Data Mining using Python
— exercises for introduction

Finn Årup Nielsen

DTU Compute
Technical University of Denmark

September 8, 2014

For loops, str and int

Write a function, `is_hashad` that determines whether a number is a *Harshad number* (for number base 10).

A Harshad number "is an integer that is divisible by the sum of its digits" (Wikipedia)

Example: $81 \rightarrow 8 + 1 = 9 \rightarrow 81/9 = 9 \rightarrow$ Harshad!

```
>>> is_hashad(81)
True
```

Hint: convert the number to a string.

Testing

Write a test function for `is_hashad` called `test_is_hashad` and run it.

Dictionaries

Count the number of items in a list with the result in a dictionary.

List example:

```
l = ['a', 'b', 'f', 'f', 'b', 'b']
```

Should give something like:

```
c = {'a': 1, 'b': 3, 'f': 2}
```

What and where is `defaultdict`?

Recursion

Implement a factorial function, $n!$, with recursion:

```
>>> factorial(4)
24
```

$(4! = 1 \times 2 \times 3 \times 4 = 24)$

See what happens with `factorial(1000)`

Classes

Construct a module with a derived dictionary class with sorted keys:

```
>>> s = SortedKeysDict({'a': 1, 'c': 2, 'b': 3, 'd': 4})
>>> s.keys()
['a', 'b', 'c', 'd']
>>> s.items()
[('a', 1), ('b', 3), ('c', 2), ('d', 4)]
```

Also implement `doctest` for the class.

Document it, test the documentation with `pep257` and extract the document with, e.g., `pydoc`. Write testing code and test it with py.test or nose.

File reading and simple computing

Consider a file with the following matrix $\mathbf{X}$:

```
1 2
3 4
```

Read and compute $\mathbf{Y} = 2 * \mathbf{X}$

Try also using the with statement in this case.

Project Euler

Project Euler is a website with mathematical problems that should/could be solved by computers.

Go to the Web-site http://projecteuler.net/ and solve some of the problems using Python.

As an example the problem number 16 can be solved in one line of Python:

```
>>> sum(map(int, list(str(2**1000))))
1366
```

Encoding

UTF-8 encoding/UNICODE

In terms of UTF-8/UNICODE what is wrong with the following code:

https://gist.githubusercontent.com/fnielsen/4183541/raw/afinn.py

Hint look at the word "naïve".

Make a correction.

See also:

http://finnaarupnielsen.wordpress.com/2011/06/20/simplest-sentiment-analysis-in-python-with-af/

UTF-8 encoding/UNICODE

Translate the AFINN sentiment word list with a language translation web service, — or perhaps just a part it — to a language you know and see if it works with with a couple of sentences.