательно, предела при $x \rightarrow x_0$ нет

Примеры пределов функций

$$f(x) = (1 + x)^{\frac{1}{x}}$$

Функция не определена в $x=0$, но её значение может быть вычислено в точках сколь угодно близких к ней

x	0.1	0.01	0.001	0.0001	...
$f(x)$	2.593..	2.704..	2.716..	2.718..	...

$$\lim_{x \rightarrow 0} (1 + x)^{\frac{1}{x}} = e.$$

Не у всех функций есть конечный предел

$$g(x) = \frac{1}{x}$$

Функция неограниченно растёт при приближении к $x = 0$

x	0.1	0.01	0.001	0.0001	...
$1/x$	10	100	1000	10000	...

$$\lim_{x \rightarrow 0} \frac{1}{x} = \infty.$$

Предел функции

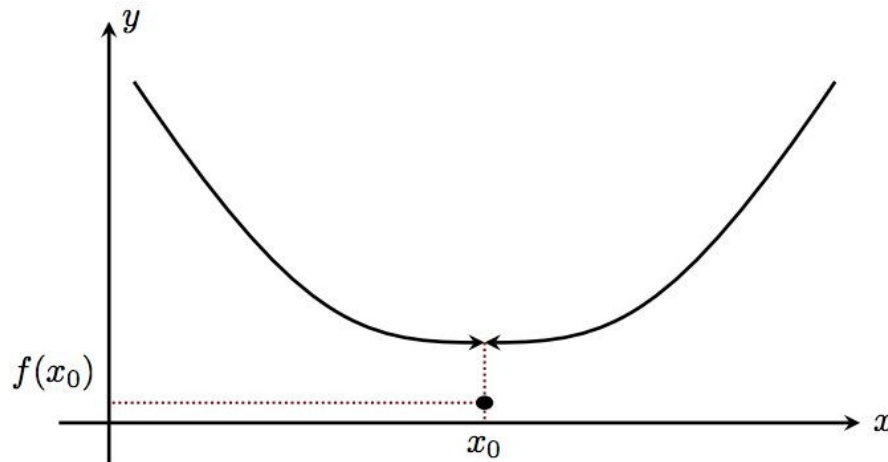
Понятие предела тесно связано с **понятием непрерывности** функции в точке.

Функция **непрерывна** в точке a , если:



$$\lim_{x \rightarrow a} f(x) = f(\lim_{x \rightarrow a} x) = f(a).$$

С помощью **понятия предела** определяется другое полезное понятие — **понятие производной**.



Производная функции

Производная - мгновенная скорость роста функции в заданной точке.

$$\frac{f(x + \Delta x) - f(x)}{\Delta x} = k.$$

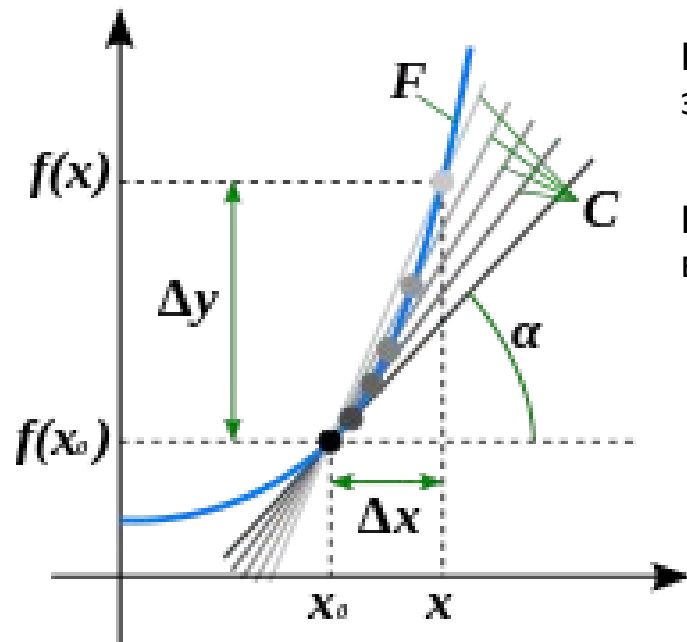
Давайте посмотрим на линейную функцию $y=kx+b$

Как понять скорость роста для произвольной функции? **Предел!**

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

Производная - предел отношения приращения функции к приращению её аргумента при стремлении приращения аргумента к нулю, если такой предел существует.

