SYSTEMS DEVELOPMENT FOR COMPUTATIONAL SCIENCE LECTURE 7

Fabian Wermelinger

Harvard University CS107 / AC207

Thursday, September 22nd 2022

LAST TIME

- Python basics (a review)
- Nested environments
- Closures (functional programming)
- Decorators

TODAY

Main topics: Towards Object Oriented Programming (OOP) in Python, Classes, Inheritance and Polymorphism

Details:

- Object Oriented Programming (OOP)
- Python classes
- Inheritance and Polymorphism

AGENDA CHECK:

• Quiz 1 takes place tonight. You have 15 minutes (12 questions). The quiz is available within a 45 minute time window (starting 7:00pm). If you start 35 minutes late, you will have 10 minutes to complete the quiz. Please be on time.

RECAP OF LAST TIME

Every name in Python is a reference to an object in memory

If you need a copy of an object, do it explicitly!

```
import copy as cp

a = [1, 2, 3]
b = cp.copy(a)  # shallow copy
c = cp.deepcopy(a) # deep copy

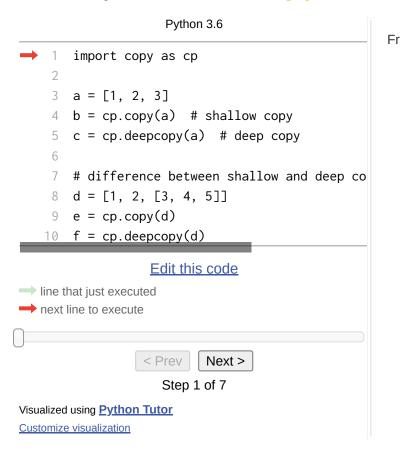
# difference between shallow and deep copy is for compound types
d = [1, 2, [3, 4, 5]] # compound type
e = cp.copy(d)

f = cp.deepcopy(d)
```

RECAP OF LAST TIME

Every name in Python is a reference to an object in memory

If you need a copy of an object, do it explicitly! (pythontutor):



Frames Objects

GENERAL PYTHON GUIDELINES

Now that you are starting to write Python code, you should be aware of some important guidelines and resources:

- Become familiar with the *Python Enhancement Proposals* called PEP for short. All enhancements in Python go through these proposals.
- One of these early proposals is the Python coding style guide (PEP8). Some projects are very strict about code formatting. It is good practice to write *consistently* formatted code.
- The Zen of Python (PEP20) and some vim koans (optional of course)

CODE FORMATTING

- It is generally good advice to have some background of PEP8 (you do not need to know all of it by heart).
- When you code, your mental focus should be on writing correct code and not worrying about consistent formatting!
- You can use tools to do the formatting for you consistently.

For Python code a good tool is called yapf (yet another Python formatter). When you work with C/C++ clang-format is a fantastic tool for code formatting. These tools can be integrated in your coding environment (e.g. vim) or used as standalone tools.

CODE FORMATTING

Example: yapf code formatting (demo in lecture codes)

```
python -m pip install [--user] yapf
```

- yapf looks for style files inside a source code directory (e.g. your git project) or a global configuration defined in your \$HOME, see the formatting style section for all details. If it can't find any customizations it will default to PEP8.
- There are four base presets: (they are named after some companies because they are used there)
 - 1. pep8 (default)
 - 2. google (based on the Google Python style guide)
 - 3. yapf (for use with Google open source projects)
 - 4. facebook
- Example . style . yapf file in the root of your git project:

```
1 [style]
2 based_on_style = pep8
3 spaces_before_comment = 4
4 split_before_logical_operator = True
5 use_tabs = False # this is a personal choice!
```

• You can process many *.py files using a Bash script (e.g. in a project directory that is dedicated to maintenance). You can also integrate yapf in your editor (e.g. vim, emacs) to automagically format your code as you type (remember: your focus is on correct code not worrying about formatting code).

