Омский Государственный ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Онлайн-курс

ВВЕДЕНИЕ В ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ

1. Функции одного аргумента и их точки минимума

Автор: Шевляков Артём Николаевич

Точки минимума функции

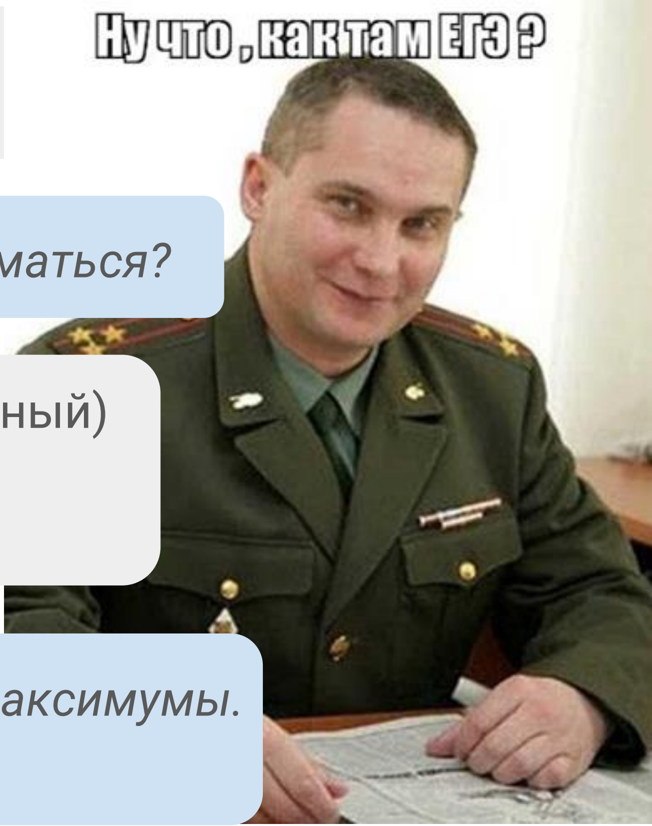
Ну, что вы помните из школы?

Есть функция *f*(*x*) аргумента *x*.
Можешь найти ее точки минимума и максимума?

А зачем? Мы же нейронными сетями будем заниматься?

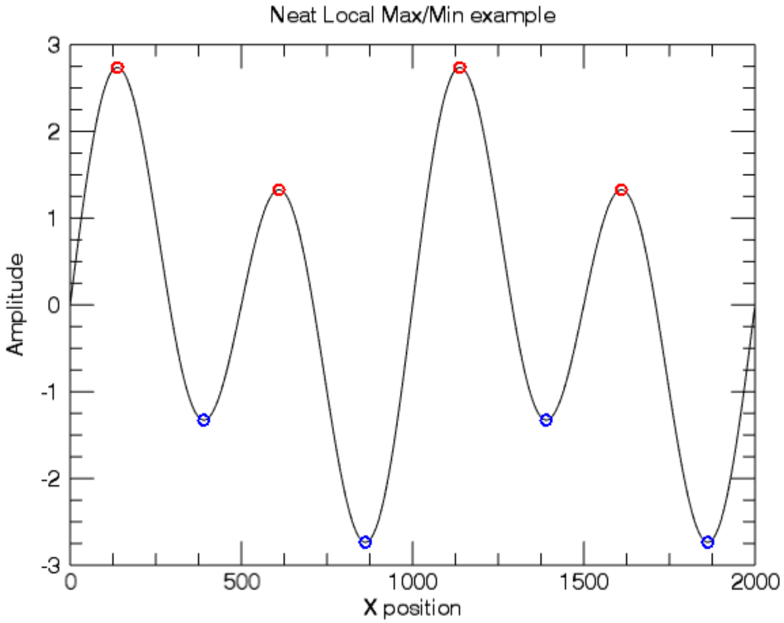
Понимаешь, любой мозг (искусственный и натуральный) есть результат некоторой минимизации и поиска оптимальной структуры.

ОК. Значит, будем учиться искать минимумы и максимумы. Это же как-то связано с производной, верно?



