Library - Pandas

Jin Wang

MSIS, Rutgers

- Library - Pandas
  - 1. Introduction
  - 2. Basic data structures - Series
  - 3. Basic data structures - DataFrame
  - 4. Obtain characteristics of dataFrame
  - 5. Column selection, addition, deletion
  - 6. Write and Read a csv file

1. Introduction

What is Pandas?

- Pandas is a Python library used for working with data sets.
- It has functions for analyzing, cleaning, exploring, and manipulating data.

Why use pandas?

- Pandas allows us to analyze big data and make conclusions based on statistical theories.
- Pandas can clean messy data sets, and make them readable and relevant.
- Relevant data is very important in data science.

Installation of Pandas

- Windows

```
pip install pandas
```

- Macbook

```
pip3 install pandas
```

Usage

- Like NumPy, we need to import pandas library

```
import pandas as pd
```

2. Basic data structures - Series

Series

- `Series` is a one-dimensional labeled array capable of holding any data type (integers, strings, floating point numbers, Python objects, etc.).
- The axis labels are collectively referred to as the **index**. The basic method to create a `Series` is to call：
  `s = pd.Series(data, index=index)`,
  - where `data` can be many different things:
    - a Python dict
    - a NumPy ndarray
    - a scalar value.
  - The passed `index` is a list of axis labels.

From ndarray

From dictionary

Series is ndarray-like

Series is dict-like

3. Basic data structures - DataFrame

DataFrame

- `DataFrame` is a 2-dimensional labeled data structure with columns of potentially different types.
- You can think of it like a spreadsheet or SQL table, or a dict of Series objects.
- It is generally the most commonly used pandas object. Like Series, DataFrame accepts many different kinds of input:
    - Dict of 1D ndarrays, lists, dicts, or `Series`
    - 2-D numpy.ndarray
    - A `Series`
    - Another `DataFrame`

From series

From dictionary

4. Obtain characteristics of dataFrame

- Shape
- Column name
- Index

5. Column selection, addition, deletion

- Choose a single column
- Choose multiple columns

6. Write and Read a csv file

End