

INT 303 BIG DATA ANALYTICS

#Lab 1: Basics

Jia WANG

Jia.wang02@xjtlu.edu.cn



Xi'an Jiaotong-Liverpool University

西安利物浦大学

	Lecture	Lab	Project
Week1	Introduction	N/A	
Week2	What is data and big data?	Preparation	
Week3	Big data Gramma	Basics	
Week4	Data collection and Data scraping	Pandas	Project 1(Start 10.01)
Week5	Data Virtualization	Data Scraping	
Week6	Infrastructure that supports Big Data processing	Data Virtualization	
Week7	How to tell a good story	Q&A	
Week8	Representing Data and Engineering Features	Q&A	<u>(Due 11.05)</u>
Week9	Dimensionality Reduction	Feature Engineering	
Week10	Big Data Analysis Models	Dimensionality Reduction	Project 2 (Start 11.15)
Week11	Bagging&Boosting	Models	
Week12	Boosting&Stacking	Bagging&Boosting	
Week13	Ethics and social issues.	Boosting&Stacking	
Week14	Review	Q&A	<u>(Due 12.24)</u>

LEARNING OUTCOMES

- Learning Goals
- Getting Started
- Lists
- Simple Functions
- Numpy
- Scipy.stats and plotting distributions
- Conclusions



QUESTION 1

- Given $a = [1, 2, 3]$, $b=a$, after the operation $a[1] = 7$, what does b refer to?
 - $b= [1, 2, 3]$
 - $b= [7, 2, 3]$
 - $b= [1, 7, 3]$

QUESTION 2

Which of the following are true of Python lists?

All elements in a list must be of the same type

A given object may appear in a list more than once

A list may contain any type of object except another list

There is no conceptual limit to the size of a list

These represent the same list:

`['a', 'b', 'c']`

`['c', 'a', 'b']`

QUESTION 3

Python

>>>

```
>>> a = ['foo', 'bar', 'baz', 'qux', 'quux', 'corge']
```

Several short REPL sessions are shown below. Which display correct output?

Python

>>>

```
>>> print(a[-5:-3])
['bar', 'baz']
```

Python

>>>

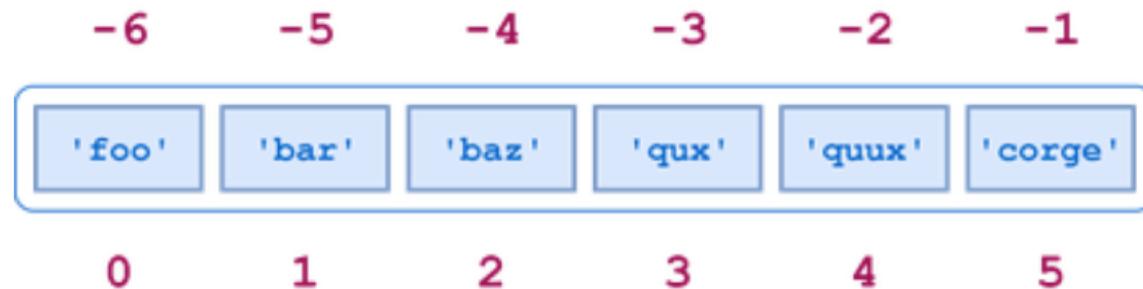
```
>>> max(a[2:4] + ['grault'])
'qux'
```

Python

>>>

```
>>> print(a[4::-2])
['quux', 'baz', 'foo']
```

Here is a diagram to remind you of the list indices:



Python

>>>

```
>>> max(a[2:4] + ['grault'])  
'qux'
```

- `a[2:4]` returns the slice `['baz', 'qux']`. The `+` operator concatenates, so the argument to `max()` is `['baz', 'qux', 'grault']`. The maximum value (for strings, the latest in alphabetical order) is `'qux'`.