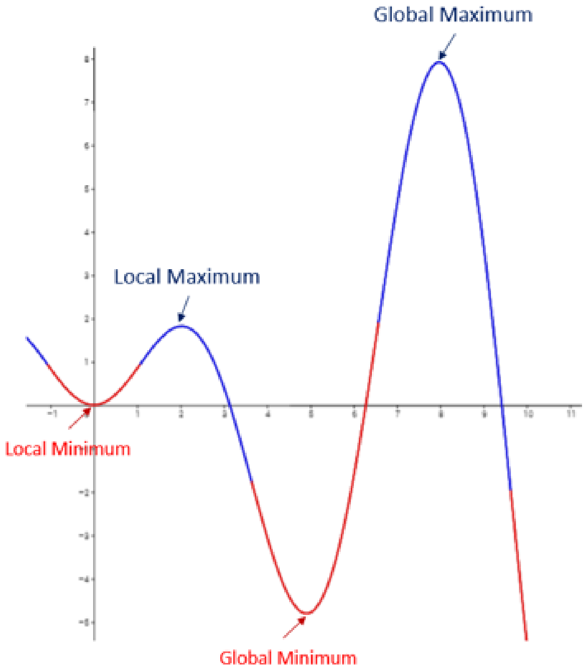
Сначала определение минимума

Точка *a* называется **локальным минимумом** функции *f(x)*, если существует окрестность точки *a*, что значение *f(a)<f(b)* для любой точки *b* из этой окрестности.



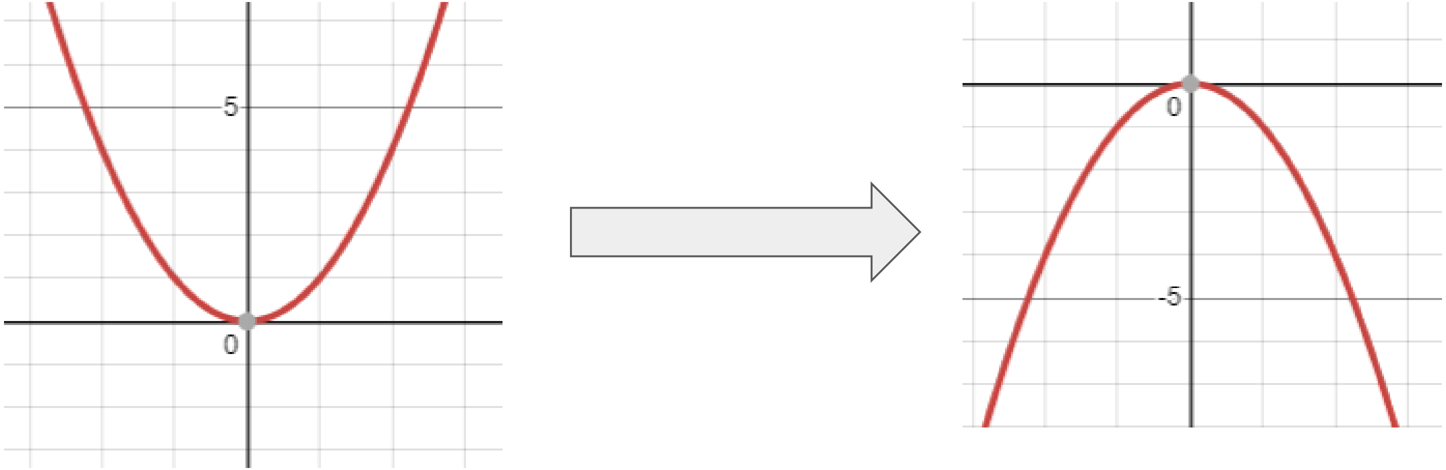
Сначала определение минимума

Точка *a* называется **глобальным минимумом** функции *f(x)*, если *f(a)<f(b)* для любой точки *b* из области определения функции.



А что точки максимума?

Можно не уметь искать точки максимума, так как точки максимума функции *f(x)* являются точками минимума функции -*f(x)*. И наоборот.



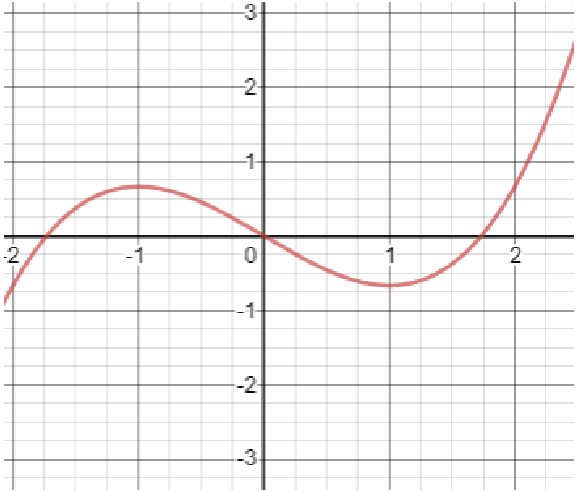
Ищем точки минимума, как в школе

- Есть функция $f(x)$.
- Находим ее производную $f'(x)$.
- Решаем уравнение $f'(x)=0$.
- Корни этого уравнения – кандидаты на звание «локального минимума».

Как же всё просто!

