

В отличие от целых чисел, вещественные числа в языке Питон имеют ограниченную длину.

Подумаем, как хранить десятичную дробь в памяти. Поскольку вещественных чисел бесконечно много (даже больше, чем натуральных), то нам придется ограничить точность. Например, мы можем хранить только несколько первых значащих цифр, не храня незначащие нули. Будем отдельно хранить целое число с первыми значащими цифрами и отдельно хранить степень числа 10, на которую нужно умножить это число.

Например, число  $5.972 \cdot 10^{24}$  (это масса Земли в килограммах) можно сохранить как 5972 (цифры числа, мантисса) и 21 (на какую степень 10 нужно умножить число, экспонента). С помощью такого представления можно хранить вещественные числа любой размерности.

Примерно так и хранятся числа в памяти компьютера, однако вместо десятичной системы используется двоичные. На большинстве аппаратных систем в языке Питон для хранения float используется 64 бита, из которых 1 бит уходит на знак, 52 бита - на мантиссу и 11 бит - на экспоненту. Это не совсем правда, но достаточно неплохо описывает реальность.

53 бита дают около 15-16 десятичных знаков, которые будут храниться точно. 11 бит на экспоненту также накладывает ограничения на размерность хранимых чисел (примерно от  $-1000$  до  $1000$  степени числа 10).

Любое вещественное число на языке Питон представимо в виде дроби, где в числителе хранится целое число, а в знаменателе находится какая либо степень двойки. Например, 0.125 представимо как  $1/8$ , а 0.1 как  $3602879701896397/36028797018963968$ . Несложно заметить, что эта дробь не равно 0.1, т.е. хранение числа 0.1 точно в типе float невозможно, как и многих других "красивых" десятичных дробей.

В целом будет полезно представлять себе вещественное число  $X$  как отрезок  $[X - \epsilon; X + \epsilon]$ . Как же определить величину  $\epsilon$ ?

Для этого нужно понять, что погрешность не является абсолютной, т.е. одинаковой для всех чисел, а является относительной. Упрощенно, аппаратную погрешность хранения числа  $X$  можно оценить как  $X \cdot 2^{-54}$ .

Чаще всего в задачах входные данные имеют определенную точность. Рассмотрим на примере: заданы два числа  $X$  и  $Y$  с точностью 6 знаков после точки (значит  $\epsilon = 5 \cdot 10^{-7}$ ) и по модулю не превосходящие  $10^9$ . Оценить абсолютную погрешность вычисления  $X \cdot Y$ . Рассмотрим худший случай, когда  $X$  и  $Y$  равны  $10^9$  и отклонились на максимально возможное значение  $\epsilon$  в одну сторону. Тогда результат вычисления будет выглядеть так:

$$(X + \epsilon) \cdot (Y + \epsilon) = XY + (X + Y) \cdot \epsilon + \epsilon^2$$

Величина  $\epsilon^2$  пренебрежимо мала,  $XY$  - это правильный ответ, а  $(X + Y) \cdot \epsilon$  - искомое значение абсолютной погрешности. Подставим числа и получим:

$$2 \cdot 10^9 \cdot 5 \cdot 10^{-7} = 10^3$$

Абсолютная погрешность вычисления составила 1000 (одну тысячу). Что довольно неожиданно и грустно.

Таким образом, становится понятно, что нужно аккуратно вычислять значение погрешности для сравнения вещественных чисел.

Пометить как выполненное

