ARM Assignment-3
Alan Israel (IMT2015526)

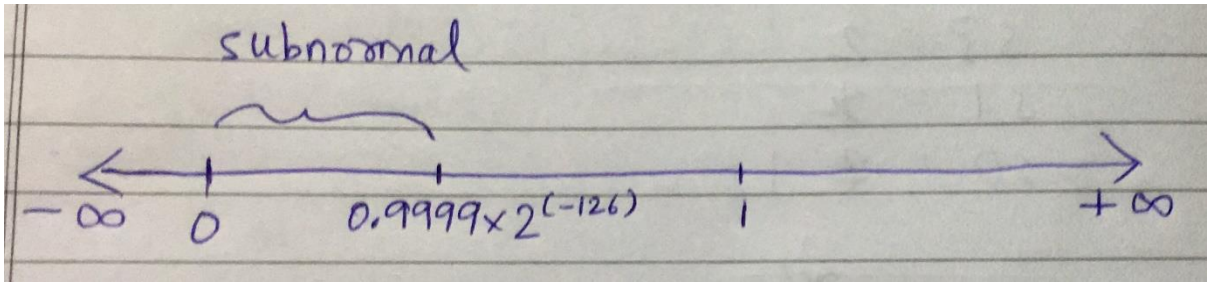
a) *Does any of the above three components play a role in the defining the Precession of the number? If so which are the component or Components which play the role in defining precession and how? Explain this with example in your own words*

Yes, the **fraction part** plays a role in defining the precision of the number. Because precision is defined as the number of digits to which a result is rounded. Hence, if we increase the number of bits in the fraction part, we get more precision.

For example, if we must represent the value of '*e*' which is **2.71828182846**, we can represent it using 0x402DF855 which is the closest representation using Cortex M4 series in Keil simulator.
The binary representation of the above value is,
01000000001011011111100001010100 - **2.7182817459106445**
But if we choose to represent it in 402DF000, which means in binary it looks like,
01000000001011011111000000000000 - **2.7177734375**
We can see that the precision is lost as we have ignored some of the bits in the end. Therefore, as we use more and more bits in the fraction part, the precision increases.

b) *What is Normal and Subnormal Values as per IEEE 754 standards explain this with the help of number line.*

A value is said to be **normal** if it is represented with a leading 1 bit. For example, numbers of the type:
-1^s * 1.f * 2^(e-127) where s is the sign bit, f the fraction part, e the exponent part. The leading binary digit is always 1.
Subnormal values do not use the entire precision available to normal number. Also, the leading binary digit is 0, which means the magnitude of the subnormal number is less than the format's smallest normal number. The subnormal value is of the type:
-1^s * 0.f * 2^(-127). The largest subnormal number is **0.999999988×2^-126**. It is close to the smallest normalized number 2^-126.



c) *IEEE 754vv defines standards for rounding floating points numbers to a represent able value. There are five methods defines by IEEE for this – Take time and understand what these five methods and explain it in your words using diagrams, illustrations of your own*.

The five methods for rounding floating point numbers are:

1) **roundTiesToEven** – Rounds to the nearest number. The result is the one with the even least significand digit.
   For example, a value of ***7.5*** rounds of to ***8***.
2) **roundTiesToAway** – Rounds to the nearest value away from zero than the value itself. It rounds to the nearest value above for positive numbers and nearest value below of negative numbers, thus taking it away from zero.
   For example, value of ***12.5*** rounds of to ***13***. And value of ***-12.5*** rounds of to ***-13***.
3) **roundTowardZero** – This is the opposite of roundTiesToAway as it rounds to the nearest value closer to zero than the value itself. It rounds to the nearest value below for positive numbers and nearest value above of negative numbers, thus taking it closer to zero.
   For example, value of ***12.5*** rounds to ***12*** and value of ***-12.5*** rounds to ***-12***.
4) **roundTowardPositive** – The value is rounded of to the nearest number towards positive infinity. For example, ***12.5*** is rounded to ***13*** and ***-12.5*** to ***-12***.
5) **roundTowardNegative** – The value is rounded of to the nearest number towards negative infinity. For example, ***12.5*** is rounded to ***12*** and ***-12.5*** to ***-13***.