Ищем точки минимума, как в школе. Пример

- Есть функция $f(x)=\frac{1}{3}x^3-x$.
- Находим ее производную $f'(x)=x^2-1$.
- Решаем уравнение $x^2-1=0$.
- Корни этого уравнения $x=-1$, $x=1$. Точка x=1 – локальный минимум, (да-да, в этой точке производная меняет знак с - на +).



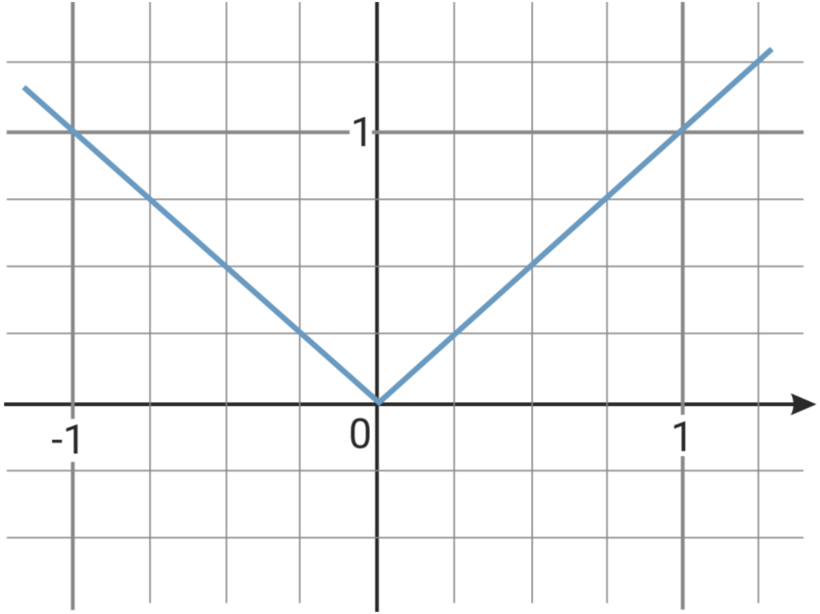
Но есть небольшая проблемка...

Указанный метод поиска точек минимума не работает, если…

- **производная определена не всюду**

(особенно весело, когда производная не определена как раз в точке минимума);

например: f(x)=|x|;

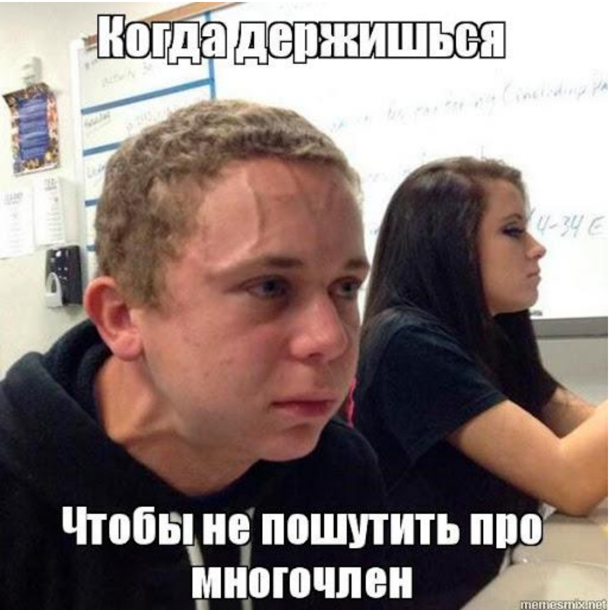


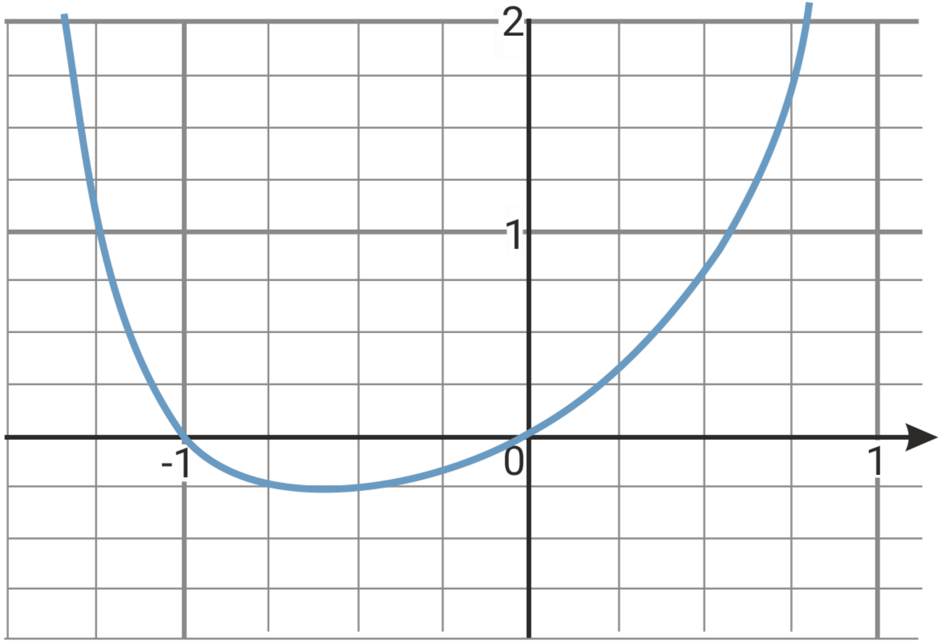
Но есть небольшая проблемка...

