

BMEG3105: Data analytics for personalized genomics and precision medicine – Lecture 2

Topic: Data & Python

Lecturer: Prof. LI Yu

Date: 9 Sep 2022

Content of Lecture 2

- Different data types
- Introduction to Python programming

Different data types

1. Sequential Data

- ✓ DNA, RNA

2. Data matrix

- Data that consists of a collection of records, each of which consists of a fixed set of attributes
- Data set can be represented by an n by m matrix, where there are n rows, one for each object, and m columns, one for each attribute

* shuffle the entire column or the entire row at one time

➡ would not change the data

* shuffle the entire column and the entire row at the same time

➡ might change the data

- ✓ 4x2 Matrix (4people, height & weight)

3. Spatial data

- Geographic locations and spatial information involved
- ✓ Map, Photographs

4. Temporal data

- With built-in support for handling data involving time
- ✓ Stock, ECG signal

5. Graph or networks

- Objects and connections
- ✓ Social network, PPI network

6. Text

- ▶ Short sentences
- ▶ Long sentences or documents
- ✓ Dictionary, Wikipedia text

7. Multi-modality data

- ✓ Video: temporal images, audio, transcript
- ✓ Electronic health records: data matrix, images, text
- ✓ Spatial transcriptomics: spatial data, sequence, data matrix

8. Unknown data type

- ▶ The data not mentioned above
- Diet & Exercise

Introduction to Python programming:

1. Definition of Programming

- Wikipedia ▶ “Computer programming is the process of designing and building an executable computer program to accomplish a specific computing result or to perform a specific task.”
- By understanding
Programming ▶ communicating with the computer (your friend), begging/asking him/her to do something for you

2. Definition of Python

- Wikipedia ▶ “Python is an interpreted high-level general-purpose programming language.”
- By understanding
Code ▶ the WeChat/WhatsApp message you send to the computer (your friend)
Python ▶ the WeChat/WhatsApp software

3. Python Programming -- computer language

- To communicate with the computer, asking it to do something for you
 - ▶ Calculate mean, variance, distance, gradient...
 - ▶ For lots of values!

4. Numpy in Python

- Additional plug-in for calculation
Function ▶ mean, variance, median, max, min...

Important syntax ► `import numpy`