Ans. to the Q. NO - 01 (a)

Given, $\beta = 2$

$m = 4$

$-2 \le e \le 6 \Rightarrow e_{min} = -2$

$e_{max} = 6$

Standard Form: We know,

$$F = \pm (0 \cdot d_1 d_2 d_3 d_4 \cdots d_m)_\beta \times \beta^e$$

Maximum:

$$\frac{15}{16} \times (2)^6$$

$$= 60$$

(Ans:)

mantissa $= 4$

$\approx (0 \cdot 1111)_2$

$= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16}$

$= \frac{15}{16}$

exponent $= 2^6 = 64$

Minimum: $\Rightarrow \frac{1}{2} \times \frac{1}{4}$

$= 0 \cdot 125$

(Ans:)

smallest mantissa,

$m = (0 \cdot 1000)_2$

exponent, $(2)^{-2} = \frac{1}{4}$

And,

smallest (most negative) $= -\left(\frac{15}{16}\right) \times (2)^6$

$= -60$

(Ans:)

[P. T. O.]

CamScanner

Denormalize Form: We Know,

$$f = \pm (1 \cdot d_1 d_2 d_3 d_4 \cdots d_m)_\beta \times \beta^e$$

Maximum $= \frac{31}{16} \times (2)^6$

$= 124$

(Ans:)

mantissa

$= (1 \cdot 1111)_2$

$= 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16}$

$= \frac{31}{16}$

Exponent $= 2^6$

$= 64$

Minimum:

$\Rightarrow 1 \times \frac{1}{4}$

$= 0.25$

(Ans:)

mantissa,

$= (1 \cdot 0000)_2$

$= (1 + 0 + 0 + 0 + 0)_2$

$= 1$

exponent $= (2)^{-2}$

$= \frac{1}{4}$

And,

Smallest (most negative) $= -\left(\frac{31}{16}\right) \times (2)^6$

$= -124$

(Ans:)

Normalize Form:

We Know,

$$F = (\pm\, 0.1 d_1 d_2 d_3 \cdots d_m)$$

Maximum $= \left(\frac{15}{16}\right) \times (2)^6$

$= 60$

(Ans:)

mantissa,

$m = (0.1111)_2$

$= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{16}$

$= \frac{15}{16}$

exponent $= 6$

$\approx (2)^6$

Minimum $= \left(\frac{1}{2}\right) \times (2)^{-2}$

$= \left(\frac{1}{2}\right) \times \left(\frac{1}{4}\right)$

$\therefore = \frac{1}{8}$

$= 0.125$

(Ans:)

Mantissa,

$m = (0.1000)_2$

$= \frac{8}{16}$

$= \frac{1}{2}$

exponent $= (2)^{-2}$

Ans. to the Q. NO - 01 (b)

Determining non-negative minimum numbers:

① Standard Form:

Minimum (non-negative) $= (0.1000)_2 \times (2)^{-2}$

$= 0.125$ [from '(a)']

(Ans:)

② Denormalize Form:

Minimum (Non-negative) $= (1.0000)_2 \times (2)^{-2}$

$= 0.25$ [From '(a)']

(Ans:)

③ Normalize Form:

Minimum (non-negative) $= (0.1000)_2 \times (2)^{-2}$

$= 0.125$ [From '(a)']

(Ans:)

Ans. to the Q. NO- 01(c)

Standard Form :

i) sign $\Rightarrow$ 2 possibilities : + (ve)
- (ve)

ii) mantissa $\Rightarrow$ d_1 must be nonzero

(1 to 15 in binary)

$\Rightarrow$ 15 choice.

$\Rightarrow d_2, d_3, d_4 \ldots$

(0 or 1)

$\Rightarrow (2)^3 = 8$ choice

$\therefore$ total mantissa's combinations,

$15 \times 8 = 120$

iii) Exponent $\Rightarrow$ Given,

$-2 < e < 6$

$\Rightarrow -2, -1, 0, 1, 2, 3, 4, 5, 6$

$\Rightarrow$ (Total 9 values)

= P (mantissa × exponent × sign)

$\therefore$ Total $= (120 \times 9 \times 2)$

$= 2160$

(Ans:)

De normalize form : We Know,

$$F = \pm \left(1 \cdot d_1 d_2 d_3 d_4 \cdots\cdots d_m\right)_\beta \times \beta^e$$

According to the given values,

$\Rightarrow$ mantissa $\Rightarrow 1 \cdot d_1 d_2 d_3$ (bits)

[1 bit fixed with 1] $\Rightarrow (2)^3 = 8$

$\Rightarrow$ exponent $\Rightarrow -2 \leq e \leq 6$

$\Rightarrow -2, -1, 0, 1, 2, 3, 4, 5, 6$

(9 values)

$\Rightarrow$ sign $\Rightarrow \left.\begin{matrix} + (ve) \\ - (ve) \end{matrix}\right\} 2$ choice

Total $= 8 \times 9 \times 2 = 144$

(Ans:)

Normalize form:

We know,

$$F = \pm(0.1\, d_1 d_2 d_3 \cdots d_m)_\beta \times \beta^e$$

i) mantissa $\Rightarrow$ 0.1 -----

[fixed]

$\Rightarrow$ for d_1, d_2, d_3 ----- $\Rightarrow (2)^3 = 8$ choices

ii) exponent $\Rightarrow$ $-2 \leq e \leq 6$

$\Rightarrow -2, -1, 0, 1, 2, 3, 4, 5, 6$

(9 values)

iii) sign $\Rightarrow$ $\left.\begin{array}{l} +(\text{ve}) \\ -(\text{ve}) \end{array}\right\}$ 2 choice.