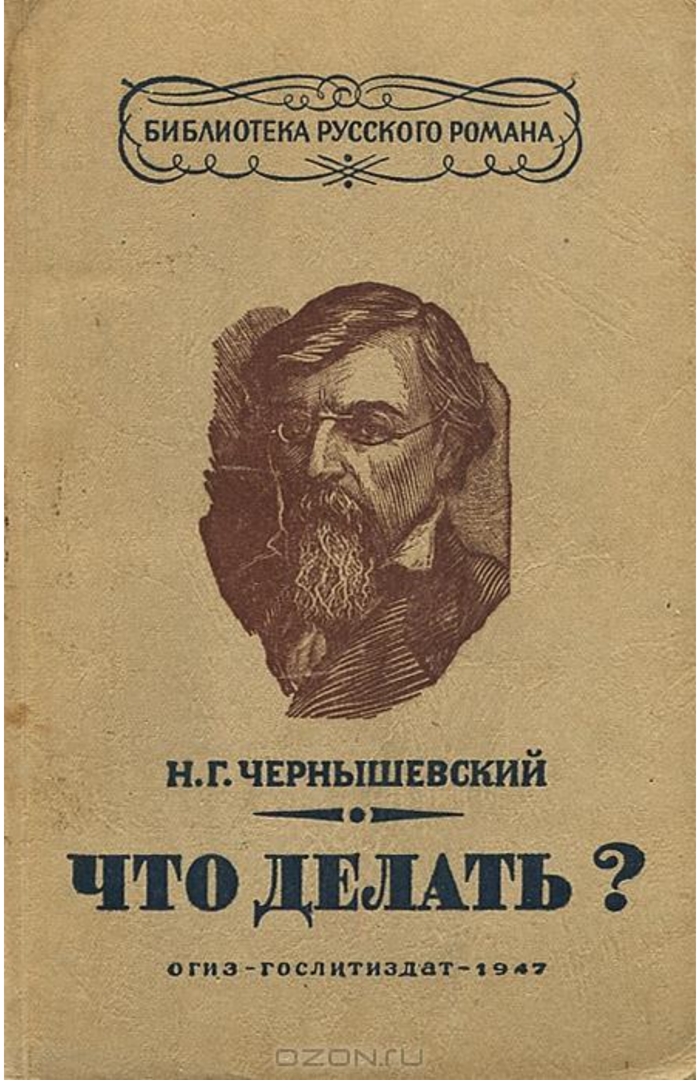
Указанный метод поиска точек минимума не работает, если…

- **уравнение *f'(x)=0* не может быть точно решено;**

например: $f(x)=x^6+x^5+x^2+x$






БИБЛИОТЕКА РУССКОГО РОМАНА
Н. Г. ЧЕРНЫШЕВСКИЙ
ЧТО ДЕЛАТЬ?
ОГИЗ - ГОСЛИТИЗДАТ - 1947
OZON.ru

Метод градиентного спуска

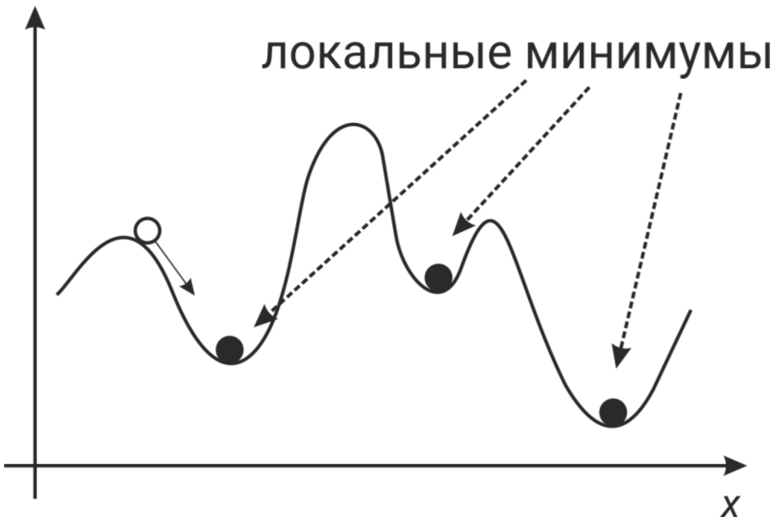
Итак, школьный метод не всегда работает

Что же делать?

