Question - 1:

Normalize the following numbers:

	Given Number	Normalized Number
i.	0.0000124678_{10}	1.24678×10^{-5}
ii.	$1584.234_{10} \times 10^5$	1.584234×10^{8}
iii.	4782.2354_{10}	4.7822354×10^{3}
iv.	110101.1111_2	1.101011111×2^{5}
v.	0.001100_2	1.100×2^{-3}
vi.	$1101.1111_2 \times 2^5$	$1.101\ 1111 \times 2^{8}$

Question - 2:

actual exp.

Find the Biased Exponent of $\underbrace{1.1011 \times 2^{34}}_{\text{normalized number}}$ in IEEE-754 $\underbrace{\text{single precision}}_{\text{Exp = 8bits}}$ format.

$$\text{Bias} = 2^{8-1} - 1 = 127$$

$$\text{Biased exponent} = 34 + 127$$

$$= 161$$

$$= 1010\ 0001 \quad \text{(Ans)}$$

Question - 3:

actual exponent

Find the Biased Exponent of $\underbrace{1.1011 \times 2^{4}}_{\text{normalized number}}$ in 12-bit IEEE-754 format where the size of the exponent field is 4 bits.

$\text{Bias} = 2^{4-1} - 1$

$= 7$

$\text{Biased exponent} = 4 + 7$

$= 11$

$= 1011$ (Ans)

Question - 4:

actual exp

Find the Biased Exponent of $\underbrace{1.1011 \times 2^{34}}_{\text{normalized number}}$ in 64-bit IEEE-754 format.

$\text{Bias} = 2^{11-1} - 1$

$= 1023$

$\text{Biased Exponent} = (34 + 1023) = 1057$

$= 100\ 0010\ 0001$ (Ans)

Question - 5:

Find the Biased Exponent of 5678.898 in 34-bit IEEE-754 format where the size of the exponent field is 10 bits.

5678.898

$= 1\ 0110\ 0010\ 1110 . 1110\ 0101\ 11 \ldots$

$= 1.0110\ 0010\ 1110\ 1110 \times 2^{12}$ actual exponent

$\text{Bias} = 2^{10-1} - 1 = 511$

$\text{Biased exponent} = 511 + 12 = 523$

$= 10\ 0000\ 1011$ (Ans)

Question - 6: sign bit = 1

Convert -0.00987$_{10}$ in 34-bit IEEE-754 floating point representation where the size of the fraction field is 23 bits.

sign bit	exponent	fraction
1	10	23

$$0.00987 = 0.0000\ 0010\ 10\ldots\ldots$$

$$= 1.\underbrace{010\ldots}_{\text{fraction}} \times 2^{-7}$$

actual exponent

$$\text{Bias} = 2^{10-1} - 1 = 511$$

Biased Exponent = 511 + (-7) = 504 = 01 1111 1000

Fraction = 0100 0000 0000 0000 0000 000

$$\underbrace{1\ 01}_{2}\ \underbrace{1111}_{F}\ \underbrace{1000}_{C}\ \underbrace{0100}_{2}\ \underbrace{0000}_{0}\ \underbrace{0000}_{0}\ \underbrace{0000}_{0}\ \underbrace{0000}_{0}\ \underbrace{000}_{0}$$

= 0x2FC200000 (Ans)

try Q7 - Q8 yourself

Question - 9:

Convert the following IEEE-754 single-precision floating point numbers into decimal.

	Given Numbers	Decimal Representation
i.	0xFF1205BA	$-1.94097021 \times 10^{38}$
ii.	3457890989_{10}	try rest yourself
iii.	23245613451_8	

(i) 0xFF1205BA

$= \underbrace{1}_{\text{Sign Bit}}\ \underbrace{111\ 1111\ 0}_{\text{Biased Exp.}}\underbrace{001\ 0010\ 0000\ 0101\ 1011\ 1010}_{\text{Fraction}}$

$\Rightarrow$ Bias $= 2^{8-1} - 1 = 127$

Biased exponent $= 1111\ 1110 = 254$

$\therefore$ Actual " $= 254 - 127 = 127$

$\Rightarrow$ Fraction $= 0.001\ 0010\ 0000\ 0101\ 1011\ 1010$

$= 0.1407\ 9976\ 0818\ 4814\ 4531$

$\Rightarrow$ Decimal value $= (-1)^1 \times (1 + 0.1407\ 9976\ 0818\ 4814\ 4531) \times 2^{127}$

$= -1.1407\ 9976\ 0818\ 4814\ 4531 \times 2^{127}$

$= -1.94097021 \times 10^{38}$ (Ans)

Question - 10:

Multiply the given numbers using IEEE-754 single-precision floating-point
representation. Check if the result has overflow or underflow. Show the result in decimal.

Note: Consider 10 decimal digits while converting from decimal to binary for the following questions.