Геометрический смысл производной



Производная функции в точке x - это угловой коэффициент касательной в данной точке

Касательная представляет собой уравнение вида $kx+b=0$

$$f'(x) = \operatorname{tg}(\alpha) = k$$

Коэффициент $k = \operatorname{tg}$

Примеры производных

$$(c)' = 0, (x)' = 1$$

$$(x^\alpha)' = \alpha x^{\alpha-1}, x > 0, \alpha \in R$$

$$\left(\frac{1}{x}\right)' = -\frac{1}{x^2}, x \neq 0$$

$$(a^x)' = a^x \ln a, a > 0, a \neq 1$$

$$(\ln x)' = \frac{1}{x}, x > 0$$

$$(\log_a x)' = \frac{1}{x \ln a}, x > 0, a > 0, a \neq 1$$

$$(\sin x)' = \cos x$$

$$(\cos x)' = -\sin x$$

$$(\operatorname{tg} x)' = \frac{1}{\cos^2 x}, x \neq \frac{\pi}{2} + \pi n, n \in Z$$

$$(\operatorname{ctg} x)' = -\frac{1}{\sin^2 x}, x \neq \pi n, n \in Z$$

$$(\operatorname{sh} x)' = \operatorname{ch} x$$

$$(\operatorname{ch} x)' = \operatorname{sh} x$$

$$(\operatorname{th} x)' = \frac{1}{\operatorname{ch}^2 x}$$

$$(\operatorname{cth} x)' = -\frac{1}{\operatorname{sh}^2 x}$$

$$(\arcsin x)' = \frac{1}{\sqrt{1-x^2}}$$

$$(\arccos x)' = -\frac{1}{\sqrt{1-x^2}}$$

$$(\operatorname{arctg} x)' = \frac{1}{1+x^2}$$

$$(\operatorname{arcctg} x)' = -\frac{1}{1+x^2}$$

$$(\sqrt{x})' = \frac{1}{2\sqrt{x}}, x > 0$$

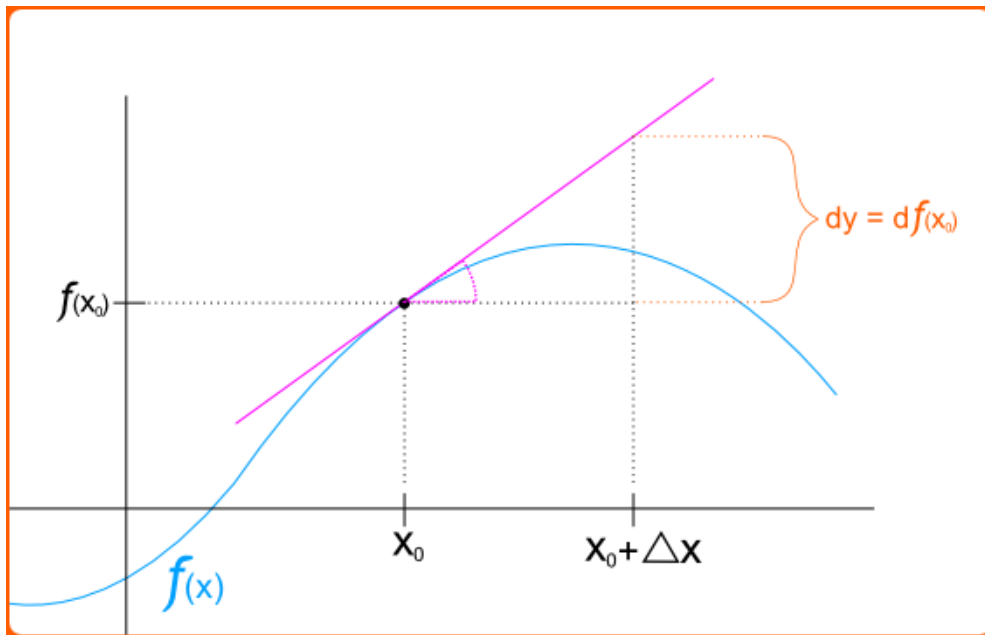
$$(e^x)' = e^x$$

$$(\ln x)' = \frac{1}{x}$$

$$(|x|)' = \operatorname{sign} x = \begin{cases} 1, & x > 0, \\ -1, & x < 0, \end{cases} x \neq 0$$

Дифференциал функции

Дифференциал - линейная часть приращения функции



Сложная функция

Функция h сложная, составленная из функций g и f , если $h(x) = g(f(x))$

$f(x)$ – внутренняя

$g(x)$ - внешняя

Производная сложной функции

$$h'(x) = g'(f) \cdot f'(x)$$

Алгоритм вычисления:

- 1) определить внутреннюю $f(x)$ и внешнюю функции $g(x)$
- 2) найти производную внутренней функции $f'(x)$
- 3) найти производную внешней функции $g'(x)$
- 4) перемножить производную внутренней и внешней функции

Производная сложной функции (пример)