Созерцать!

Обычный мячик под действием силы тяжести скатывается в самую глубокую область поверхности.

Неужели мы тупее мячика и не сможем смоделировать такое движение по графику функции?



Моделируем поведение шарика

Выберем длину шага *h* (например, *h=0.01*). Сделав шаг такой длины наш шарик будет останавливаться и думать, в какую сторону ему дальше катиться.

Пусть шарик находится в точке с координатой a. Тогда новая a точка равна:

$$a := a - f'(a) \cdot h$$

После этого процесс повторяется, мы получаем новую позицию для шарика итд.

Моделируем поведение шарика

В общем случае этот процесс можно описать формулой. Пусть a_n - координата шарика после *n* шагов. Тогда координата шарика на следующем шаге равна:

$$a_{n+1} := a_n - h \cdot f'(a_n)$$

Стандартное значение шага *h=0.01.*

Когда нужно остановиться?

Когда значение $f'(a_n)$ будет (примерно) равно *0*. Либо когда пройдет заданное заранее максимальное число итераций.

Пример

Пусть *f(x)=x³-x*. Начальное положение *a₀=0*, шаг *h=0.1*, *f'(x)=3x²-1*.

Следующая координата равна:

$$a_1 = a_0 - f'(a_0)h = 0 - (-1)0.1 = 0.1$$

Далее:

$$a_2 = a_1 - f'(a_1)h = 0.1 - (-0.97)0.1 = 0.197$$

Потом:

$$a_3 = 0.197 - (-0.88)0.1 = 0.28$$

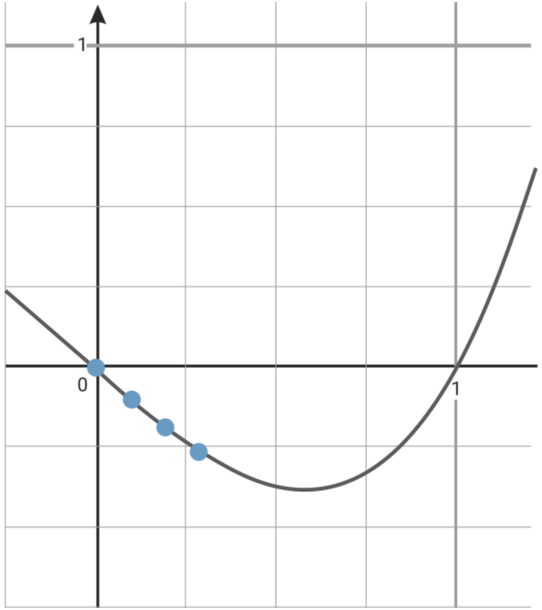
и так далее...

Посмотрим результат процесса на графике функции...

Пример

Мы действительно по шагам подходим к точке минимума!

Кстати, после 13-ти итераций мы придем в точку с координатой $a_{13}=0.57$, при этом точная точка минимума 0.577



Пример

Информацию о шагах нашего алгоритма приведем в виде таблицы

a_n	0.1	0.197	0.285	0.361	0.422	0.469	0.503	0.527	0.544	0.555	0.563	0.568	0.571
$f'(a_n)$	-0.97	-0.884	-0.756	-0.609	-0.466	-0.34	-0.241	-0.167	-0.112	-0.076	-0.049	-0.032	-0.022
$f(a_n)$	-0.1	-0.19	-0.26	-0.31	-0.35	-0.37	-0.38	-0.38	-0.38	-0.38	-0.38	-0.38	-0.38

Можно заметить:

- производная (по модулю) уменьшается и приближается к 0;
- значения функции уменьшаются;
- точка a_n приближается к истинной точке минимума;
- расстояние между соседними точками a_n уменьшается.

$$a_{n+1} := a_n - h \cdot f'(a_n)$$

Метод градиентного спуска

Такой способ нахождения точки минимума называется **градиентным спуском (ГС)**. И он работает, даже когда школьный метод ломается.

