STATS 507
Data Analysis in Python

Week5-2: Test and Debugging

Dr. Xian Zhang

Adapted from slides MIT 6.100L

Recall Scope of this class

Part 1: Introduction to Python

Data types, functions, classes, objects, testing and debugging

Part 2: Numerical Computing and Data Visualization

numpy, scipy, matplotlib, scikit-learn, Seaborn

Part 3: Dealing with structured data

pandas, regular expressions, SQL, real datasets

Part4: Intro to Deep Learning

PyTorch, Perceptron, Multi-layer perceptron, SGD, regularization, ConvNets

Programming in Python so far…

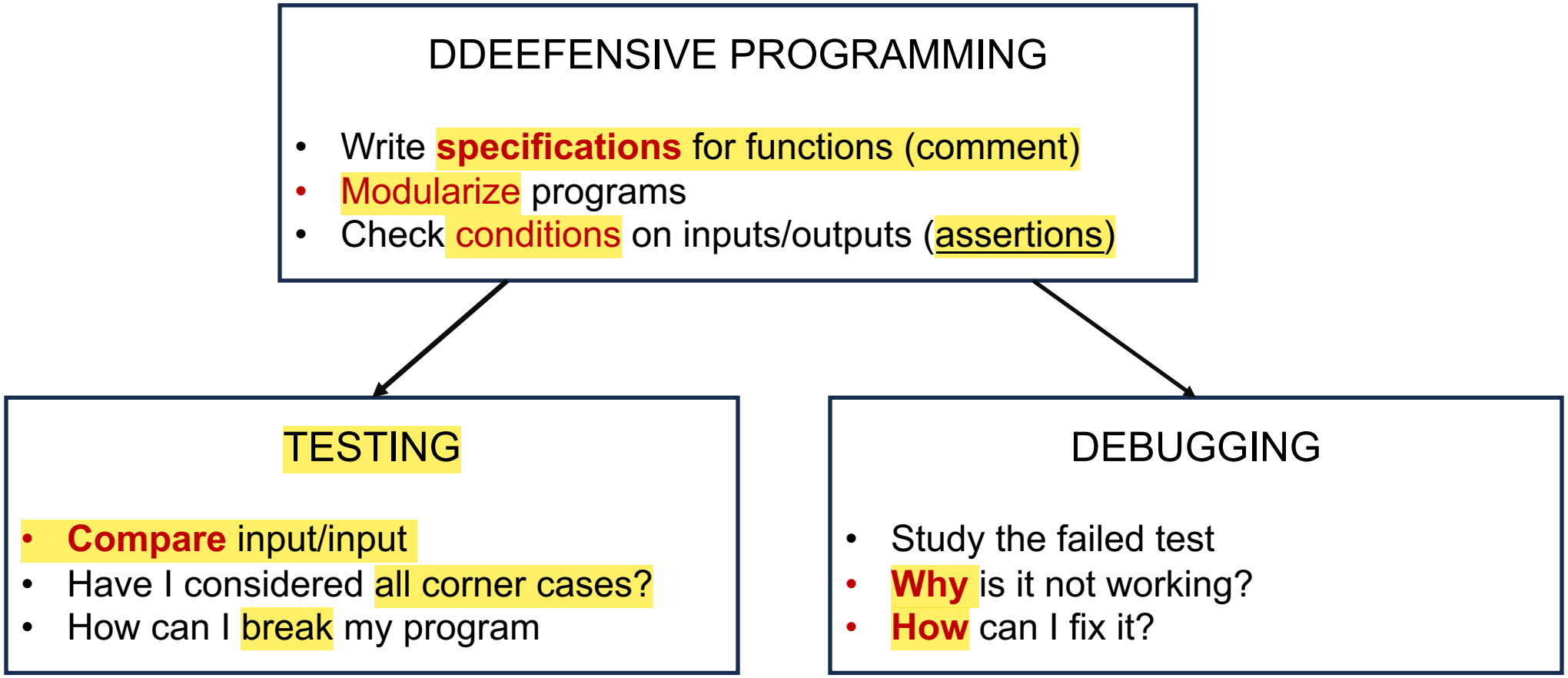
Expectation

https://www.linkedin.com/pulse/vibe-coding-ai-powered-creativity-non-coders-dr-shantanu-kumar-qpmrc/

https://www.pinterest.com/tieamsilenteye/expectation-vs-reality-students-life/

Reality

The idea of testing and debugging



Set yourself up for testing and debugging

At the beginning design code to ease the test and debugging:

Break programs up into modules/functions that can be tested and debugged **separately**.

Document constraints on modules/functions.

- Input
- Outputs

Document **assumptions** behind the code design.

- Expected data types
- Dimensions/shapes

Once you have written the code…

- Ensure the **code runs**
  - Remove syntax/semantics errors
  - Python interpreter can usually find those for you

- Have a set of **expected results**
  - Different input/output pairs

Different classes of tests

- Unit test (UI)
  - Validate each piece of program
  - Testing **each functions** separately

- Integration test / Continuous integration test
  - Validate **overall** program

Black box testing

Glass box testing

- Regression test
  - Add test for bugs as you find them.
  - Catch **reintroduced errors** that were previously fixed.

Debugging

- Study/Read/Write codes
  - **How** did I get the wrong results?
  - Is it part of a certain pattern

- Scientific method
  - Study the available tests
  - Form hypothesis
  - Experiment
  - Pick/choose simplest input to test

`PRINT` statement

Good way to **test hypothesis**

Where to print?

- Functions
  - Input
  - Output
- Parameter

Use **bisection** method

- Put print halfway in code
- Decie where bug may be

In class practice

Other things

In person midterm on **Wednesday, 10/08/24**

Read chap1-3 of Intro to Algorithm

Coming next:

Intro to Numpy.