

计算机系统概论（2024 秋）作业 1

1. 在所有由五个“1”和三个“0”组成的 8 位二进制整数（补码形式）中，最小的数是_____，最大的数是_____（答案用二进制表示）。
2. 已知 $[X]_{\text{补}} = 0x2024$ ， $[Y]_{\text{补}} = 0xAE77$ ，则 $[X+Y]_{\text{补}} = \underline{\hspace{2cm}}$ ， $[X-Y]_{\text{补}} = \underline{\hspace{2cm}}$ （X、Y 的数据位宽均为 16 位，计算结果用 16 进制的补码表示）。
3. BF16 是一种 16 位浮点数类型（符合 IEEE 浮点数标准），常见于大模型混合精度训练。BF16 的 exp 位数是 8，frac 位数是 7，符号位数是 1，其所能表示的最大的非规格化数的 exp 是___，frac 是 ___。2024（十进制数）的 exp 是 ___，frac 是 ___（请用 0、1 位串表示答案）。
4. 假设存在一种 9 位浮点数（符合 IEEE 浮点数标准），符号位数是 1，exp 位数是 4，frac 位数是 4。其数值被表示为 $V = (-1)^S \times M \times 2^E$ 形式。请在下表中填空。

描述	Binary	M	E	Value
5.0				5.0
最小的大于 0 的浮点数				

5. FP16 是另一种 16 位浮点数（符合 IEEE 浮点数标准），也常见于大模型混合精度训练。FP16 的 exp 位数是 5，frac 位数是 10，符号位数是 1。某同学对该格式的一个数 x 执行了（整数的）按位右移操作，得到了 80.5。若右移操作按有符号数执行（算术右移），原来的数可能是_____，若右移操作按无符号数执行（逻辑右移），原来的数可能是_____（列出所有情况或填入“不存在”，数可以用小数或分数来表示，必须精确）。
6. 给定相同的字长（例如 16 位），
 - (a) 能表示的定点数个数多还是浮点数个数多？为什么？
 - (b) BF16 格式表示的浮点数个数多还是 FP16 格式表示的浮点数个数多？在模型训练中如果遇到梯度爆炸（模型参数数值大小趋向 ∞ ）或者梯度消失（模型参数数值大小趋向 0），希望通过改变浮点数格式来缓解数值问题时，应该选用 BF16 还是 FP16？

7. 使用不超过 4 次位运算或加减运算完成整数运算 $y = x \times 85$ （允许引入临时变量，不需要考虑溢出的情况）。