An introduction to data science with

Håkon Hapnes Strand

Contents

[0. The How, Why and What of Python 1](#_Toc501282552)

[1. Anaconda distribution 1](#_Toc501282553)

[2. Jupyter notebook 1](#_Toc501282554)

[3. Basic Python 1](#_Toc501282555)

[4. Core data science libraries 2](#_Toc501282556)

[5. Machine learning 2](#_Toc501282557)

[6. Natural Language Processing (NLP) 2](#_Toc501282558)

# The How, Why and What of Python

import this

# Python download and setup

conda list

conda install nb\_conda

conda create -n py27 python=2.7 ipykernel

activate py26

Python 2 vs. 3

# Jupyter notebook

A convenient way to prototype and share Python code.

Basic keyboard shortcuts

Markdown

Line/cell magic

%time

%%time

%%timeit

%%javascript

# Basic Python syntax

2+2

print()

for, if, while

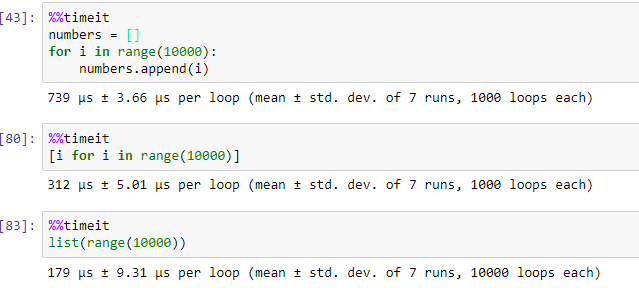
import this

The list data structure with different data types

Tuple

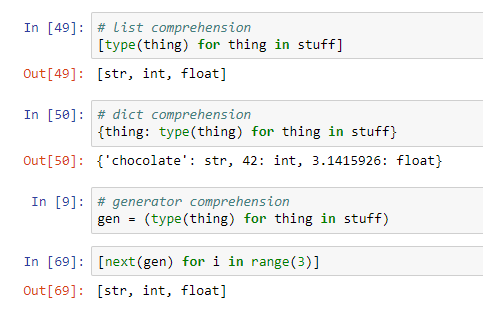
Dictionary

List comprehension



Dict comprehension

Generators (comprehension)



Basic add function

# Core data science libraries

Basics about importing libraries

* Numpy
  + Add two lists, add two np arrays (element-wise)
  + Shapes and linear algebra
  + Dot product
* Pandas
  + Read csv
  + pd.dummies
  + pd.concat
  + mean, max, min
  + df[df.Age>50]
  + df.loc
* matplotlib/seaborn
  + %matplotlib inline

# Machine learning

(scikit-learn example)

(talk about TensorFlow/Keras, CNTK, PyTorch)

# Natural Language Processing (NLP)

conda install gensim

conda install nltk

word2vec