

1. Yes, the Fractional part plays an imp. role in deciding the precision of a Floating-point number.

Ex: In single precision representation of a Floating-point number, out of 32 bits, 1 bit MSB is for sign representation, next 8 for exponent, last 23 for fractional part which is the precision of the number.

More bits for fractional part imply more precise.

To double the precision 52 bits are for fractional part in 64-bit representation.

2. Non-zero numbers whose adjusted exponents are less than minimum exponent E_{\min} are called Subnormal numbers.

When all the exponent bits are 0 and the leading hidden bit of the significand is 0, then the floating-point number is called a subnormal number. Thus, one logical representation of a subnormal number is

$(-1)^s \times 0.f \times 2^{-127}$ (all 0s for the exponent),

where f has at least one 1 (otherwise the number will be taken as 0). However, the standard uses -126 , i.e., bias $+1$ for the exponent rather than -127 which is the bias for some not so obvious reason, possibly because by using -126 instead of -127 , the gap between the largest subnormal number and the smallest normalized number is smaller

The interpretation of a subnormal a number is different. The content of the exponent part (e) is zero and the significand part (m) is non-zero. The value of a subnormal number is

$$(-1)^s \times 0.m \times 2^{-126}$$

The smallest magnitude of a normalized number in single precision is $\pm 0000\ 0001\ 000\ 0000\ 0000\ 0000\ 0000\ 0000$, whose value is

$$1.0 \times 2^{-126}.$$

The largest magnitude of a normalized number in single precision is $\pm 1111\ 1110\ 111\ 1111\ 1111\ 1111\ 1111\ 1111$, whose value is $1.99999988 \times 2^{127} \approx 3.403 \times 10^{38}$

The smallest magnitude of a subnormal number in single precision is $\pm 0000\ 0000\ 000\ 0000\ 0000\ 0000\ 0000\ 0001$, whose value is $2^{-126+(-23)} = 2^{-149}$.

The largest magnitude of a subnormal number in single precision is $\pm 0000\ 0000\ 111\ 1111\ 1111\ 1111\ 1111\ 1111$, whose value is $0.99999988 \times 2^{-126}$.

3. There are 5 methods

- Rounding to nearest even number
9.9 to 10, -9.9 to -10
- Rounding to nearest, away from zero
9.9 to 10, -9.9 to -10
- Rounding towards zero
9.9 to 9, -9.9 to -9
- Rounding towards minus infinity
9.9 to 9, -9.9 to -10
- Rounding towards plus infinity
9.9 to 10, -9.9 to -9