

# COMPUTACIÓN CIENTÍFICA

## PRÁCTICA 1

INSTITUTO DE EDUCACIÓN



UNIVERSIDAD  
NACIONAL DE  
HURLINGHAM

- 1) Sea  $m = 112$ . Escribir todos los números en coma flotante con una precisión  $t = 4$  y un exponente  $e \in \{-1, 0, 1, 2, 3\}$  que pueden generarse con dicho  $m$ .
- 2) Escribir los siguientes números en la forma normalizada  $0.d_1d_2\ldots \times 10^e$ :

- |          |          |
|----------|----------|
| a) 101.2 | b) 355   |
| c) 122.4 | d) 0.901 |

Reescribirlos como  $\pm m \times 10^{e-t}$

- 3) Escribir los siguientes números en el estándar IEEE754 de 32 bits y 64 bits:
- |                                |                         |
|--------------------------------|-------------------------|
| a) Los números del ejercicio 2 | b) 1.001                |
| c) 10100                       | d) 0.001                |
| e) 1                           | f) +0                   |
| g) -0                          | h) $\infty$ y $-\infty$ |
- 4) Convertir los siguientes números IEEE 754 a decimal. Cuando haya puntos suspensivos, indican que se completa con ceros.
- a) 0 10000011 010000000000000000000000
- b) 1 01111111 000000000000000000000000
- c) 1 00110100001 1010... 00101 (64 bits)
- d) 0 00000100001 1010010... 0010 (64 bits)

Investigue qué representa una codificación con todos unos en el exponente y mantisa distinta de cero.

- 5) Sea  $\beta$  la base en la que trabaja una máquina imaginaria y llamemos  $F$  a al conjunto de todos los números de máquina. Pruebe que la distancia entre 1 y el siguiente número de máquina es  $\varepsilon_M = \beta^{1-t}$ . Esta cantidad se llama *épsilon de máquina*. Cuando no haya ambigüedad escribiremos simplemente  $\varepsilon$ . Describa a  $\varepsilon$  para el estándar de precisión simple y doble. (Recuerde que si queremos ser consistentes con la notación de punto flotante normalizada para cualquier base, en precisión simple tenemos  $t = 24$  y en precisión doble  $t = 53$ , sin embargo, sabemos que al pasar al estándar IEEE754, se optimiza no guardar el bit que vale 1, por lo que allí, la longitud de la mantisa es 23 y 52 respectivamente.)
- 6) (Continuación) En la bibliografía se define a  $\varepsilon$  como el mínimo elemento  $y$  de  $F$  tal que  $1 + y \neq 0$ . Explique por qué tiene sentido esta definición.
- 7) Hallar el error (absoluto) y el error relativo en los siguientes casos:
- |  |                                       |
|--|---------------------------------------|
| a) $x = 3.141592$ y $\hat{x} = 3.14$     | b) $y = 1000000$ y $\hat{y} = 999996$ |
| c) $z = 0.000012$ y $\hat{z} = 0.000009$ |                                       |

Determine para cada caso, si se está en presencia de una buena aproximación o no (Sugerencia: interprete el error relativo como porcentaje de  $x$ ). Considere ahora una máquina que maneja la cantidad de dígitos significativos de las entradas. ¿Se

cumplen las cotas para los errores relativos? ¿Qué sucede si se utiliza aritmética de dos dígitos?

- 8) En una máquina de cinco decimales que redondea correctamente los números al número de máquina más cercano, ¿qué números reales  $x$  tendrán la propiedad  $\text{fl}(1 + x) = 1$ ?
- 9) Supongamos que estamos utilizando una máquina que maneja los números de la forma  $x = 0.1d_2d_3d_4 \times 2^e$ , significando de 4 dígitos). Calcule el error relativo al realizar los siguientes cálculos:
- |                                |                                     |
|--------------------------------|-------------------------------------|
| a) $\frac{1}{3} + \frac{1}{5}$ | b) $\frac{1}{3} \times \frac{1}{5}$ |
| c) $\frac{1}{5} - \frac{1}{3}$ | d) $\frac{1}{8} + 4$                |
| e) $(\frac{1}{8} + 4) - 4$     | f) $\frac{1}{8} + (4 - 4)$          |
- 10) Realizar las siguientes operaciones en Python y comparar el resultado esperado con el obtenido.
- |  |  |
|--|--|
| a) $1 + \frac{\varepsilon}{2}$                                 | b) $(1 + \frac{\varepsilon}{2}) + \frac{\varepsilon}{2}$       |
| c) $1 + (\frac{\varepsilon}{2} + \frac{\varepsilon}{2}) - 1$   | d) $((1 + \frac{\varepsilon}{2}) + \frac{\varepsilon}{2}) - 1$ |
| e) $1 + (\frac{\varepsilon}{2} + (\frac{\varepsilon}{2} - 1))$ |  |