Python Numeric Types

DS 5110/CS 5501: Big Data Systems

Spring 2024

Lecture 2b

Yue Cheng

University of Virginia

Some material taken/derived from:

- Wisconsin CS 544 by Tyler Caraza-Harter.

Learning objectives

- Know how machine stores floats
- Compare different numeric types in terms of memory space cost, range, and precision

Python numeric types (built in)

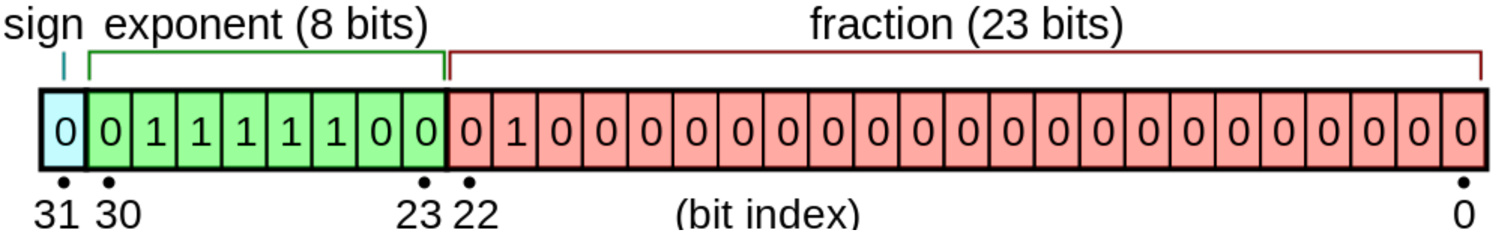
https://docs.python.org/3/library/stdtypes.html#numeric-types-int-float-complex

Python numeric types

- int
  - No max/min size (Python is unusual in this way)
  - Bigger values -> more bits necessary
- float
  - Defaults 64 bits (double precision)
    - You can also use float32 given a certain framework (e.g., PyTorch, numpy, etc.)
  - Most pre-trained ML models use float32 for parameters

float32

- Standard IEEE format (float32)

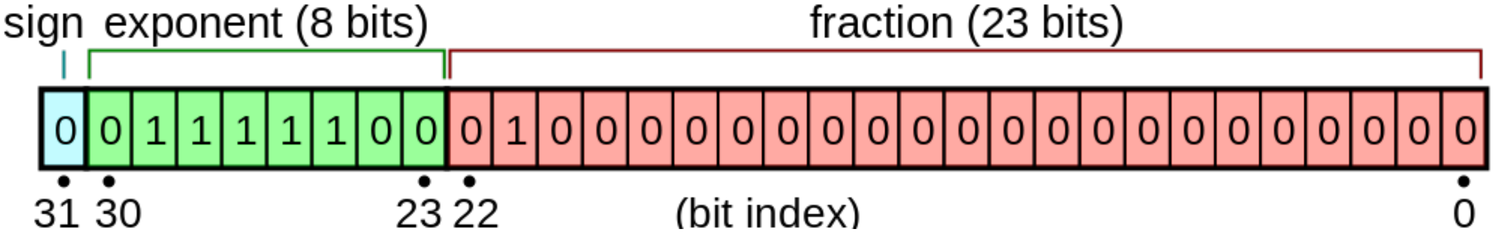


$$p = (-1)^s \times 2^{e-127} \times (1.m_1 m_2 \ldots m_{23})_2$$

$$= (-1)^s \times 2^{e-127} \times (1 + \sum_{i=1}^{23} m_i \times 2^{-i})$$

float32

- Standard IEEE format (float32)



$$p = (-1)^s \times 2^{e-127} \times (1.m_1 m_2 \ldots m_{23})_2$$

$$= (-1)^s \times 2^{e-127} \times (1 + \sum_{i=1}^{23} m_i \times 2^{-i})$$

$$(-1)^0 \times 2^{124-127} \times (1 + 1 \cdot 2^{-2}) = (1/8) \times (1 + (1/4)) = 0.15625$$

Python numeric types (built in)

https://docs.python.org/3/library/stdtypes.html#numeric-types-int-float-complex

Python numeric types

- int
  - No max/min size (Python is unusual in this way)
  - Bigger values -> more bits necessary
- float
  - Defaults 64 bits (double precision)
    - You can also use float32 given a certain framework (e.g., PyTorch, numpy, etc.)
  - Most pre-trained ML models use float32 for parameters
  - Min/max, Inf, -Inf, NaN have special bit combinations

Python numeric types (built in)

https://docs.python.org/3/library/stdtypes.html#numeric-types-int-float-complex

Python numeric types

- int
  - No max/min size (Python is unusual in this way)
  - Bigger values -> more bits necessary
- float
  - Defaults 64 bits (double precision)
    - You can also use float32 given a certain framework (e.g., PyTorch, numpy, etc.)
  - Most pre-trained ML models use float32 for parameters
  - Min/max, Inf, -Inf, NaN have special bit combinations
- complex

Other (commonly used) numeric types

- Common numeric types that (a) CPU can directly manipulate and (b) popular Python frameworks (e.g., PyTorch) support
    - ints: uint8, int8, int16, int32, int64
    - floats: float16, float32, float64
    - dtype (data type)

Demos …