180010011

CS 301

Assignment-1

1) a) Signed Binary Numbers use the MSB as a sign bit to display a range of either positive numbers (MSB = 0) or negative numbers (MSB = 1).

As, we know

- In two's complement form, a negative number is the 2's complement of its positive number with the subtraction of two numbers being $A - B = A + (\text{2's complement of } B)$.
- Also, 2's complement of an n bit binary unsigned number with MSB as '0' is $2^n - |u|$,
  where u is the n bit binary unsigned number.

2's complement is defined as follows:

$$F(u) = \begin{cases} u, & 0 \le u \le 2^{n-1} - 1 \text{ ; for +ve numbers} \\ 2^n - |u|, & -2^{n-1} \le u < 0 \text{ ; for -ve numbers.} \end{cases}$$

Eg:- -5 representation

$\Rightarrow 2^4 - |-5| \Rightarrow 16 - 5 \Rightarrow 11 \Rightarrow (1011)_2$

b) By using sign-magnitude method we can result in the possibility of two different bit patterns having the same binary value.

Eg:- As, we know there are two representations for '0'.
i.e 1000 for negative zero.
0000 for positive zero.

This cause a big complications.
Also, operations such as addition and subtraction becomes difficult.

So, we can use sign-magnitude method (or) 2's complement notation for representing signed binary numbers.

notes
Phone
email
website

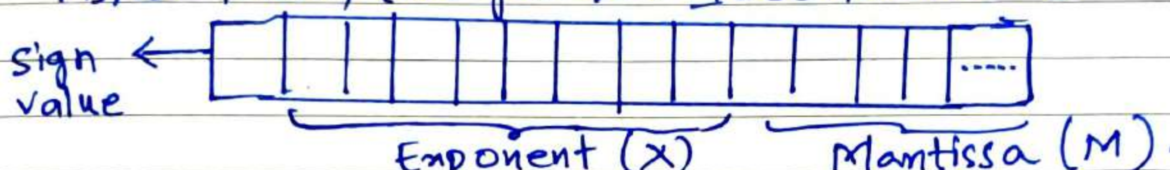
2) Represent −2.25 in the IEEE 754 format. Give your answer in hexadecimal format. Show all steps clearly.

sol) As we see the sign is negative, therefore our number will start with '1'.

Now, ignoring the sign, let's convert 2.25 to binary:
The binary code for 2.25 is 10.01

$\therefore\ 2.25 = 10.01$

As, we know, the general IEEE format is



As we know sign is -ve, therefore

sign = 1

To find Exponent (x) -

We know Normal floating point i.e

$$A = (-1)^{s} * (1.frac) \times 2^{(Exp-127)}$$

Mantissa = 10.01

$$\Rightarrow \underbrace{1.001}_{\text{Mantissa.}} \times 2^{1} \leftarrow E$$

Let's Equate

$$\Rightarrow (-1)^{s} \cdot (1.frac) \times \left(2^{(Exp-127)}\right) = (-1)^{s} \cdot (1.frac) \cdot 2^{1}$$

$$\Rightarrow Exp - 127 = 1$$

$$Exp = 127 + 1$$

$$\boxed{Exp = 128}$$

⇒ Binary form of 128 = 10000000.

⇒ lets subtitute in IEEE format we get

1100 0000 0001 0000 0000 0000 0000 0000 0000
12 0 1 0 -----

∴ Hexadecimal form for −2.25 is C0100000.

3) What are denormal numbers in IEEE 754 standard. What is the need for them?

Ans) IEEE 754 Standard uses denormal numbers to
- fill the gap between 'Zero' and the smallest normalized float.
- Smallest floating point we can have

  S=0 i.e signed bit

  E= 000000001 i.e Exponent

  M= 000...... 0 i.e Mantissa

  $F = 2^{(-126)}$.

  $G = F/2$ can clearly be not represented in our normal system of representing floating numbers.

  To represent them, we use denormal set of no.

  i.e $A = (-1)^S \times P \times 2^{(-126)}$

  where $P = 0+M$; $0 \le M < 1$.

  Smallest floating number : $2^{(-126)}$

  Largest denormal number : $(1-2^{(-23)}) \cdot 2^{(-126)}$.
- Provide gradual underflow to zero.
- Value of denormalized number (S, 0, F)

  Single precision : $(-1)^S \times (0.F)_2 \times 2^{-126}$

  double precision : $(-1)^S \times (0.F)_2 \times 2^{-1022}$