OBJECT ORIENTED PROGRAMMING (MOTIVATION)

Programs = Algorithms + Data Structures

Niklaus Wirth

- We would like to find a way to represent complex, structured *data* in the context of our programming language.
- If you want to model a *particle* for example, you would want to associate every particle with a position x, y, z and possibly a velocity u, v, w for each of the three spatial coordinates.
- In C you could then define a compound *data structure* that describes a particle like this:

```
1 struct Particle {
2    double x, y, z; // particle position
3    double u, v, w; // particle velocity
4 };
```

OBJECT ORIENTED PROGRAMMING (MOTIVATION)

```
Programs = Algorithms + Data Structures

Niklaus Wirth
```

• In C you could then define a compound *data structure* that describes a particle like this:

```
1 struct Particle {
2    double x, y, z; // particle position
3    double u, v, w; // particle velocity
4 };
```

- In reality you will have many particles to deal with (think atoms for example). In addition to the Particle data structure, you would need an *algorithm* that, for example, describes the *interaction* between particles.
- The algorithm and data structure together would allow you to formulate a program to simulate this system of particles.

OBJECT ORIENTED PROGRAMMING (MOTIVATION)

Loosely speaking, we could say that we have defined an *abstract object* to describe a physical particle and we have called this "object" Particle. Any program we formulate with this abstraction will be *oriented* towards this object because it encodes the *data* we need in order to describe a physical process (using *algorithms*). When we write code using such abstract objects we speak of *Object Oriented Programming (OOP)*.

```
1 struct Particle {
2    double x, y, z; // particle position
3    double u, v, w; // particle velocity
4 };
```

OBJECT ORIENTED PROGRAMMING

- The C language is not an object oriented language!
- Although we can encapsulate data in compound objects, we can not form an abstraction of the data because the data in a struct is always accessible by anyone and we must initialize it explicitly.
- OOP goes further than that. We actually want to protect the data from being accessible by anyone.
- Instead, OOP offers an *interface* to perform certain operations with the data that is *private* to the object.
- An interface is defined through *methods* or *member functions* (synonyms) which are accessible by anyone (from the outside).
- This way, the actual *implementation* of how you perform the *transformation of the data* is hidden from the user of your object.
- The Application Programming Interface (API) of your code/library defines the interface you offer the users of your code and therefore the abstraction of your object.
- OOP further consists of concepts such as inheritance (inherit data and interfaces from other objects) and polymorphism (make objects with the same parent behave differently).
 The Greek word "poly" means "many". ευχαριστώ πολύ

 The basic type in Python that we can use to compound data of possibly different types is the tuple:

```
1 def four():
2    return 0x4
3
4 t = (1, 2.0, '3', four)
5 for item in t:
6    print(type(item))

<class 'int'>
    <class 'float'>
    <class 'str'>
    <class 'function'>
```

- Can we do that with a list too?
- What is the difference between a tuple and a list?

Let's use the tuple to model a complex number:

```
def Complex(r, i):
       Construct a complex number
       Arguments:
           r: real part
           i: imaginary part
       return (r, i)
   def real(c):
        """Get the real part of a complex number c"""
       return c[0]
   def imag(c):
        """Get the imaginary part of a complex number c"""
16
       return c[1]
17
18
   def string(c):
        """Represent a complex number c as a string"""
20
21
       return f"{c[0]:e} + i{c[1]:e}"
```

• We can use our complex object like this:

```
1 >>> z = Complex(1, 2)
2 >>> print(real(z), imag(z), string(z))
3 1 2 1.000000e+00 + i2.000000e+00
```

• But data is not private at all:

```
1 >>> z[0]; z[1]
2 1
3 2
```

- It means that we can easily bypass the interface and do things on our own! No proper data encapsulation.
- Furthermore, a tuple is immutable:

```
1 >>> z[0] = 3
2 Traceback (most recent call last):
3 File "<stdin>", line 1, in <module>
4 TypeError: 'tuple' object does not support ...
```

 Not being able to assign other values to a complex number is not useful at all.

Let's try a closure instead: (a closure captures the names from the enclosing scope)

```
def Complex(r, i):
        Construct a complex number
        Arguments:
 5
            r: real part
            i: imaginary part
        11 11 11
        def implementation(method):
            if method.lower() == 'real':
10
                return r
            elif method.lower() == 'imag':
                return i
12
            elif method.lower() == 'string':
13
                return f"{r:e} + i{i:e}"
14
15
        return implementation
16
```

We can use our complex object like this:

```
1 >>> z = Complex(1, 2)
2 >>> print(z('real'), z('imag'), z('string'))
3 1 2 1.000000e+00 + i2.000000e+00
```

- This is better because the implementation (closure) defines all the *methods* allowed for data transformations of Complex objects.
- The implementation defines the interface for this example.
- We need to extend it such that we can *set* new values. These methods are called *setters*. The ones we already implemented are called *getters*.