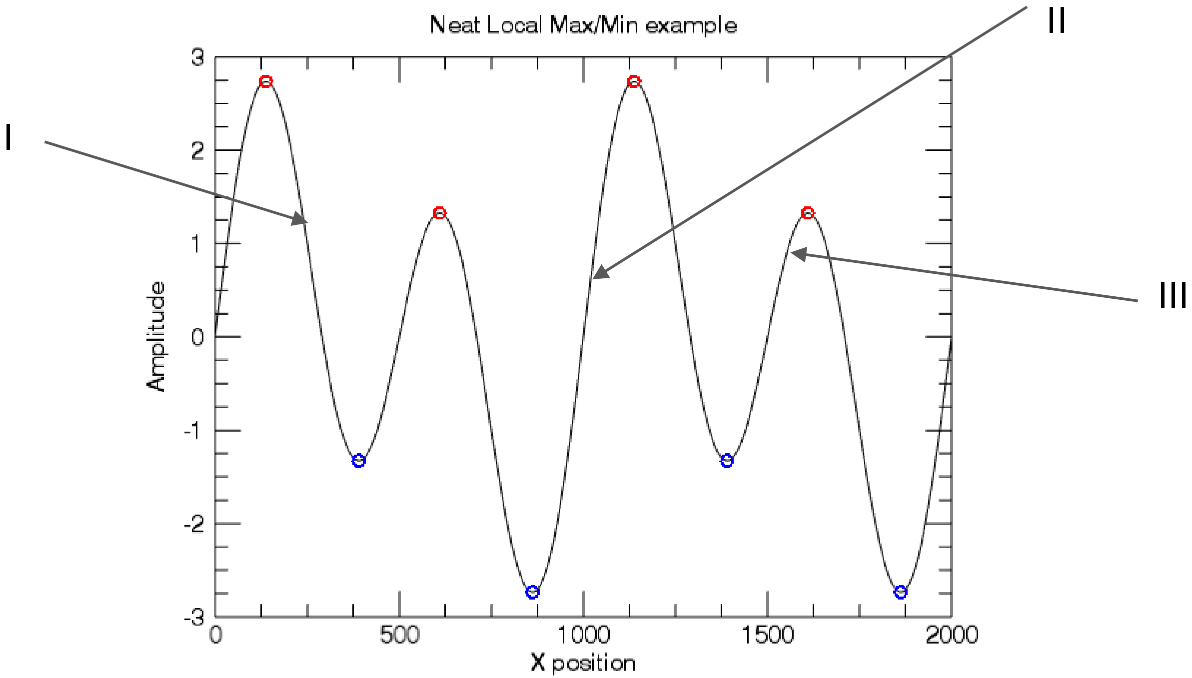
Например, если $f(x)=|x|$, то $$f'(x) = \begin{cases} 1, & x > 0 \\ -1, & x < 0 \end{cases}$$ и ГС работает.

Для $a_0=1, h=0.1$ имеем $a_1=1-1*0.1=0.9$, $a_2=0.9-1*0.1=0.8$, $a_3=0.8-1*0.1=0.7$ и тд.

В конечном итоге мы придём в истинную точку минимума $x=0$.

Недостатки ГС

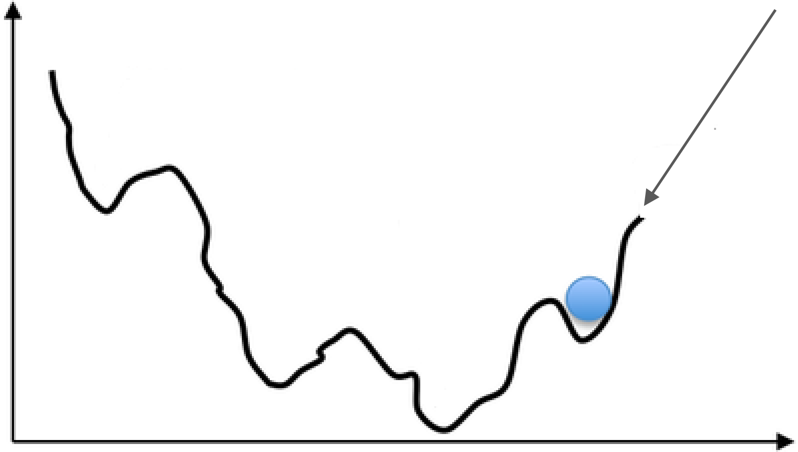
- всё зависит от расположения начальной точки a_0.



