Floating-Point Numbers

Due at 11:59pm ET on 23 January, 2022

What you need to get

- `YOU_a1q1.ipynb`: a jupyter notebook for Q1
- `YOU_a1q2.ipynb`: a jupyter notebook for Q2
- `YOU_a1q3.ipynb`: a jupyter notebook for Q3
- `YOU_a1q4.ipynb`: a jupyter notebook for Q4
- `YOU_a1q5.ipynb`: a jupyter notebook for Q5

What you need to know

The notebook has a function called `dec2fp` that takes a numerical value as input and generates a binary floating-point representation of it. The inputs t, L and U specify a floating-point number system (FPNS), which we will denote $\mathcal{F}(2, t, L, U)$, containing elements

$$b = \pm 0\,.\,d_1\,d_2\,d_3\ \ldots\ d_t\ \times\ 2^p\ ,$$

where $d_k \in \{0, 1\}$, $d_1 \neq 0$, and $p \in \mathbb{Z}$ with $L \leq p \leq U$. If a value falls outside the range of values in the FPNS, then it returns an exception: `Inf`, `-Inf`, `NaN`, or 0 (for underflow). The value of zero is a special code in which the mantissa is all zeros and the exponent is zero.

The floating-point numbers will be stored as strings. For example,

- 0.1101×2^{-3} will be represented by the string '`+0.1101b-3`'
- -0.100010×2^4 will be represented by the string '`-0.100010b4`'.

Note that the first character is always either a '+' or '-'. The number after the 'b' is the exponent for the base (the base is 2), although the exponent itself is represented in base-10. For example,

```
b = '+0.11100b4'
```

represents the number 0.11100×2^4, which has a value of 14. Hence,

```
b2 = dec2fp(14, 7, -20, 20)
```

returns the string '`+0.1110000b4`'. Type "`? dec2fp`" for more information.

You can perform arithmetic operations involving these binary strings using the function `fpMath` (also supplied in the notebook). The function takes two binary strings, a function, and t, L, and U. The output is another binary string. Note that functions in Python can be defined inline using the `lambda` notation. For example, the Python code

```
(lambda z1,z1: z1-z2)
```

returns a function that subtracts its second argument from its first argument. Thus, the call

```
fpMath(b1, b2, (lambda z1,z2: z1-z2), 3, -10, 10)
```

returns the binary code for the number that corresponds to `b1-b2`. Type "`? fpMath`" for more information.

What to do

1. [20 marks] Complete the Python function `randfp` in the `YOU_a1q1` notebook so that it randomly generates normalized binary floating-point numbers from the number system $\mathcal{F}(2, t, L, U)$. Your function should work for values of t up to 52, and $-1022 \leq L < U \leq 1023$. You can read the function's documentation for more information (type "`? randfp`").

   Hint:
   To append strings in Python, simply 'add' strings. For example,

   ```
   b = 'hi' + ' there ' + str(15)
   ```

   will construct the string '`hi there 15`'.

2. [20 marks] Complete the function `fp2dec` (in the `YOU_a1q2` notebook) so that it converts binary floating-point numbers in $\mathcal{F}$ to their decimal equivalents. An incomplete version of the function is supplied as starter code. Its input is a string representing a binary floating-point number (as described in *What you need to know* above). It is sufficient to output an IEEE double-precision number as the decimal value.

   Hints:
   For this question, you might find the Python functions `find`, and `int` useful. Also, you can extract substrings using indexing. For example, if `b='+0.1001b3'`, then `b[2]` will return the string `'.'`, and `b[6:]` will return `'1b3'`. Furthermore, the Boolean expression `b[3]=='1'` would return a value of `True`. You **cannot**, however, use any other function that does the conversion for you. You must implement it yourself based on first principles.

3. [20 marks] Consider the normalized floating-point number system $\mathcal{F}(\beta = 6, t = 6, L = -6, U = 6)$, with elements of the form
$$\pm 0 \,.\, d_1\, d_2\, d_3\, d_4\, d_5\, d_6 \times 6^p$$
   where $-6 \leq p \leq 6$. The number system is normalized, so $d_1 \neq 0$. The only exception is the zero element, in which all the mantissa digits are zero.

   (a) What is the largest value in $\mathcal{F}$?

   (b) What is the value of $\frac{0.5453334 5_6}{100_6}$ using this number system.

   (c) What is machine epsilon for $\mathcal{F}$? Express your answer in normalized base-6 format.

   (d) What fraction of the normalized numbers in $\mathcal{F}$ are smaller in magnitude than 1?

   Put your answers in the notebook `YOU_a1q3`.

4. [20 marks] Let $\mathcal{F}$ be a floating-point number system with machine epsilon E, and suppose that a, b and c are all elements of $\mathcal{F}$. Show that the relative error for the expression $ab - c$ has the upper bound
$$\frac{|(a \otimes b) \ominus c - (ab - c)|}{|ab - c|} \leq \frac{|ab|}{|ab - c|} E(1 + E) + E\,.$$
   Justify each inequality that you introduce. Put your solution in the notebook `YOU_a1q4`.

5. [20 marks] During a routine audit of First National Bank, auditors noticed that the accounts owned by the bank appeared to be missing a significant amount of money. Alarmed by this, the manager of the bank has alerted the police to investigate. You, as a forensic specialist, have been assigned to look through the software the bank uses to process its credit and debit transactions. Mathematically, the bank's net income is
$$\text{Net Income} = \sum_i \text{Credit}_i + \sum_i \text{Debit}_i\,.$$

The jupyter notebook `YOU_a1q5` loads 10,000 credit transactions and 10,000 debit transactions from the bank's database (using the function `ReceiveTransactions`), and calls the function `CalculateNet` to add up the credits and debits to arrive at a net income.

The auditors have asked you to investigate the function `CalculateNet` closely. In that function, you will see that there are three methods for calculating the net income, labelled A, B, and C. Write a short police report (a few sentences) that answers the following questions:

(a) Which method is the most accurate?

(b) Why is that method more accurate than the others. Justify your claim in (a).

(c) In your opinion as a forensic specialist, what does the function `CalculateNet` accomplish. Is a crime being committed?

What to submit

Rename each of your jupyter notebooks, replacing "`YOU`" with your WatIAM ID. For example, I would rename `YOU_a1q1.ipynb` to `kfountou_a1q1.ipynb`. Export each jupyter notebook as a PDF, and submit each PDF to Crowdmark. If you want, you can typeset your solutions to Q3 and Q4 in a LaTeX or Word document, or write electronically (as on a tablet), and hand in your document as a PDF. **Photographs or scans of handwritten solutions should be legible, otherwise, the TAs might deduct marks.**

Finally, upload your Python notebooks on Learn dropbox <—— Do not forget this, otherwise I will apply 10% penalty to the assignment.