Let's try a closure instead with setters:

```
def Complex(r, i):
       Construct a complex number
       Arguments:
 5
            r: real part
            i: imaginary part
        11 11 11
       def implementation(method, z=None):
            nonlocal r, i
            if method.lower() == 'set_z':
10
                assert z is not None
12
                r, i = z
            elif method.lower() == 'real':
13
14
                return r
            elif method.lower() == 'imag':
15
                return i
16
            elif method.lower() == 'string':
17
                return f"{r:e} + i{i:e}"
18
19
20
        return implementation
```

• We can use our complex object like this:

```
1 >>> z = Complex(1, 2)
2 >>> print(z('real'), z('imag'), z('string'))
3 1 2 1.000000e+00 + i2.0000000e+00
4 >>> z('set_z', (3, 4))
5 >>> print(z('real'), z('imag'), z('string'))
6 3 4 3.000000e+00 + i4.0000000e+00
```

- Now with support for setting new values through the set_z method.
- There are still many operations missing. **How** would we multiply two complex numbers?
- The nonlocal keyword is used to reset a name (variable) that was bound in a different scope.
- You are studying the nonlocal keyword in more detail in a homework problem. (See also PEP3104)

- We have created our own object system before using tuples and closures.
- We figured that there are still limitations. For example, when working with complex numbers we would like to do arithmetic:

```
1 >>> z0 = Complex(1, 2)
2 >>> z1 = Complex(3, 4)
3 >>> z2 = z0 + z1
4 >>> print(z2('string'))
5 4.000000e+00 + i6.000000e+00
6 >>> z3 = z0 * z1
7 >>> print(z3('string'))
8 -5.000000e+00 + i1.000000e+01
```

- Python already implements a type system for *user defined* types which provides support for such operations.
- User defined types are called classes in OOP.

Let us return to the complex number type:

• The bare minimum (data only, no methods/interface):

```
1 class Complex():
2    """Complex number type"""
3    def __init__(self, real, imag):
4        self.real = real # real part
5        self.imag = imag # imaginary part
```

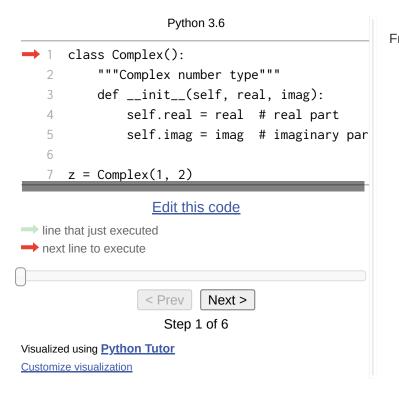
__init__: is a special method automatically run when you execute code like this:

```
1 >>> z = Complex(1, 2) # calls the constructor of the class
2 >>> z.real
3 1
```

For our purposes we call it the *constructor* of the class (see this thread for a lower level discussion).

self: is a reference to the *instance* of the object. Note that unlike to the equivalent this pointer in C++, self *is passed explicitly* to any class method in Python (as the first argument). It is therefore *not a keyword*. You can name it anything you want, "self" is the universally accepted convention. A good read by the creator of Python about the explicit nature of self.

Let us return to the complex number type: pythontutor



Frames Objects

An instance of a class is an allocated object:

• Printing the *class type*:

```
1 >>> print(Complex)
2 <class '__main__.Complex'>
```

• Printing an **instance of the class**:

```
1 >>> z = Complex(1, 2)
2 >>> print(z)
3 <__main__.Complex object at 0x7f669070c2e0>
```

An instance of a class is an object that *occupies memory*. The hex number is its memory address.