Недостатки ГС

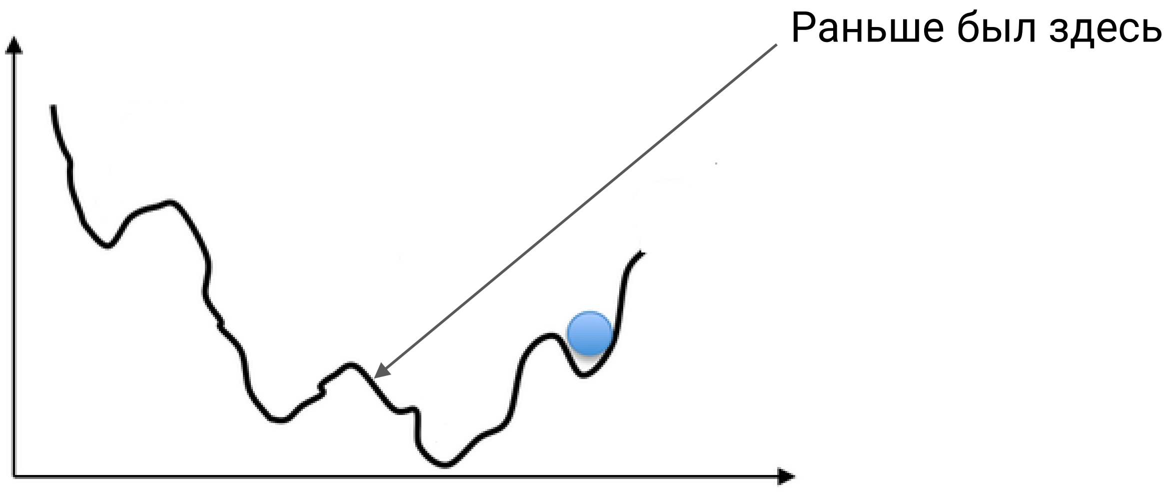
- **слишком маленький шаг *h* плох;**

Раньше был здесь



Недостатки ГС

- **слишком большой шаг *h* плох;**



Недостатки ГС

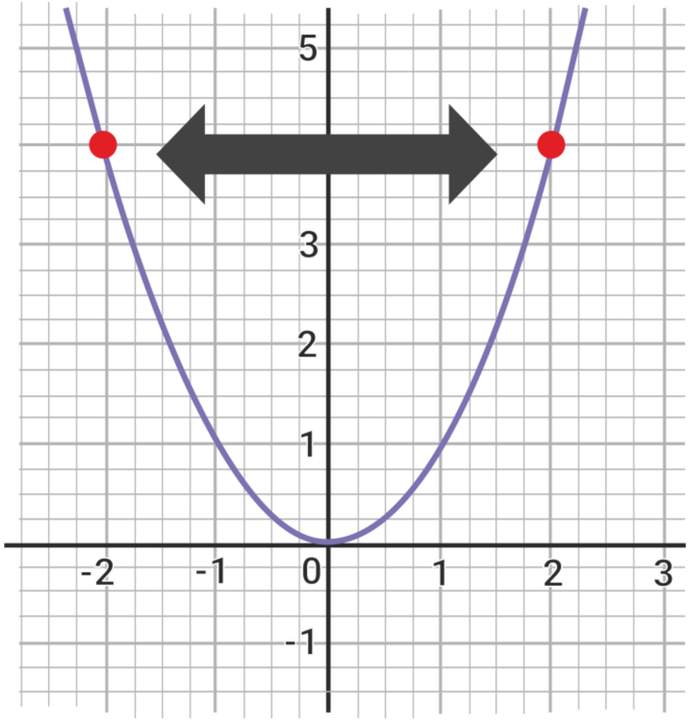
- **точность зависит от длины шага *h*;**

Пусть $f(x)=x^2$, $f'(x)=2x$, $a_0=-2$, $h=1$.

Тогда $a_1 = -2 - 2 \cdot (-2) \cdot 1 = 2$

$$a_2 = 2 - 2 \cdot 2 \cdot 1 = -2$$

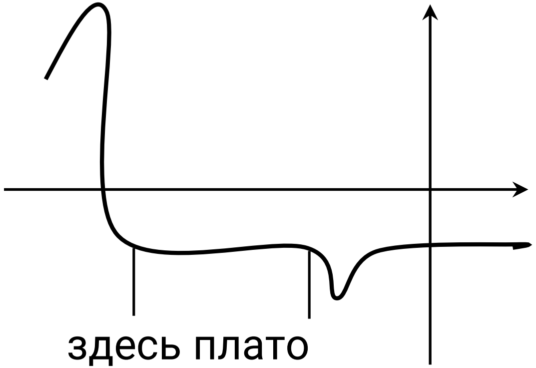
Мы зациклились!