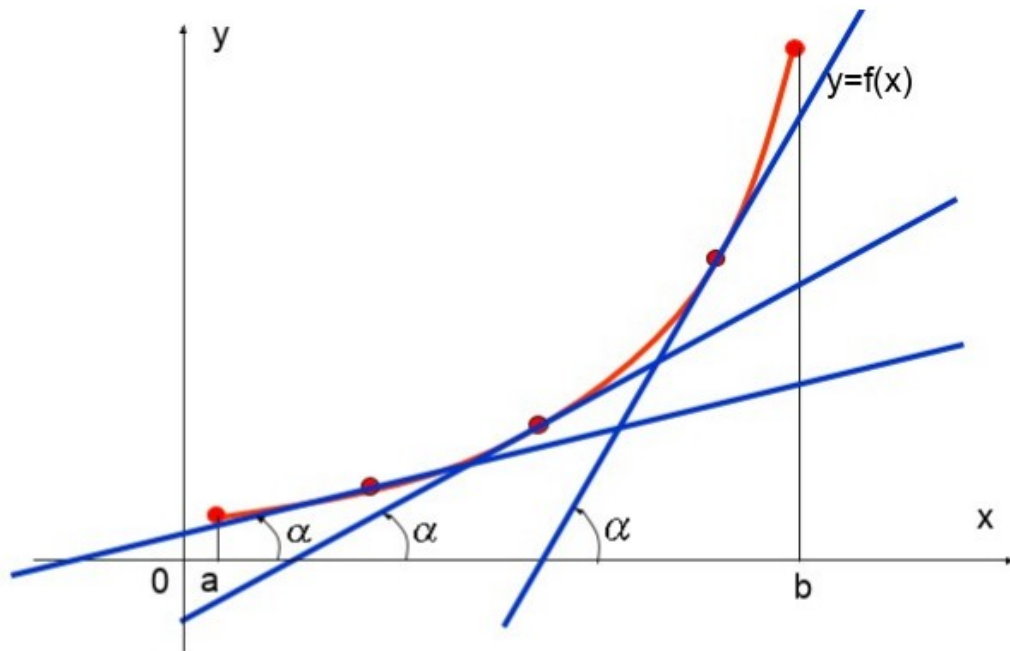
$$f(x) = \sin(\ln(x)+5x)$$

$$f'(x) = \cos(\ln(x)+5x) * (\ln(x)+5x)'$$

$$f'(x) = \cos(\ln(x)+5x) * (1/x + 5)$$

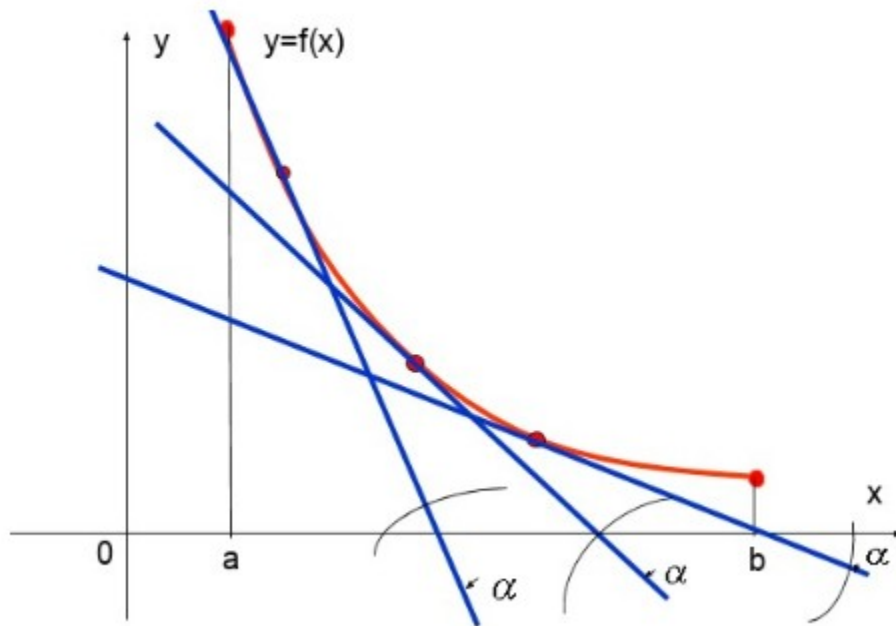
Исследование функции с помощью производной