Printing the type of an instance:

```
1 >>> print(type(z))
2 <class '__main__.Complex'>
```

Accessing class attributes:

- Private variables do not exist in Python.
- You can access class attributes with the "." operator:

```
1 >>> print(z.real); print(z.imag)
2 1
3 2
```

• Because Python does not prevent access to data, we can print all attributes of a class instance using the vars built-in:

```
1 >>> vars(z)
2 {'real': 1, 'imag': 2}
```

which returns a Python *dictionary* with the attribute name and its current value.

- See this list of the Python built-in functions. Up to here we have seen two:
 - type: return the type of an object
 - vars: return the __dict__ attribute for any object that implements it (we will discuss what this
 means in more detail later).

INHERITANCE AND POLYMORPHISM

- We have covered the main conception of object oriented programming and looked at how "data encapsulation" works in Python.
- At this point you should know how to create simple classes in your Python code.
- We have not talked about inheritance and polymorphism yet.
 Recall:
 - Inheritance: inherit data and interfaces from parent class(es).
 - Polymorphism: make objects with the same parent class(es) behave differently.

INHERITANCE AND POLYMORPHISM

- We have not talked about inheritance and polymorphism yet. Recall:
 - Inheritance: inherit data and interfaces from parent class(es).
 - Polymorphism: make objects with the same parent class(es) behave differently.

A less abstract example: Both, a cat and a dog are animals. Because they are both animals, they share some common features that we describe by data (e.g. name, age, fur color, they speak, and so on). The two speak different languages, however. Cats meow and dogs bark.

- Dogs and cats inherit from an animal class → inheritance.
- Dogs and cats behave differently. Dogs bark and cats meow → polymorphism.

Base class (superclass): modeling cats and dogs

```
1 class Animal():
2    """Base class for animals"""
3    def __init__(self, name, age):
4        self.name = name
5        self.age = age
6
7    def speak(self): # class method, note how 'self' is passed as argument
8    """Sounds animals can make"""
9        raise NotImplementedError
```

- Our intention is to use the Animal class to derive specific animals from. It is called a base class or superclass.
- Because the Animal class is *generic*, we do not know a priori what sounds an inherited animal instance will speak. It raises a NotImplementedError.
- Instances are created by calling the class with arguments defined in the __init__
 method:

```
1 >>> animal = Animal("Generic animal", 5) # instance names are lower case
2 >>> print(vars(animal))
3 {'name': 'Generic animal', 'age': 5}
```

Base class (superclass): modeling cats and dogs

```
1 class Animal():
2    """Base class for animals"""
3    def __init__(self, name, age):
4        self.name = name
5        self.age = age
6
7    def speak(self): # class method, note how `self` is passed as argument
8    """Sounds animals can make"""
9        raise NotImplementedError
```

Instances are created by calling the class constructor:

```
1 >>> animal = Animal("Generic animal", 5) # instance names are lower case
2 >>> print(vars(animal))
3 {'name': 'Generic animal', 'age': 5}
```

 But the speak method is not implemented and the behavior is therefore not defined and the NotImplementedError is thrown:

```
1 >>> animal.speak()
2 Traceback (most recent call last):
3  File ... # output omitted
4  raise NotImplementedError
5 NotImplementedError
```

Derived class (subclass): modeling cats and dogs

• We can create *specific* animals by *inheritance* from the base class:

```
class Dog(Animal): # Dog is a derived class, it inherits from Animal
       """Dog is a derived Animal class"""
       def speak(self):
           """Special sounds dogs make"""
           return "Woof"
   class Cat(Animal): # Cat is a derived class, it inherits from Animal
       """Cat is a derived Animal class"""
       def __init__(self, name, age): # re-define in the derived class
           self.name = f"A very special cat: {name}" # cats have a custom name string
10
11
12
       def speak(self):
           """Special sounds cats make"""
13
           return "Meow"
```

• We can use a *derived class* in the same way as we did with a base class:

```
1 >>> dog = Dog("Snoopy", 6)
2 >>> cat = Cat("Kitty", 4)
3 >>> print(vars(dog)); print(vars(cat))
4 {'name': 'Snoopy', 'age': 6}
5 {'name': 'A very special cat: Kitty'} # Kitty has no age?
```

