

LẬP TRÌNH PYTHON

Matplotlib & Ứng dụng

NGUYỄN HẢI TRIỀU¹

¹Bộ môn Kỹ thuật phần mềm,
Khoa Công nghệ thông tin, Trường ĐH Nha Trang

NhaTrang, February 2022

Nội dung

- 1 Matplotlib Pyplot
- 2 Machine learning with scikit-learn

Giới thiệu Matplotlib

Matplotlib là một Python module dùng để vẽ đồ thị, trực quan hóa dữ liệu. Để cài đặt Matplotlib, t sử dụng câu lệnh pip như sau: `pip install matplotlib` và để sử dụng matplotlib cần import module như sau: `import matplotlib`

matplotlib.pyplot

Hầu hết các tiện ích của Matplotlib đều nằm trong submodule pyplot. Để đơn giản khai báo, ta sử dụng alias **plt**:

```
import matplotlib.pyplot as plt
```

Matplotlib thường kết hợp với Numpy để vẽ đồ thị.

- 1 Matplotlib Pyplot
- 2 Machine learning with scikit-learn

```
1 import matplotlib.pyplot as plt
2 plt.plot([-1, -4.5, 16, 23, 78, 22, 3])
3 plt.show()
```

Các giá trị trong danh sách là giá trị của trục y . Giá trị trục x sẽ sinh tự động.

```
1 import matplotlib.pyplot as plt
2 plt.plot([-1, -4.5, 16, 23, 78, 22, 3], "ob") # marker là hình
        tron o, b là màu blue
3 plt.show()
```

Vẽ đồ thị với các giá trị rời rạc

Ví dụ 1.1

Vẽ đồ thị biểu diễn nhiệt độ trong vòng 8 ngày.

```
1 import matplotlib.pyplot as plt
2 days = range(1, 9)
3 celsius_values = [25.6, 24.1, 26.7, 28.3, 27.5, 30.5, 32.8, 33.1]
4 fig, ax = plt.subplots()
5 ax.plot(days, celsius_values)
6 ax.set(xlabel='Day', # nhan cho trục x
        ylabel='Temperature in Celsius', # nhan cho trục y
        title='Temperature Graph') # ten cho đồ thị
9 plt.show()
```

Vẽ nhiều đối tượng trong một đồ thị

```
1 import matplotlib.pyplot as plt
2 days = list(range(1,9))
3 celsius_min = [19.6, 24.1, 26.7, 28.3, 27.5, 30.5, 32.8, 33.1]
4 celsius_max = [24.8, 28.9, 31.3, 33.0, 34.9, 35.6, 38.4, 39.2]
5
6 fig, ax = plt.subplots()
7
8 ax.set(xlabel='Day',
9        ylabel='Temperature in Celsius',
10        title='Temperature Graph')
11 ax.plot(days, celsius_min, "blue", # vẽ đường nhiệt độ nhỏ nhất
12         days, celsius_min, "oy", # vẽ các điểm rơi rạc nhiệt độ
13         days, celsius_max, "black",
14         days, celsius_max, "or")
15
16 plt.show()
```

Vẽ đồ thị cột

```
1 # Bar Plots
2 import matplotlib.pyplot as plt
3 import numpy as np
4 years = [str(year) for year in range(2010, 2021)]
5 visitors = (1241, 50927, 162242, 222093,
6             665004, 2071987, 2460407, 3799215,
7             5399000, 5474016, 6003672)
8 plt.bar(years, visitors, color="blue")
9 plt.xlabel("Years")
10 plt.ylabel("Values")
11 plt.title("Bar Chart Example")
12 plt.plot()
13 plt.show()
```

```
1 # Scatter Plots
2 import numpy as np
3 x = np.arange(0, 11)
4 y1 = np.random.randint(2, 7, (11,))
5 y2 = np.random.randint(9, 14, (11,))
6 y3 = np.random.randint(15, 25, (11,))
7 # Markers: https://matplotlib.org/api/markers\_api.html
8 plt.scatter(x, y1)
9 plt.scatter(x, y2, marker='v', color='r')
10 plt.scatter(x, y3, marker='^', color='m')
11 plt.title('Scatter Plot Example')
12 plt.show()
```


Vẽ nhiều đồ thị trong một hình

function subplot

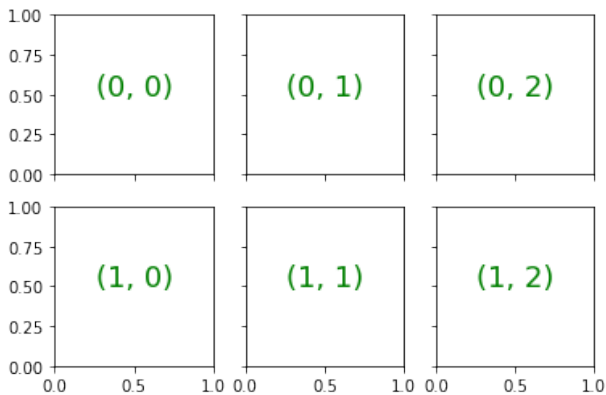
Hàm subplot() nhận ba đối số mô tả bố cục của hình. Hai đối số đầu tiên biểu thị số hàng và cột của hình. Đối số thứ 3 mô tả vị trí hiện tại của đồ thị. Ví dụ

```
1 import matplotlib.pyplot as plt
2 import numpy as np
3 #plot 1:
4 x = np.array([0, 1, 2, 3])
5 y = np.array([3, 8, 1, 10])
6 plt.subplot(1, 2, 1)
7 plt.plot(x,y)
8 #plot 2:
9 x = np.array([0, 1, 2, 3])
10 y = np.array([10, 20, 30, 40])
11 plt.subplot(1, 2, 2)
12 plt.plot(x,y)
13 plt.show()
```