Недостатки ГС

- **плохое поведение на плато функции *f(x)*.**

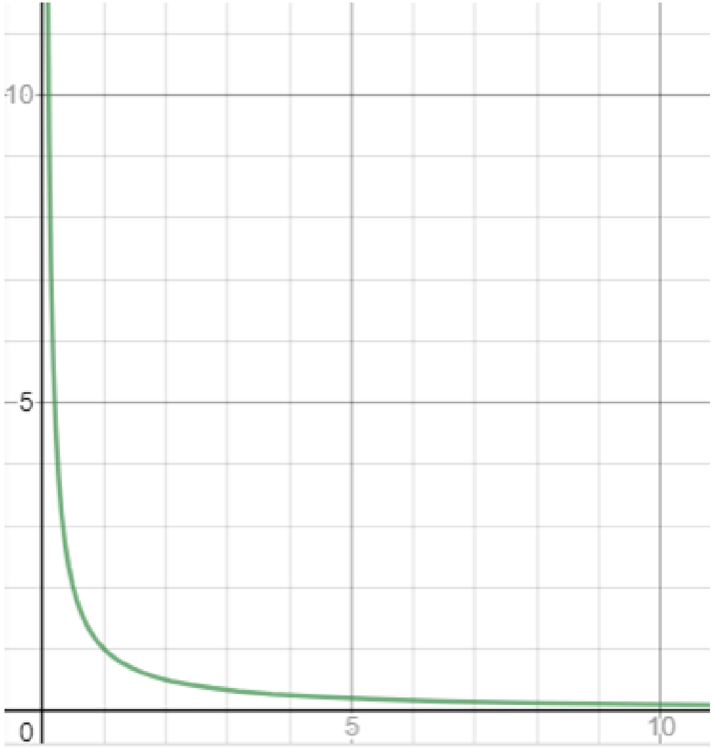
На плато *f'(x)* по модулю почти равна нулю (это еще называют **затуханием градиента**). Нет стимула хоть куда сделать шаг.



Шарик, кстати, по плато тоже не катится))))

Недостатки ГС

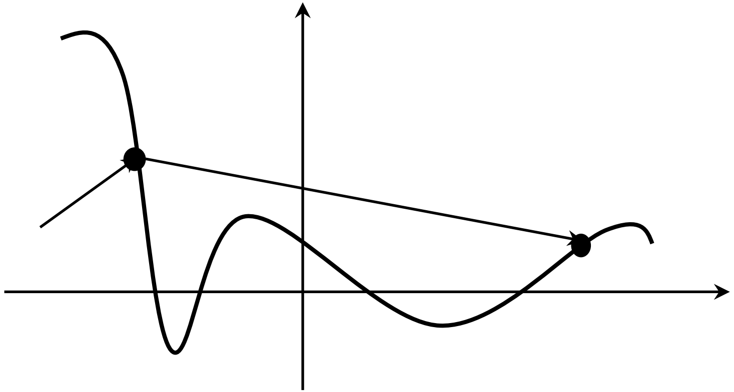
Например, функция $f(x)=1/x$ имеет плато при больших значениях x.
В частности, значение производной $f'(x)=-1/x^2$ в точке 10 равно: -0.01



Недостатки ГС

- **взрыв градиента**, когда значение *f'(a)* очень велико.

Это означает, что следующий шаг будет очень большой (это следует из формулы *a:=a-f'(a)*h*)!



очень большая по модулю f'(x) в левой точке

Преимущества ГС

- Других общих методов нахождения точек минимума у нас нет.

- Проблемы с выбором h вполне решаемы: мы можем изменять величину h на каждой итерации. На первых итерациях h может быть достаточно большим, и потом плавно уменьшаться.

- Метод может работать, даже когда производная неизвестна! См. след. слайд.

Численное дифференцирование

Пусть δ – небольшое положительное число. Тогда производная в точке a приближенно равна:

$$f'(a) \approx \frac{f(a+\delta) - f(a)}{\delta}$$

Например, для $f(x)=x^3$ при $\delta=0.05$ в точке $a=1$ имеем:

$$f'(1) \approx \frac{f(1+0.05)^3 - 1^3}{0.05} = 3.15$$

Истинное значение $(f'(x)=3x^2)$ равно: $3*1^2=3$

Численное дифференцирование

А еще более точная формула для дифференцирования такая:

$$f'(a) \approx \frac{f(a-2\delta) - 8f(a-\delta) + 8f(a+\delta) - f(a+2\delta)}{12\delta}$$

Например, для *f(x)=x³* при δ =0.05 в точке *a=1* имеем:

$$f'(1) \approx \frac{f(1-2\cdot 0.05)^3 - 8(1-0.05)^3 + 8(1+0.05)^3 - (1+2\cdot 0.05)^3}{12\cdot 0.05}$$

$$f'(1) \approx 3.01$$

Что достаточно близко к истинному значению *(f'(1)=3)*.

Выводы

Выводы:

- Мы изучили новый метод нахождения точки минимума – градиентный спуск.
- Обсудили его плюсы и минусы.
- Показали на примерах, как градиентный спуск итеративно ищет точку минимума.