

# Exercises 1

1. a) The largest number that can be represented by a floating point system (2, 5, -10, 10) is  $0.11111 \times 2^{10}$  which is equal to 992 (base 10). The largest value of  $n$  where  $n! \leq 992$  is 6 (since  $6! = 720$ ). Hence  $n=6$ .

- b) 7 has exponent 3 in the base-2 system. Hence we have

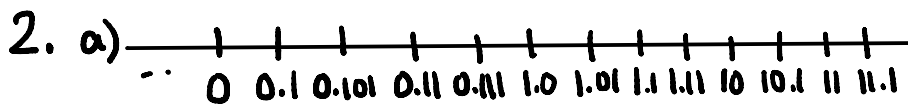
$$2^{-12} = 2^{3-t} \rightarrow t = 15$$

70 has exponent 7. Hence its distance from the next largest floating point number is

$$2^{7-15} = 2^{-8}$$

- c)  $8$  (base 10)  $= 0.1 \times 2^4$ . Since 8 is a number in this floating point system and we have  $x < 8 < y$ , to have the smallest  $y-x$  we must have that  $x$ , 8, and  $y$  are consecutive numbers. The distance from  $x$  to 8 is  $2^{4-t}$ . The distance from  $y$  to 8 is  $2^{4-t}$ . Hence we have that the smallest  $y-x$  is

$$2^{4-t} - 2^{3-t} = 2^{3-t}$$



Negative numbers are a mirror image of the positive ones (didn't want to draw them lol)  
 \*\*Not to scale

- b) There are 25 elements in  $S$ .  
 OFL = largest positive number = 11.1  
 UFL = smallest positive number = 0.1

$$|f(x) - x| \leq 0.1$$

$$|x| \geq 0.1$$

$$\frac{|f(x) - x|}{|x|} \leq 1 \rightarrow \text{machine epsilon}$$

3.  $1.5 \times 10^8 = 0.100011110000110100011 \times 2^{28}$   
 The length on the last bit of this mantissa represents  $2^8$ .

$$75 \text{ (base 10)} = 0.1001011 \times 2^7$$

The length on the last bit of this mantissa represents 2.