	Given Numbers	Result	Overflow/ Underflow
i.	7.01_{10} and 0.71_{10}		
ii.	0.000101_2 and 10.1_2		
iii.	$0.000101_2 \times 2^{70}$ and $10010.000101_2 \times 2^{60}$		
iv.	1584.234_{10} and 1584.234_{10}		
v.	0.001100_2		
vi.	$1101.1111_2 \times 2^5$ and $110.000101_2 \times 2^6$		

(iii) $0.000101 \times 2^{70} \times 10010.000101 \times 2^{60}$

$\Rightarrow 0.000101 \times 2^{70}$

$= 1.01 \times 2^{70-4}$

$= 1.01 \times 2^{66}$

$\Rightarrow 10010.000101 \times 2^{60}$

$\Rightarrow 1.0010\ 0001\ 01 \times 2^{60+4}$

$= 1.0010\ 0001\ 01 \times 2^{64}$

$\Rightarrow 1.01 \times 2^{66} \times 1.0010\ 0001\ 01 \times 2^{64}$

$\Rightarrow (1.01 \times 1.0010\ 0001\ 01) \times 2^{130}$

$\Rightarrow 1.01101 \times 2^{130} \leftarrow$ actual exponent

$$\text{Bias} = 2^{8-1} - 1 = 127$$

$$\text{Biased Exponent} = 127 + 130 = 257$$

Range of Biased exponent $= 0$ to $2^8 - 1$

$= 0$ to 255

$= (1$ to $254)$ [0 & 255 reserved]

(upper range) 254 is smaller than biased exponent (257)

Hence, Overflow.

(iv) 1584.234×1584.234

$1584 = 110\ 0011\ 0000$

$.234 = .0011\ 1011\ 11\ldots$

$1584.234 = 110\ 0011\ 0000.0011\ 1011\ 11\ldots$

$= 1.10\ 0011\ 0000\ 0011\ 1011\ 11\ldots \times 2^{10}$

$\Rightarrow 1.10\ 0011\ 0000\ 0011\ 1011\ 11\ldots \times 2^{10} \times 1.10\ 0011\ 0000\ 0011\ 1011\ 11\ldots \times 2^{10}$

$= 10.01100 \times 2^{20}$

$= 1.001100 \times 2^{21}$

$$\text{Bias} = 2^{8-1} - 1 = 127$$

$$\text{Biased Exponent} = 127 + 21 = 148$$

Range of Biased exponent $= 0$ to $2^8 - 1$

$= 0$ to 255

$= (1$ to $254)$ [0 & 255 reserved]

$1 < 148 < 254$

Hence, No overflow / underflow

Question - 11:

Multiply the given numbers using IEEE-754 double-precision floating-point
representation. Check if the result has overflow or underflow.

Note: Consider 10 decimal digits while converting from decimal to binary for the following questions.

	Given Numbers	Result	Overflow/ Underflow
i.	7.01_{10} and 0.71_{10}	try rest yourself	
ii.	$0.000101_2 \times 2^{-850}$ and $10.1_2 \times 2^{-900}$		
iii.	$0.0101_2 \times 2^{790}$ and $10010.0101_2 \times 2^{680}$		

(iii) $0.0101 \times 2^{790} \times 10010.0101 \times 2^{680}$

$\Rightarrow 0.0101 \times 2^{790}$

$= 1.01 \times 2^{790-2}$

$= 1.01 \times 2^{788}$

$\Rightarrow 10010.0101 \times 2^{680}$

$= 1.0010\,0101 \times 2^{680+4}$

$= 1.0010\,0101 \times 2^{684}$

$\Rightarrow 1.01 \times 2^{788} \times 1.0010\,0101 \times 2^{684}$

$= (1.01 \times 1.0010\,0101) \times 2^{788+684}$

$= 1.01101 \times 2^{1472}$ ← actual exponent.

Bias $= 2^{11-1} - 1 = 1023$

Biased Exponent $= 1023 + 1472$

$= 2495$

Range of Biased exponent: 0 to $2^{11} - 1$

$= 0$ to 2048

$= (1$ to $2047)$ [0 & 2048 reserved]

$2495 > 2047$ (upper range)

Hence, overflow.

Question - 12:

Multiply the given numbers using 18bit IEEE-754 floating-point representation where the size of the fraction field is 12 bits. Check if the result has overflow or underflow.
Note: Consider 10 decimal digits while converting from decimal to binary for the following questions.

S-bit	exp.	frac.
1	5	12

	Given Numbers	Result	Overflow/ Underflow
i.	7.01_{10} and 0.71_{10}		
ii.	$0.000101_2 \times 2^{-85}$ and $10.1_2 \times 2^{-90}$		
iii.	$0.0101_2 \times 2^{79}$ and $10010.0101_2 \times 2^{68}$		

(ii) $0.000101 \times 2^{-85} \times 10.1 \times 2^{-90}$

$\Rightarrow 0.000101 \times 2^{-85}$
$= 1.01 \times 2^{-85-4}$
$= 1.01 \times 2^{-89}$

$\Rightarrow 10.1 \times 2^{-90}$
$= 1.01 \times 2^{-90+1}$
$= 1.01 \times 2^{-89}$

$\Rightarrow 1.01 \times 2^{-89} \times 1.01 \times 2^{-89}$
$= (1.01 \times 1.01) \times 2^{-89-89}$
$= 1.1001 \times 2^{-178}$ actual exponent

Bias $= 2^{5-1} - 1 = 15$

Biased Exponent $= 15 - 178$
$= -163$

Range of Biased exponent $= 0$ to $2^5 - 1$
$= 0$ to 31
$= (1$ to $30)$ [0 & 31 reserved]

$-163 < 1$ (lower range)

Hence, underflow.