MONORINA MUKHOPADHYAY
COMP 350 A# 1.

1. Number:

1111 1111 1111 1111 1111 1111 1010 1010

This is a negative number.

To get the 2's complement:
Flip all bits:

0000 0000 0000 0000 0000 0000 0101 0101

Add 1:

0000 0000 0000 0000 0000 0000 0101 0110

Convert to decimal: $2^6 + 2^4 + 2^2 + 2 = 86$

$\Rightarrow$ The number is -86.

2.

a) If the exponent width is 4 then the range of exponents is from -6 to 7. and the bias is 7.

b) The largest normalized floating pt. # is $(1.1111)_2 \times 2^7$

c) The largest subnormal floating pt # is $(0.1111)_2 \times 2^{-7}$

d) ϵ = gap b/w 1 and next smallest #
$= 1 \times 2^{-4}$

e) $a = -(10.01101)_2$
$= -(1.001101)_2 \times 2$
$= -(1.0011|\underbrace{01}_{\text{bits to be rounded}})_2 \times 2$

$\Rightarrow$ Round

- down $= -(1\cdot0011_2 + 0\cdot0001_2) \times 2$

$= -(1\cdot0100_2) \times 2.$

- Up: for -ve #s rounding up $= -(1\cdot0011)_2 \times 2$

- Round towards zero : $-(1\cdot0011)_2 \times 2$ (b/w 0 and a)

- Round to nearest: $-(1\cdot0011)_2 \times 2$ is closer to a

3.

a) In IEEE single format:

$1_{10} = 1\cdot000\ldots\ldots0_2 \times 2^0$

$2_{10} = 1\cdot000\ldots\ldots0_2 \times 2^1$

$3/2_{10} = 1\cdot1000\ldots\ldots0_2 \times 2^0$

Range: $1 \leq x < 2$

$1\cdot000\ldots0_2 \times 2^0 \leq x < 1\cdot000\ldots0_2 \times 2^1$

$\Rightarrow$ there are 2^{23} #s in this range.

$1 \leq x < 3/2 \Rightarrow 1\cdot000\ldots0_2 \times 2^0 \leq x < 1\cdot1000\ldots0_2 \times 2^0$

$\Rightarrow 2^{22}$ #s in this range

|||ly, $2^{23} - 2^{22} = 2^{22}$ #s in $3/2 \leq x < 2$ range.

b) $\frac{1}{2} = 1.00\,0 \ldots 0_2 \times 2^{-1}$

$1 : 1.00 \ldots 0_2 \times 2^0$

There are 2^{23} #s in this range.

$2/3$ is not a floating pt. # $\Rightarrow$ it has to be rounded.

$2/3_{10} = 0.10101010\ldots_2$

$= 1.010101\ldots_2 \times 2^{-1}$

Converting to IEEE single format:

$a_- \approx 1.010101 \ldots 10_2 \times 2^{-1}$

$\approx 0.666667_{10}$

The gap b/w a_- and the next nearest FPN is
$2^{-23} (=\varepsilon) \times 2^{-1} (=E) = 2^{-24}$

Between $2/3$ and $\frac{1}{2}$ theres are $(2/3 - \frac{1}{2}) \times \frac{1}{2^{-24}}$ #s $= \frac{1}{6} \times 2^{+24}$ #s.

$\Rightarrow$

Illy b/w $2/3 < x \leq 1 \Rightarrow 2^{23} - 2^{+24} \cdot \frac{1}{6} = \frac{2}{3}\, 2^{23}$

(3c. is later)

5.

$-\infty/0 = -\infty$ $\qquad 1^{-NaN} = NaN$

$\infty/-\infty = NaN$ $\qquad -0/NaN = NaN$

4. Let $x = 11.11_2$ and $y = 1.1_2$

$\Rightarrow x - y = 10.01_2$
$= 1.001_2 \times 2^1$

Now if we round up: round $(x-y) = x \ominus y = (1.00_2 + 0.01_2) \times$
$= 1.01_2 \times 2^1$
$= 10.1_2$

$y - x = -10.01_2$
$= -1.001_2 \times 2^1$

If we round up, $y \ominus x = \text{round}(y - x)$
$= \text{round}_{up}(-1.001_2 \times 2^1)$

$= -1.00_2 \times 2^1$
$= -10.0_2$

$-(y \ominus x) = 10.0_2 \neq x \ominus y$

$\Rightarrow$ Not true in general. Also in decimal, rounding up -1.25_{10} gives -1.2_{10} and rounding up 1.25_{10} gives 1.3 so this is obviously untrue.

3c. $1 \leq x_1, x_2 \leq 2$

Pick $x_1 \in [3/2, 2)$

There are from (a) 2^{22} FPNs in $[3/2, 2)$. $\frac{1}{x} > 1/2$ and $\frac{1}{x} \leq 2/3$

$3/2 \leq x < 2$

$\Rightarrow 1/2 < 1/x \leq 2/3 \Rightarrow \Rightarrow 1/2 < 1 \oslash x \leq a$ where

a is smallest FPN larger than $2/3$. From (b), there are $1/6\ 2^{24}$ $= 2/3 \cdot 2^{22}$ #s b/w $1/2$ and $a < 2^{22}$. Since there are less numbers in $1 \oslash x \in (1/2, 2/3]$ than there are x's in $[3/2, 2) \Rightarrow$ there must be two #s x_1 & x_2 in $[3/2, 2)$ s.t. $1 \oslash x_1 = 1 \oslash x_2$.