Если $f'(x) > 0$
на интервале,
то функция возрастает
на данном промежутке



Исследование функции с помощью производной

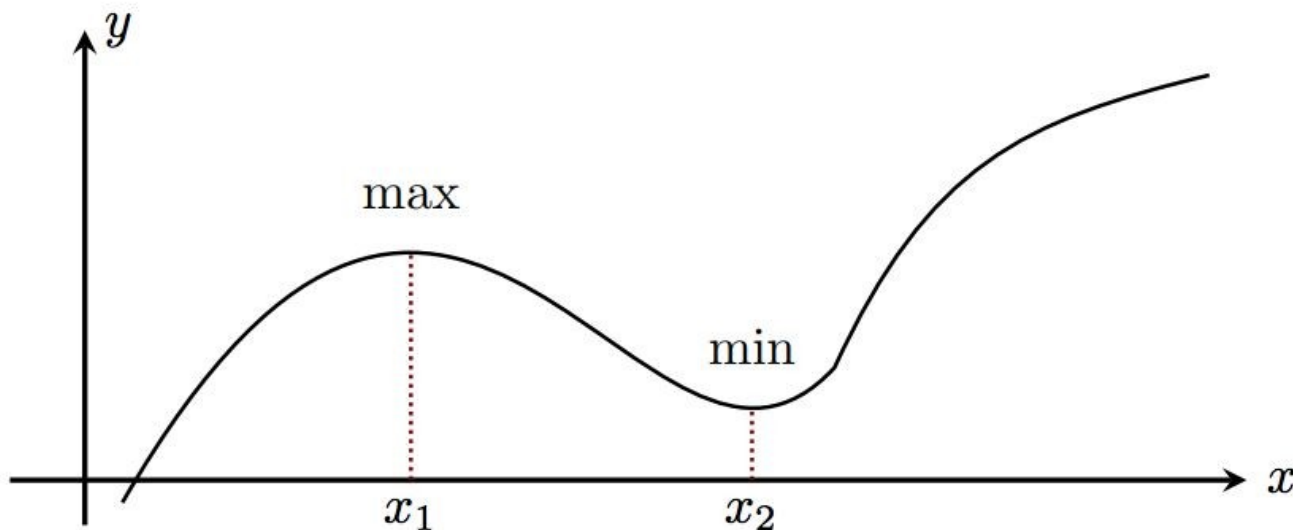
Если $f'(x) < 0$
на интервале,
то функция убывает
на данном промежутке



Экстремум функции

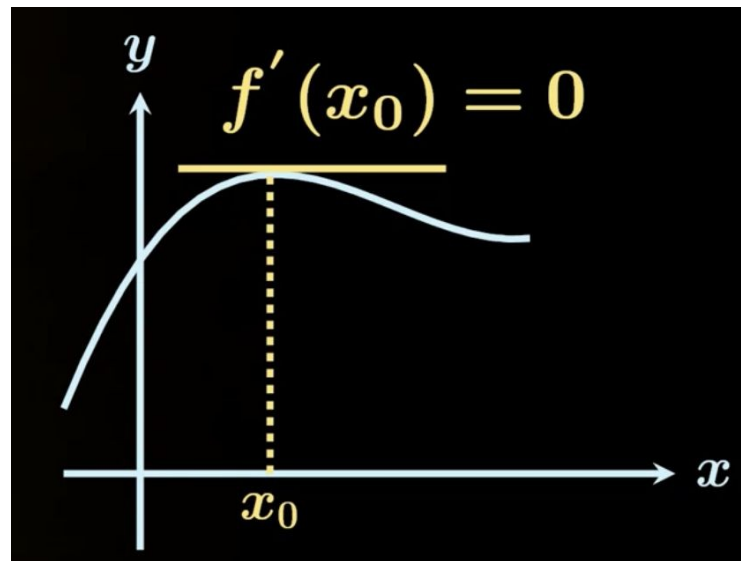
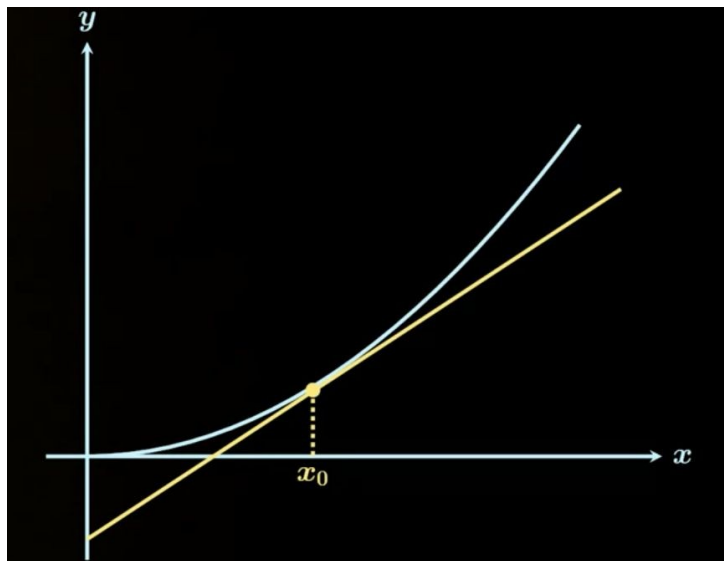


Точка x_0 - называется **локальным минимумом** функции $f(x)$, если существует такая окрестность $U(x_0)$, для которой $f(x) > f(x_0)$, x из $U(x_0)$. Аналогично для максимума. В случае глобального минимума $U(x_0) = D(f)$.



