Introduction to Python

Lecture 4: Packages

Daniel Kadyrov

July 7th, 2023

Python Packages

- One of the most powerful features of Python is the availability of a large number of libraries and packages that can be imported into your code.
- These packages are open-source and managed by the Python community.
- The Python Package Index (PyPI) is a repository of software for the Python programming language. Packages are usually installed from the Python Package Index using the `pip` command.

Installing a package

The following command installs the `numpy` package. This command is performed in the terminal, not in the Python interpreter. Numpy is a package for scientific computing in Python that extends the functionality of Python lists to arrays.

```
pip install numpy
```

To use the package in your code, you need to import it. The following code snippet imports the `numpy` package and assigns it the alias `np`.

```
import numpy as np
```

Numpy

- Numpy is a package for scientific computing in Python that extends the functionality of Python lists to arrays.
- Numpy arrays are more efficient than Python lists.
- Numpy arrays are homogeneous, i.e. they can only contain elements of the same type.
- Numpy arrays can be multidimensional.
- Numpy arrays can be created from Python lists using the `array()` function.

Numpy

The following example shows some of the basic operations that can be performed on numpy arrays. The `linspace` function creates an array of 5 elements between 10 and 14. The `arange` function creates an array of 5 elements between 1 and 5. The `add` function adds the two arrays element-wise.

```
1  import numpy as np
2  a = np.linspace(10, 14, 5)
3  print(a)
4  b = np.arange(1, 6)
5  print(b)
6  print(np.add(a,b))
```

```
[10. 11. 12. 13. 14.]
[1 2 3 4 5]
[11. 13. 15. 17. 19.]
```

Pandas

- Pandas is a Python package for data manipulation and analysis.
- Pandas provides data structures and functions that make working with structured data easier.
- Pandas is built on top of Numpy.
- Pandas provides two data structures: Series and DataFrame.
- A Series is a one-dimensional array of indexed data.
- A DataFrame is a two-dimensional array of indexed data.

Pandas

- Pandas provides functions to read data from different file formats, such as CSV, Excel, JSON, HTML, etc.
- Pandas provides functions to write data to different file formats, such as CSV, Excel, JSON, HTML, etc.
- Pandas provides functions to manipulate data, such as merging, reshaping, selecting, etc.
- Pandas provides functions to perform statistical analysis on data.
- Pandas provides functions to visualize data.

Pandas

The following example shows how to create a DataFrame from a CSV file. The `read_csv` function reads the CSV file and creates a DataFrame. The `head` function displays the first 5 rows of the DataFrame.

```
1  import pandas as pd
2  df = pd.read_csv('data.csv')
3  df.head()
```

	id	age	weight	height
0	1	22	65	170
1	2	25	70	175
2	3	28	75	180
3	4	31	80	185
4	5	34	85	190

Pandas

There are a multitude of functions available from Panda to manipulate data. The following example shows how to select a subset of the data. The `loc` function selects rows and columns by label. The `iloc` function selects rows and columns by position.

```
1  import pandas as pd
2  df = pd.read_csv('data.csv')
3  df.loc[0:2, ['age', 'weight'
       ]]
4  df.iloc[0:2, 1:3]
```

```
   age  weight
0   22      65
1   25      70
2   28      75
   age  weight
0   22      65
1   25      70
```

Pandas

The following example shows how to perform statistical analysis on data. The `describe` function computes summary statistics for numerical columns. The `value_counts` function counts the number of occurrences of each value in a column.

```
1  import pandas as pd
2  df = pd.read_csv('data.csv')
3  df.describe()
4  df['age'].value_counts()
```

Pandas

The following example shows how to visualize data. The `plot` function plots the data in a DataFrame. The `plot` function can be used to plot different types of plots, such as line plots, bar plots, pie plots, scatter plots, etc.

```
1  import pandas as pd
2  df = pd.read_csv('data.csv')
3  df.plot()
4  df.plot(kind='bar')
5  df.plot(kind='pie')
6  df.plot(kind='scatter', x='age', y='weight')
```

Graphing

There are several plotting libraries for Python. The most popular ones are Matplotlib, Seaborn, and Plotly.

Writing Packages

- A package is a collection of Python modules.
- A package is a directory containing `__init__.py` file.
- This file can be empty, and it indicates that the directory it contains is a Python package, so it can be imported the same way a module can be imported.
- A package can contain subpackages, which are subdirectories containing again a `__init__.py` file, and submodules, which are Python scripts like any other Python modules.
- A package can also contain `c-extensions`, which are compiled C code.