Name: ________________________ CSCI 237

Practice Problems :: Floating Point

For all questions, please show your work. Clearly indicate your final answers.

1. **IEEE 754 FP:** For this question, please consider the following **32-bit floating point** representation of a number **`x`**:

   `x` `= 0100 0001 0100 1000 0000 0000 0000 0000`

   Notes: For a 32-bit floating point number, the ***exp*** *field has 8 bits. Therefore, the bias in a* ***normalized*** *representation is* 2^{8}*-1=127.*

   What is the decimal value of **`x`**?

2^0=1, 2^1=2, 2^2=4, 2^3=8, 2^4=16, 2^5=32, 2^6=64, 2^7=128, 2^8=256, 2^9=512, 2^{10}=1024, 2^{11}=2048, 2^{12}=4096, 2^{13}=8192, 2^{14}=16384, 2^{15}=32678, ...

2. For this question, consider all *non-negative* numbers representable using the IEEE 754 Floating point representation from the previous question (32-bits with a 1-bit `sign` field followed by an 8-bit `exp` field, followed by a 23-bit `frac` field).

    a. What is the bit representation of the **smallest** *normalized* number?

       What is the value of that number expressed in the form $v = (-1)^s \times M \times 2^E$?

       S =

       M = 1.

       E =

       v =

    b. What is the bit representation of the **largest** *denormalized* number?

       What is the value of that number expressed in the form $v = (-1)^s \times M \times 2^E$?

       S =

       M = 0.

       E =

       v =