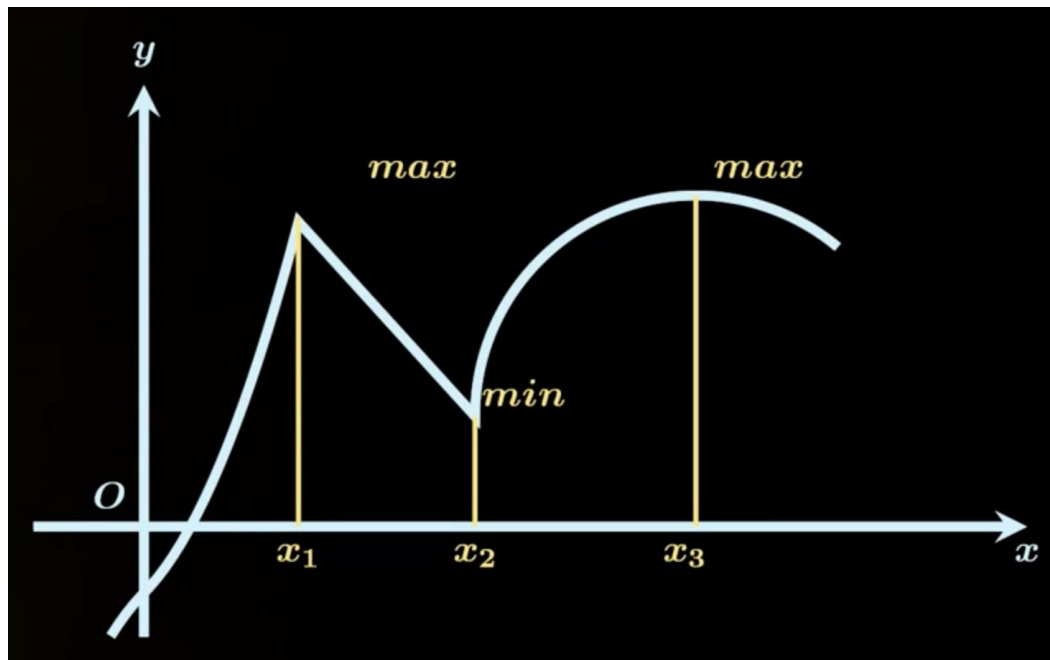
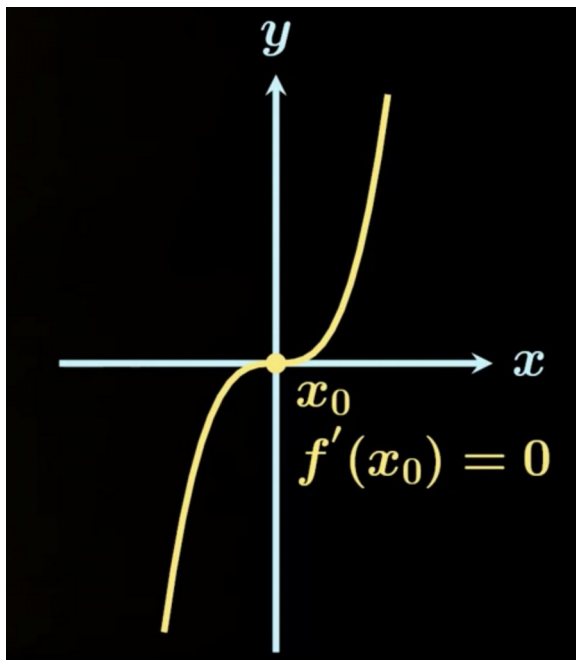
<https://github.com/rtriangle/Netology-statistics>

Экстремум функции и производная



В точках локальных экстремумов производная (если она определена (!), пример дальше) обязана равняться нулю.
Это **необходимое** условие.

Экстремум функции и производная

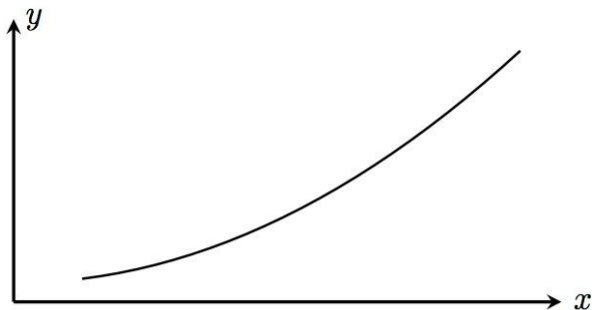


Однако равенство нулю производной **не является достаточным** условием локального экстремума. Также производная может быть вовсе не определена в точках локальных экстремумов.

