Learn Data Science in Python

Loops and Functions

Feb 25, 2017

Welcome!

February 4	1	Setup & Basics
February 11	2	Matrix & Data Frame
February 25	3	Loop & Function
March 4	4	Report & Data Viz

Welcome!

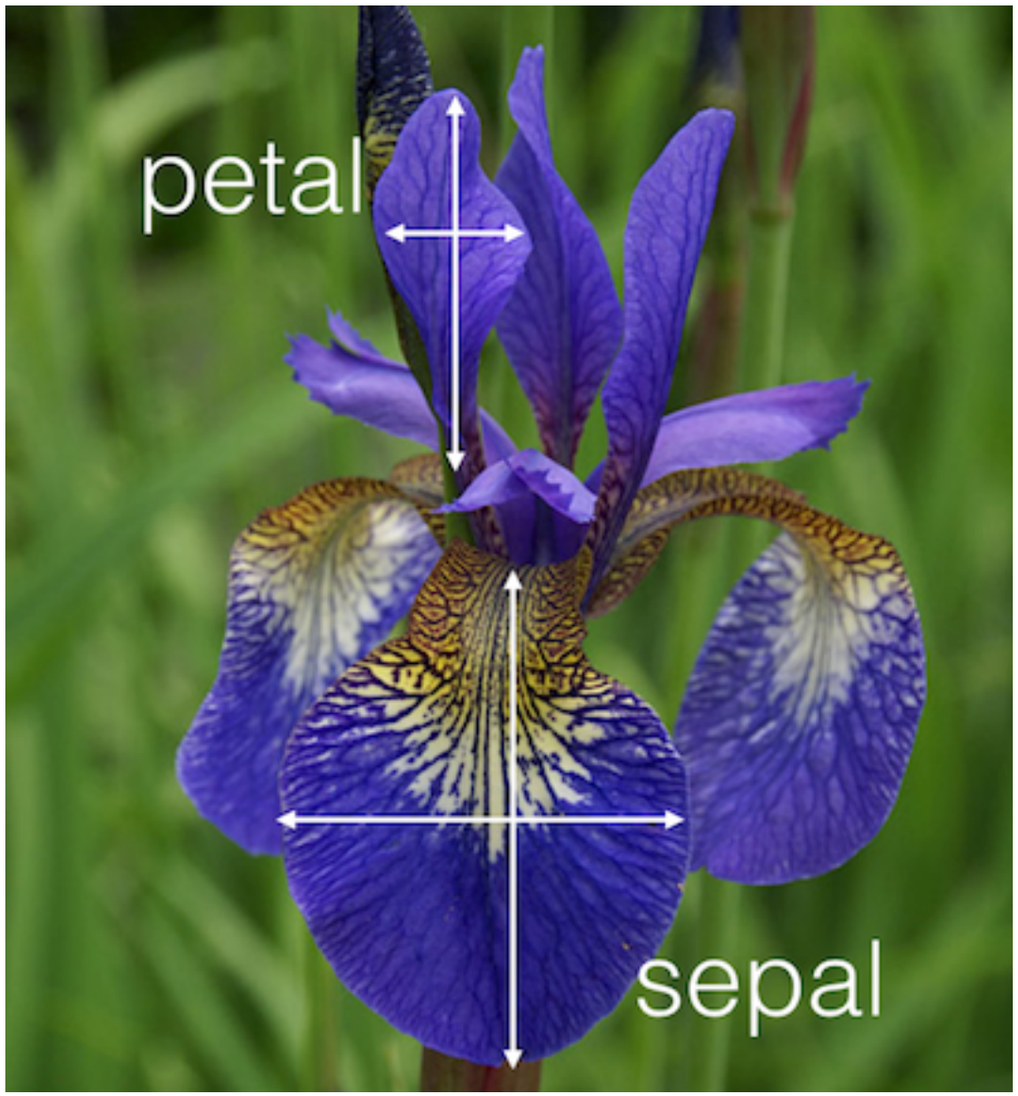
What we've learned previously:

Create Array and Data Frame

Index and Select Rows and Columns

Import Files as Data Frame

Iris Dataset



- Collected by Edgar Anderson
- Analyzed by Ronald Fisher in his 1936 paper
- Focus on classifying 3 different species of Iris based on their petal and sepal
- Shape: `(150, 5)`
- Use `matplotlib/plotly` for Data Visualization
- Use `scikit-learn` for Data Analysis

> Photo from Kaggle's *Machine learning first steps with the Iris dataset*

Data Frame

Import the *Iris* data set and recap how to index, explore and visualize the data in a data frame.

Reading in Data

Most Common Format

CSV	`pandas.read_csv()`
EXCEL	`pandas.read_excel()`

Useful Methods

`.head()` `.unique()` `.describe()` `.hist()`

Loops

Design an algorithm to perform automated iteration of data processing.

Three Types of Loops

While

Most flexible, but slowest form of loop

For in

Quite effective and is faster than while loop

List Comprehension

Least flexible, but ridiculously fast

Three Types of Loops

While

Most flexible, but slowest form of loop

For in + `break, continue, pass`

Quite effective and is faster than while loop

List Comprehension

Least flexible, but ridiculously fast

Functions

Learn how to use and write your own functions.

Where Do You Find Functions

```
?function
```

```
help("function")
```

```
dir("")
```

- Definition
- Parameters
- Examples
- Attributes
- Methods

Two Types of Custom Functions

Traditional

- name (function name)
- parameters (inputs for function)
- return statement (outputs of function)

Lambda

anonymous functions, useful for conciseness

Lab: Libraries

After learning loops and functions, you can import these libraries to do visualizations and statistical analysis.

Cool Libraries

Numpy	Pandas
Array & Matrix	Data Frame

Matplotlib	scikit-learn
Data Visualization	Machine Learning