$\therefore$ Total $= 8 \times 9 \times 2 = 144$

(Ans:)

Ans. to the Q. NO - 02 (a)

Given,

real number $x = (5.625)_{10}$

Integer Part	Fraction part
$2 \mid 5$ $2 \mid 2 - 1$ $2 \mid 1 - 0$ $0 - 1$	$0.625 \times 2 = 1.25$ $0.625 \times 2 \Rightarrow$ Integer 1, fraction 0.25
$\therefore (5)_{10} = (101)_2$	$0.25 \times 2 \Rightarrow 0.5$ $\Rightarrow$ Integer 0, fraction 0.5 $0.5 \times 2 = 1.0 \Rightarrow$ Integer 1 fraction 0.0 $\therefore (0.625)_{10} = (.0 \cdot 101)_2$

$\therefore (5.625)_{10} = (101.101)_2$

(Ans:)

Ans. to the Q. NO- 02(b)

We know,

Denormalize form $= \pm (1.d_1 d_2 d_3 \cdots d_m)_\beta \times \beta^e$

Given,

$m = 3$

[only 3 fraction bits are stored]

Where,
β = base
e = exponent

from (a);

$$(5.625)_{10} = (101.101)_2$$

In Normalized form $\Rightarrow$ if $m = 3$;

$$= (1.01101)_2 \times 2^2$$

Denormalize form:

$$= \pm (1.d_1 d_2 d_3 \cdots d_m)_\beta \times \beta^e$$

$$= (1.011)_2 \times 2^2 \quad [\because m = 3]$$

here,
$d_1 = 0$
$d_2 = 1$
$d_3 = 1$

$\therefore$ the stored floating-point value,

$$fl(x) = (1.011)_2 \times 2^2$$

(Ans:)

Ans. to the Q. NO- 02(C)

i) converting $fl(x)$ back to decimal:

from '(b)': we get,

$$fl(x) = (1.011)_2 = \left(1 + 0/2 + 1/4 + 1/8\right) \times 4$$

$$= \left(1 + 0 + 0.25 + 0.125\right) \times 4$$

$$= 1.375 \times 4$$

$$\therefore fl(x) = (5.5)_{10}$$

(Ans:)

ii) Rounding Error $= \left| x - fl(x) \right|$

$$= \left| 5.625 - 5.5 \right|$$

$$= 0.125$$

(Ans:)

iii) Determining Machine Epsilon:

Given,

mantissa, $m=3$

Maximum Possible Total error,

$$= 0.125 \times (10)^{-3} \times (10)^{0}$$

[P.T.O.]

CamScanner

We Know,

Machine Epsilon, $\epsilon_M = \dfrac{0.125 \times (10)^{-3} \times (10)^{0}}{0.125 \times (10)^{-1} \times (10)^{0}}$

$= \dfrac{1}{8} \times (10)^{-2}$

$= 1.25 \times 10^{-3}$

$= 0.125$

(Ans:)

Answer to the Q. NO-03(a)

Given,

quadratic equation,

$$x^2 - 60x + 1 = 0 \quad \text{—— (1)}$$

Here,

for the given equation,

$a = 1$
$b = -60$
$c = 1$

We know,

for $ax^2 + bx + c = 0$,

solving formula,

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Now,

using $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

in equation (1),

here

$\sqrt{899}$
$= 29.9833287$

≈ 29.9833
(after rounding)

$$x_1 = 30 + \sqrt{899} = 59.9833$$

$$x_2 = 30 - \sqrt{899} = 0.01667$$

[computers generally calculate upto 4 sigbit value]

Now, while calculating x_2, the loss of significance occurs ⇒ in the subtraction ⇒ $|(0.0167) - (0.01677)|$

$= 2.81 \times 10^{-5}$

(result a small number)

Ans. to the Q. NO - 03(b)

Given equation,

$$x^2 - 60x + 1 = 0 \quad \text{--- (1)}$$

In '(a)',

we get,

loss of significance occurs.

Now,

Evaluation of correct roots without occurance of Loss of significance :

$$x_1 = \frac{60 + 59.9833287}{2} = 59.9833287$$

Now,

rounding to 6 significant figures :-

$$\boxed{x_1 \approx 59.9833}$$

Again, for equation (1) ⇒

$$\beta = \frac{1}{x_1} = x_2$$

$$\Rightarrow x_2 = \frac{1}{59.9833287}$$

$$= 0.0166713243$$

rounding to 6 significant figures :-

$$\boxed{x_2 \approx 0.0166713}$$

We know,

for $ax^2+bx+c=0$,

$$\alpha+\beta = -\frac{b}{a}$$

$$\alpha-\beta = \frac{c}{a}$$

fundamental property of a polynomial.

$\therefore$ The correct roots where no loss of significance occurs:

$$x_1 \approx 59.9833$$

$$x_2 \approx 0.0166713$$

(Ans:)