Derived class (subclass): modeling cats and dogs

• We can use a derived class in the same way as we did with a base class:

```
1 >>> dog = Dog("Snoopy", 6)
2 >>> cat = Cat("Kitty", 4)
3 >>> print(vars(dog)); print(vars(cat))
4 {'name': 'Snoopy', 'age': 6}
5 {'name': 'A very special cat: Kitty'} # Kitty has no age?
```

- What is wrong with Kitty? We have written a special constructor for the Cat class (this is a common scenario), but we did not inherit the age attribute from the base class! (no data inheritance)
- The speak methods sure work!
- But the age attribute for cats is broken:

```
1 >>> print(cat.age)
2 Traceback (most recent call last):
3 File "/home/fabs/harvard/CS107/lecture07/code/06.py", line 40, in <module>
4 print(cat.age)
5 AttributeError: 'Cat' object has no attribute 'age'
```

Derived class (subclass): modeling cats and dogs

- The speak methods sure work!
- But the age attribute for cats is broken:

```
1 >>> print(cat.age)
2 Traceback (most recent call last):
3 File "/home/fabs/harvard/CS107/lecture07/code/06.py", line 40, in <module>
4 print(cat.age)
5 AttributeError: 'Cat' object has no attribute 'age'
```

• It seems that inheritance did not work properly for the derived Cat class. We will talk more about this in a few slides.

Calling member functions (methods): modeling cats and dogs

- Let us focus first on how calling class methods (member functions) works in Python.
- Recall: we have this for dogs:

```
1 class Dog(Animal):
2    """Dog is a derived Animal class"""
3    def speak(self): # takes one argument `self`
4          """Special sounds dogs make"""
5         return "Woof"

1 >>> dog = Dog("Snoopy", 6)
2 >>> print(dog.speak())
3 Woof
```

Calling member functions (methods): modeling cats and dogs

```
1 >>> dog = Dog("Snoopy", 6)
2 >>> print(dog.speak())
3 Woof
```

We are calling the member function speak() without passing any arguments! Why is Python not complaining?

Remember: self is a reference to the object in memory. Since dog is an instance of the Dog class, the instance dog and self reference the same object. If we were to add the following method to Dog:

```
1 def my_id(self):
2    print(id(self))

then we get:
```

```
1 >>> print(id(dog)); dog.my_id()
2 140320390688928
3 140320390688928
```

Calling member functions (methods): modeling cats and dogs

An instance of a class has its methods bound to self and Python knows about it.
 The first argument of class methods is therefore not needed when you call them.

```
1 >>> print(dog.speak); print(Dog.speak)
2 <bound method Dog.speak of <__main__.Dog object at 0x7f9d32bdc0a0>> # dog is an instance of Dog
3 <function Dog.speak at 0x7f9d32ad3160> # Dog is a class type
```

• We can use member functions either way: when *bound* to an instance of the class *or* by passing an instance to Dog. speak. In the latter case, Python does not have a valid self reference, so we *must* pass the instance as the first argument:

```
1 >>> print(dog.speak()); print(Dog.speak(dog))
2 Woof
3 Woof
```

• See:

```
1 >>> Dog.speak()
2 Traceback (most recent call last):
3 File "/home/fabs/harvard/CS107/lecture07/code/06.py", line 42, in <module>
4 Dog.speak()
5 TypeError: speak() missing 1 required positional argument: 'self'
```

Calling member functions (methods): modeling cats and dogs

Summary:

- dog was an instance of the Dog class. It is a reference to an object in memory.
- dog. speak is a bound member function. It is bound to the instance dog. In this case, the name self in the class code behaves as a reference to the *instance* dog.
- Dog was a class type. Member function definitions within this class always have at least one argument. The first argument is a reference to an instance of the class and is conventionally called self. This argument will be consumed for bound methods, i.e., you do not need to pass it when calling the bound method.
- Dog. speak is just a regular function. This function has one argument which is an explicit reference to an instance of the class Dog.
- You can use the isinstance() built-in to check if a name is an instance of a particular class. See the documentation here.

