刘程华　　2018011687　　计91

（1）解释为什么“允许的最大和最小双精度浮点数分别为±1.8×10^308和±2.2×10^(-308)?

假设浮点数系统F采用β进制，表示形式为$\pm\left(d_0+\dfrac{d_1}{\beta}+\dfrac{d_2}{\beta^2}+\ldots+\dfrac{d_{p-1}}{\beta^{p-1}}\right)\times\beta^E \quad d_0\neq 0,$

此题中$\beta=2$双精度浮点数系统中$p=53$（有一个隐藏位），有11个bit用来表示指数，1个bit用来表示正负。E上限值为$U=2^{10}-1=1023$，下限值为$L=-(2^{10}-2)=-1022$（剩余三个取值用来处理特殊情况0，NaN，Inf）

所以表示的最大浮点数为d_i取1，E取U，即$\pm(2-\dfrac{1}{2^{52}})\times 2^{1023}\approx\pm 1.8\times 10^{308}$

表示的最小浮点数为d_0取1其余d_i取0，E取L，即$\pm 1\times 2^{-1022}\approx\pm 2.2\times 10^{-308}$

（2）最小的非规格化数字是1.4e-45，为什么?

$p=24$，表示的数是非规格化形式时，指数段全为0，此时最小的数字为$d_{p-1}=1$，其余d_i取0，即为$\dfrac{1}{2^{23}}\times 2^{-126}\approx 1.4\times 10^{-45}$

（3）证明浮点符号的关键概念，即“相邻字符之间的间隙数字随数字的大小而缩放”。

由（1）中浮点数表示形式可知，相邻的浮点数之间距离是变化的，离0越远相邻浮点数之间距离越大。值得一提的是，相邻的2的整数次幂之间，浮点数均匀分布。

（4）查找真实的二进制单精度（32位）浮点表示形式数字“ -9.625”，然后尝试输入“ ID.ID”。 报告您的发现。

$-9.625=-(1\times 2^3+0\times 2^2+0\times 2^1+1\times 2^0+1\times 2^{-1}+0\times 2^{-2}+1\times 2^{-3})=-1.001101\times 2^3$

指数段$127+3=130$，所以最终结果为：

$1\ 10000010\ 00110100000000000000000$

2018011687转换为二进制1111000010010000110101000100111

0.2018011687转换为二进制0.0011001110101001001111011100101111011100011011111010101

ID.ID转换为二进制为

1111000010010000110101000100111.0011001110101001001111011100011011111010101

$1.11100001001000011010100\times 2^{30}$指数段为$127+30=157$，最终结果为：

$0\ 10011101\ 11100001001000011010100$

再次转为浮点数我们发现这与原始输入值相差较大，单精度浮点数表示数字能力有限，在转换为单精度二进制过程中截断了很多位，小数部分信息甚至全部被截断。一般来说表示越大的数字，差值越大。