MT2017510 Assignment 3

1. Yes.

The fractional part has a very significant role on deciding the precision of a floating-point number. For example Lets take a single precision representation of floating point number. Out of the total 32 bits, 1 bit from MSB is reserved for representing the sign of the number. The next 8 bits are for deciding the exponent whereas the las 23 bits are for deciding the fractional part i.e the precision of the number being represented. The larger the number of bits used to represent this fractional part, the more precise is the representation. For example, Double Precision, which uses 64 bits to represent a floating-point number, uses 52 bits to represent a number.

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

For binary numbers, in single precision floating point, we generally represent a number as $(-1)^s * 1.m * (2)^{e-127}$ so, this number with the leading 1 in the mantissa i.e. 1.m is called a normalized number. The minimum being say, $(-1)^s * 1.m * (2)^{0-127}$.

So, a subnormal number is a number which are smaller than this which can be represented, but without a leading 1 in the mantissa.

These numbers are represented as $(-1)^s * 0.m * (2)^{e-127}$. Here, the precision decreases as the number of leading zeroes increase in the mantissa.

< Normal Numbers >| < Subnormals > | < Subnormals > | < Normal Numbers >

Number line representation

3. The first two methods defined by IEEE is to round to the nearest value whereas the remaining 3 methods are called direct rounding.

a. Round to nearest even number.

e.g. +6.7 is rounded to +6.0

-3.5 is also rounded to -4.0 as nearest even number is -4.0

b. Round to nearest, away from zero

   The number lying midway is rounded to nearest number (even or odd, doesn't matter), which is at more distance from 0.

   e.g. +4.6 is rounded to +5.0

   +5.5 is rounded to +6.0

c. Rounding towards zero

   e.g. +5.5 is rounded to +5.0

   +8.7 is rounded to +8.0

   -6.5 is rounded to -6.0

d. Rounding towards the plus infinity is called rounding up or ceiling or greatest integer.

   e.g. +3.5 is rounded to +4

   +7.5 is rounded to +8.0

e. Rounding towards also called rounding down or flooring

   e.g. +9.5 is rounded to +10.0

   +6.5 is rounded to +7.0