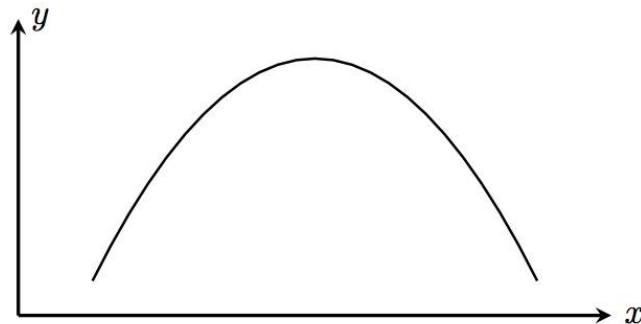
Выпуклость функции и вторая производная

Давайте спустимся на уровень ниже: какому свойству функции соответствует монотонная производная?

1. $f''(x) \geq 0$ — функция $f(x)$ выпукла,
2. $f''(x) > 0$ — функция $f(x)$ строго выпукла,
3. $f''(x) \leq 0$ — функция $f(x)$ вогнута,
4. $f''(x) < 0$ — функция $f(x)$ строго вогнута.



Выпуклая



Вогнутая

Выпуклость функции и вторая производная

Помните **необходимое** условие локального экстремума?



Наложив некоторые условия на вторую производную, можно сделать их **достаточными**!

Достаточное условие экстремума Пусть выполнено необходимое условие экстремума, то есть в некоторой точке x_0 значение $f'(x_0) = 0$. Если в таком случае

1. $f''(x) > 0$ — функция будет строго выпукла и реализуется строгий минимум.
2. $f''(x) < 0$ — функция будет строго вогнута и реализуется строгий максимум.

Функция нескольких переменных

Пусть теперь x - не число из R , а вектор (x_1, \dots, x_n) , где каждое x_i из R , а весь x из R^n .

$D(f)$ - область определения функции (ничего не изменилось!)

$E(f)$ - область значений функции (ничего не изменилось!)

Снова будем работать только с функциями, у которых $D(f)$ и $E(f)$ - подмножество R^n .

Частная производная



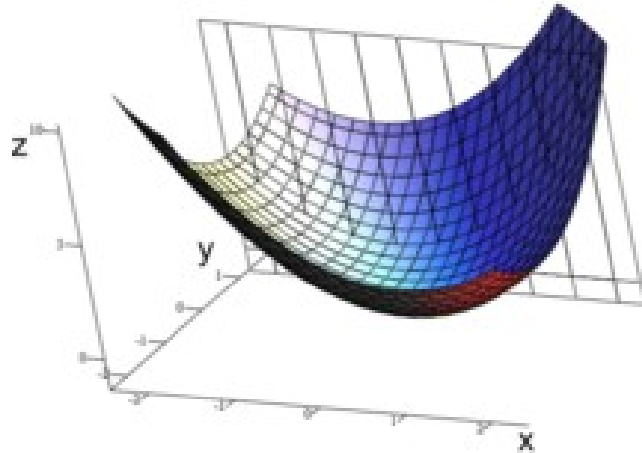
Частная производная — это одно из обобщений понятия производной на случай функции нескольких переменных.



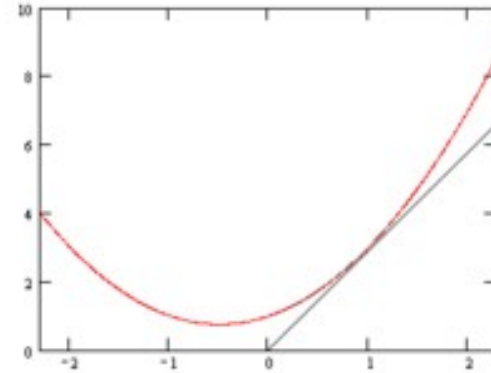
Частная производная функции $f(x, y)$ по x определяется как производная по x , взятая в смысле функции одной переменной, при условии **постоянства** оставшейся переменной y .

$$f'_x(x, y) = \lim_{h \rightarrow 0} \frac{f(x + h, y) - f(x, y)}{h}, \quad f'_y(x, y) = \lim_{h \rightarrow 0} \frac{f(x, y + h) - f(x, y)}{h}.$$