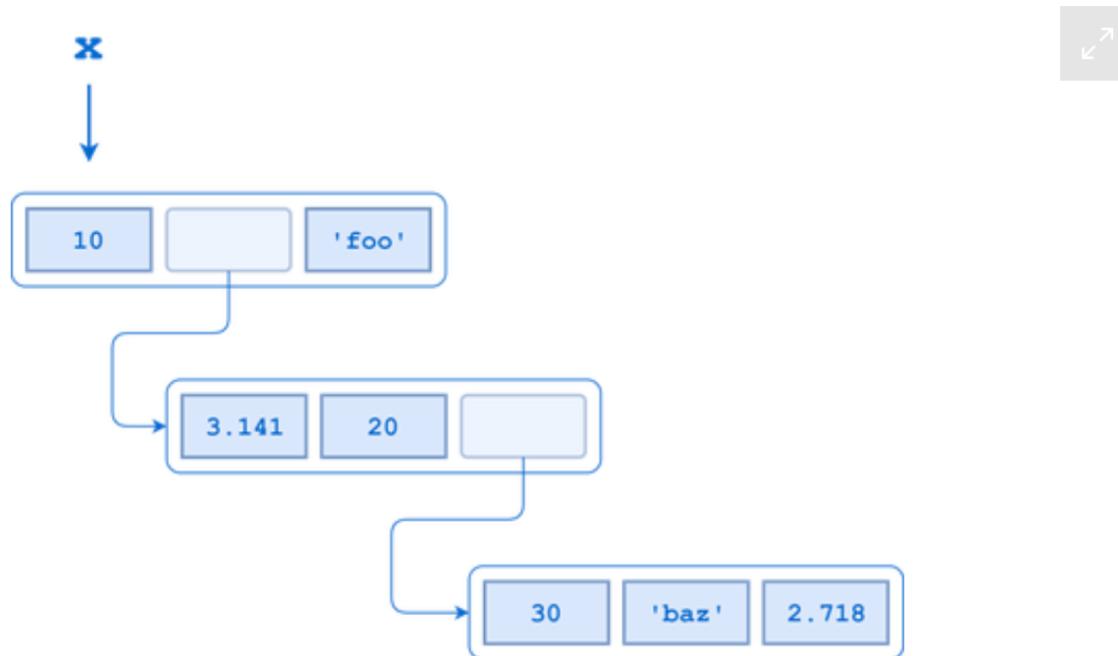
QUESTION 4

Consider the following nested list definition:

Python

```
x = [10, [3.141, 20, [30, 'baz', 2.718]], 'foo']
```

A schematic for this list is shown below:

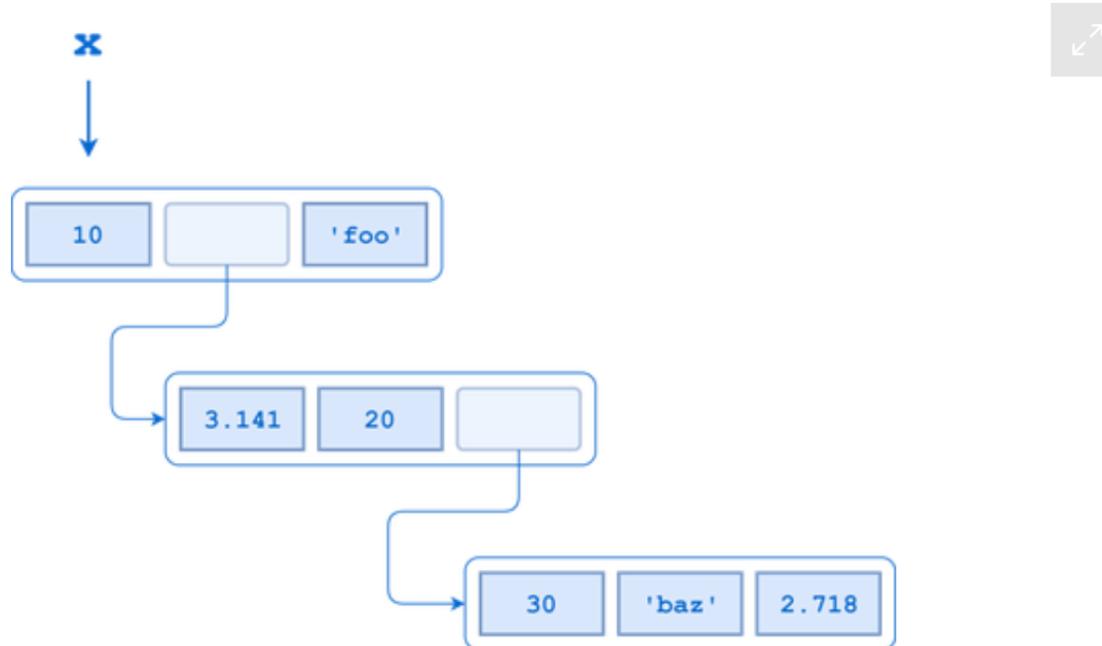


QUESTION 5

Same nested list as the previous question:

Python

```
x = [10, [3.141, 20, [30, 'baz', 2.718]], 'foo']
```



What expression returns the list `['baz', 2.718]`? `x[1][2][1:]`

QUESTION 6

Python

```
a = [1, 2, 3, 4, 5]
```

Select all of the following statements that remove the middle element 3 from a so that it equals [1, 2, 4, 5]:



Python

```
a[2:3] = []
```



Python

```
a.remove(3)
```



Python

```
a[2] = []
```



Python

```
del a[2]
```

QUESTION 7

```
def hypotenuse(x, y):  
    return(x*x + y*y)  
  
hypotenuse(3,4)
```

The output of this function **is** the same **as** which of the following statements?

- A. `hypotenuse = lambda x, y: x*x + y*y`
- B. `hypotenuse = lambda y, x: x*x + y*y`
- C. `square = lambda x: x*x`

QUESTION 8

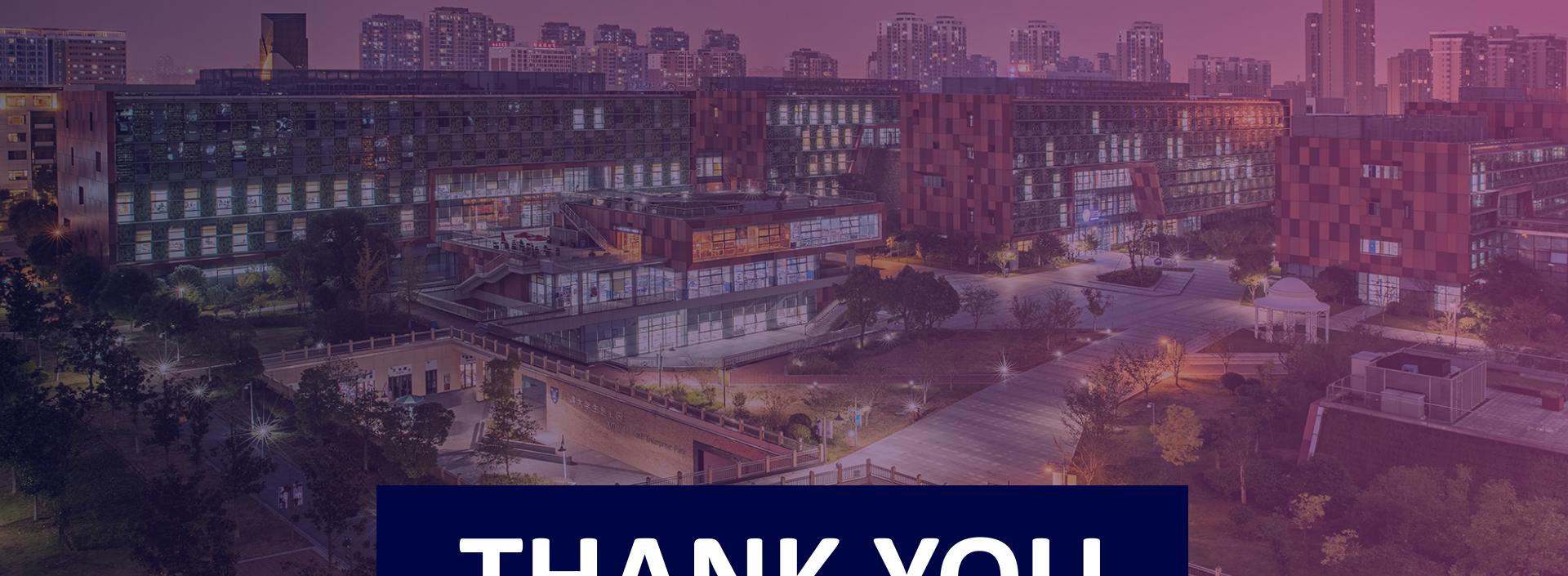
With the following code

```
import numpy as np
import matplotlib.pyplot as plt

x = np.linspace(0, 10, 1000)
plt.plot(x, np.sin(x) + x)
```

the result is

- An error as "np.sin" only works for a single value
- A plot with 1000 elements
- A plot with 2000 elements as "np.sin(x)" and "x" is concatenated



THANK YOU



VISIT US

WWW.XJTLU.EDU.CN



FOLLOW US

@XJTLU



Xi'an Jiaotong-Liverpool University
西安利物浦大学