Step through code on pythontutor

Superclass initialization:

- We know that when we inherit from a base class we inherit its data and interface.
- If we recall the derived class Cat from earlier, we would expect when we create an instance of Cat that:
 - 1. The base class is initialized.
 - 2. Additional initialization of the derived class takes place.
- If we do not define another __init__ method in the derived class, then __init__
 of the base class is called automatically.
- If we **do define** another __init__ method in the *derived class* (because we need to perform other initialization there), then the superclass initialization **is not performed automatically anymore** (which is what happened to Cat (Kitty) earlier).
- In that case we have to use the super() built-in function to access the superclass explicitly.

Rule for polymorphism in Python:

If a method exists in the base class as well as in the derived class (both with the same name), then if an instance of the derived class calls the method, Python will execute the implementation of the method in the derived class and vice versa for a base class instance. The super() built-in can be used to call methods in the parent class. The actual method call is determined by the Method Resolution Order (MRO) specified in the __mro__ attribute of the class type.

See this post for additional reading:

https://rhettinger.wordpress.com/2011/05/26/super-considered-super/

Example: rule for polymorphism in Python

```
class Base():
       """Base class"""
 3
       def __init__(self, a):
           self.a = a # some data stored in the base class
 4
 5
6
       def explain(self):
           print(f"Executing from base class: data=`{self.a}`")
8
   class Derived(Base):
       """Derived class"""
10
       def __init__(self, a, b):
12
           super().__init__(a) # properly initialize the base class
           self.b = b # some data specific to the derived class
13
14
15
       def explain(self):
           # 1. Call the base class method first
16
17
           super().explain()
18
           # 2. Then do special work required for the derived class
           print(f"Executing from derived class: data=`{self.b}`")
19
```

Example: rule for polymorphism in Python

 When you call super() it will follow this resolution order to find the matching method in the parent classes:

```
1 >>> print(Base.__mro__)
2 (<class '__main__.Base'>, <class 'object'>)
3 >>> print(Derived.__mro__)
4 (<class '__main__.Derived'>, <class '__main__.Base'>, <class 'object'>)
```

• Lets see how this works for the explain() method of an instance from the Derived class type (previous slide):

```
1 >>> a = "base class data"  # some data we pass for initializing the base class
2 >>> b = "derived class data" # some data we pass for initializing the derived class
3 >>> derived = Derived(a, b) # here we create an instance of the derived class
4 >>> derived.explain()  # and we call the explain() method
5 Executing from base class: data=`base class data`
6 Executing from derived class: data=`derived class data`
```

Example: rule for polymorphism in Python

```
1 class Derived(Base):
2    """Derived class"""
3    def __init__(self, a, b):
4        super().__init__(a) # properly initialize the base class
5        self.b = b # some data specific to the derived class
6
7    def explain(self):
8        # 1. Call the base class method first
9        super().explain()
10        # 2. Then do special work required for the derived class
11        print(f"Executing from derived class: data=`{self.b}`")
```

- In the Cat class from earlier, we **did not** initialize the base class like this, hence the age attribute was missing!
- If you don't do this, you will not enter the base class __init__ method and consequently Python will not bind self.a = a. This causes an error whenever we try to access self.a later on, e.g. when we call the explain() method.
- This agreement to define class methods and attributes **on the fly** as execution flow encounters them, is called **duck typing**. Two instances of a Derived class may suddenly have different attributes defined because of this.

COMMENTS ON DUCK TYPING

- When using duck typing, you are specifying an *implicit* interface. It can speed up the short-term development process as sometimes you do not have a clear picture of a current design.
- Explicit software design (interface is defined before implementation work starts) is more stable especially in large projects. It is also more difficult to design because it requires thinking further into the future at the beginning of the project compared to an implicit duck typing approach where you could inject an interface on the fly.
- When you have an implicit interface through duck typing, make sure to write extensive tests.
- The Python data model is designed around duck typing: If it walks like a duck and it quacks like a duck then it must be a duck.

RECAP

- *Object Oriented Programming:* data encapsulation, inheritance, polymorphism
- Classes: base classes and derived classes
- Inheritance and Polymorphism: class methods, interfaces, method resolution order

Further reading:

- Chapter 10 (Protocols and Duck Typing) and 12 in Luciano Ramalho, "Fluent Python: Clear, Concise, and Effective Programming", O'Reilly Media, 2015
- Explicit self in Python: Why explicit self has to stay
- Method resolution order: super() considered super!