Частная производная



$$z = x^2 + xy + y^2$$



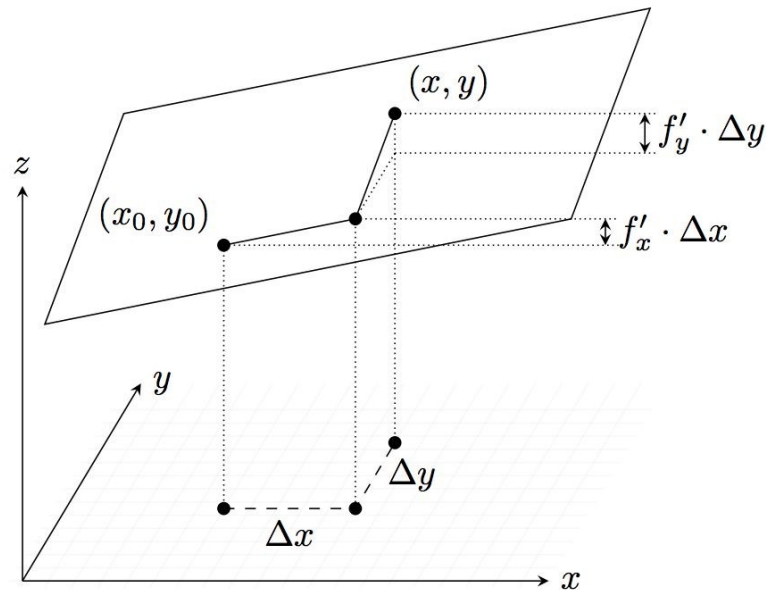
Сечение графика плоскостью $y=1$

Частная производная в точке $(1, 1, 3)$ при постоянном y соответствует углу наклона касательной прямой, параллельной плоскости xz

Касательная плоскость

Пусть дана некоторая функция двух переменных $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. Она, вообще говоря, определяет некоторую поверхность $z = f(x, y)$ в трехмерном пространстве.

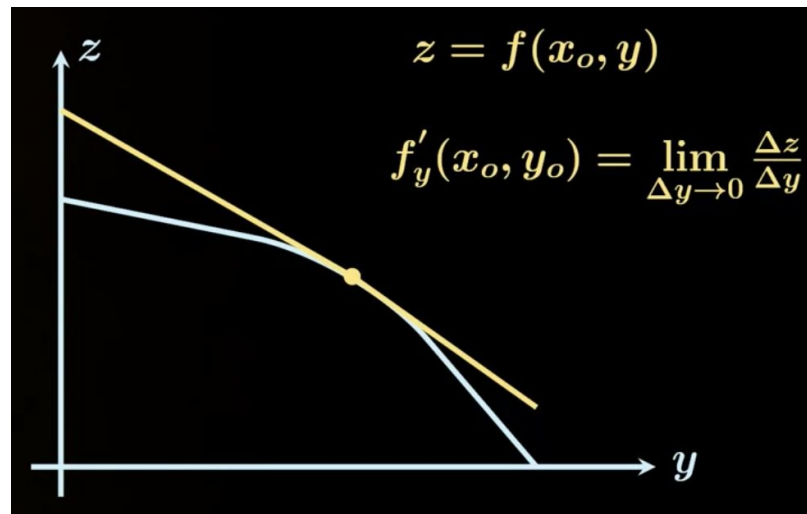
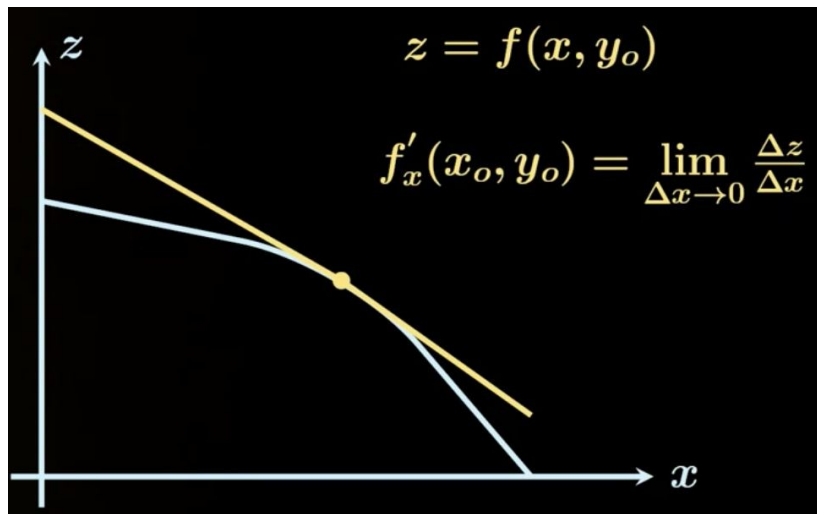
Если в некоторой точке (x_0, y_0) функция дифференцируема как функция многих переменных, то в этой точке можно рассмотреть **касательную плоскость** к данной поверхности.



Касательная плоскость

Таким образом, график функции $f(x, y)$ в окрестности точки можно приблизить касательной плоскостью:

$$f(x_0 + \Delta x, y_0 + \Delta y) \approx f(x_0, y_0) + f'_x(x_0, y_0)\Delta x + f'_y(x_0, y_0)\Delta y.$$



Градиент и линии уровня функции

Если $f(x_1, \dots, x_n)$ — функция n переменных x_1, \dots, x_n , то n -мерный вектор из частных производных:

$$\text{grad } f = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right)$$



называется **градиентом функции**.



Линией уровня функции называется множество точек, в которых функция принимает одно и то же фиксированное значение. Оказывается, что **градиент перпендикулярен линии уровня**.

Градиент в задачах оптимизации

Задачей оптимизации называется задача по нахождению экстремума функции, например минимума:

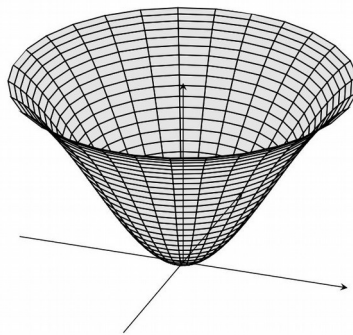
$$f(x_1, \dots, x_n) \rightarrow \min$$

Такая задача часто встречается в приложениях, например при выборе оптимальных параметров рекламной компании, а также в задачах классификации.

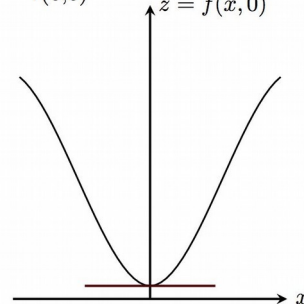
Вспомним необходимые условия экстремума из прошлой лекции!

Градиент в задачах оптимизации

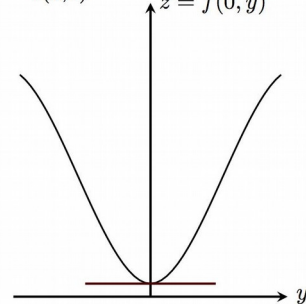
Но не всегда задачу можно решать аналитически. В таком случае используется численная оптимизация. Наиболее простым в реализации из всех методов численной оптимизации является метод градиентного спуска.



$$\left. \frac{\partial f}{\partial x} \right|_{(0,0)} = 0$$



$$\left. \frac{\partial f}{\partial y} \right|_{(0,0)} = 0$$



Функция двух переменных достигает минимума в начале координат

Градиентный спуск

Это итерационный метод. Решение задачи начинается с выбора начального приближения $\vec{x}^{[0]}$

После вычисляется приблизительное значение \vec{x}^1

Затем \vec{x}^2

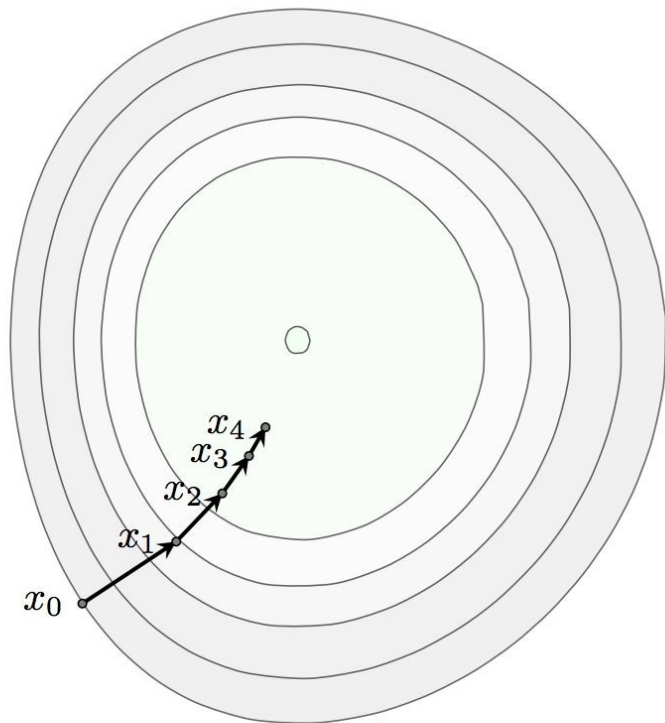
и так далее...



$\vec{x}^{[j+1]} = \vec{x}^{[j]} - \gamma^{[j]} \nabla F(\vec{x}^{[j]}),$ где $\gamma^{[j]}$ — шаг градиентного спуска.

Идея: идти в направлении наискорейшего спуска, а это направление задаётся антиградиентом $-\nabla F$.

Градиентный спуск



Аналогия: домик в низине

Если заблудились, то верным решением будет двигаться в направлении наискорейшего спуска

Спасибо за внимание!