Vẽ nhiều đồ thị trong một hình

```
1 import matplotlib.pyplot as plt
2 plt.figure(figsize=(6, 4))
3 fig, (ax1, ax2) = plt.subplots(1, 2, sharey='row')
4 ax1.text(0.5, 0.5, "left", color="green", fontsize=18, ha='center')
5 ax2.text(0.5, 0.5, "right", color="green", fontsize=18, ha='center')
6 plt.show()
```

```
1 # vẽ đồ thị với số liệu cụ thể
2 import numpy as np
3 import matplotlib.pyplot as plt
4 x = np.linspace(0, 2*np.pi, 400)
5 y = np.sin(x**2) + np.cos(x)
6 f, (ax1, ax2) = plt.subplots(1, 2, sharey=True)
7 derivative = 2 * x * np.cos(x**2) - np.sin(x)
8 ax1.plot(x, y)
9 ax1.set_title('Sharing Y axis')
10 ax2.plot(x, derivative)
```



Hình 1: Hãy vẽ hình như trên

Setting the Plot Range

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3 fig, axes = plt.subplots(1, 3, figsize=(10, 4))
4 x = np.arange(0, 5, 0.25)
5 axes[0].plot(x, x**2, x, x**3)
6 axes[0].set_title("default axes ranges")
7
8 axes[1].plot(x, x**2, x, x**3)
9 axes[1].axis('tight')
10 axes[1].set_title("tight axes")
11
12 axes[2].plot(x, x**2, x, x**3)
13 axes[2].set_ylim([0, 60]) #thiet lap vung hien thi x
14 axes[2].set_xlim([2, 5]) #thiet lap vung hien thi y
15 axes[2].set_title("custom axes range");
```

- 1 Matplotlib Pyplot
- 2 Machine learning with scikit-learn

Machine learning with scikit-learn

Machine learning–ML

Machine learning is about adapting models to data. How data can be represented in order to be understood by the computer?

Iris Dataset, "Hello World" of Machine Learning

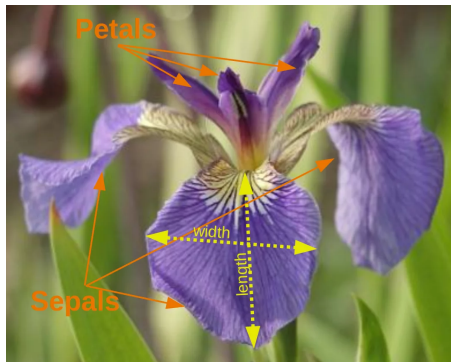
Bộ dữ liệu phổ biến nhất để bắt đầu học ML là *Iris Dataset*. Bộ dữ liệu này được tích hợp trong [scikit-learn](#). Cài đặt scikit-learn: `pip install -U scikit-learn`. Mô tả của bộ dữ liệu này như sau:

① Features in the Iris dataset

- ▶ sepal length in cm
- ▶ sepal width in cm
- ▶ petal length in cm
- ▶ petal width in cm

② Target classes to predict

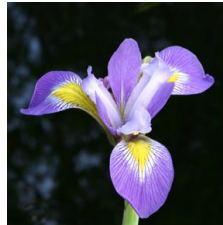
- ▶ Iris Setosa
- ▶ Iris Versicolour
- ▶ Iris Virginica



(a) Iris Setosa



(b) Iris Versicolour



(c) Iris Virginica

Loading the Iris Data with Scikit-learn

Sau khi load iris dataset từ sklearn, tập dữ liệu kết quả là một Bunch object

```
1 #Loading the Iris Data with Scikit-learn
2 from sklearn import datasets
3 iris = datasets.load_iris()
4 type(iris)
5 print(iris.keys())
6 #dict_keys(['data', 'target', 'target_names', 'DESCR', '
    feature_names', 'filename'])
7 print(iris['feature_names'])
8 #[ 'sepal length (cm)', 'sepal width (cm)', 'petal length (cm)', '
    petal width (cm)']
9 print(iris["target_names"])
10 #[ 'setosa' 'versicolor' 'virginica']
11 n_samples, n_features = iris.data.shape
12 print('Number of samples:', n_samples) #Number of samples: 150
13 print('Number of features:', n_features) #Number of features: 4
14 print(iris.data[0]) #[5.1 3.5 1.4 0.2]
15 print(iris.target[0]) #0
```

Visualising the Features of the Iris Data Set

- Hãy in ra các đặc trưng của hoa virginica?
- Hãy vẽ đồ thị histogram của ba loại hoa.

```
1 #histogram for 3 types
2 import matplotlib.pyplot as plt
3 fig, ax = plt.subplots()
4 x_index = 3
5 colors = ['blue', 'red', 'green']
6 for label, color in zip(range(len(iris.target_names)), colors
7                           ):
8     ax.hist(iris.data[iris.target==label, x_index],
9             label=iris.target_names[label], color=color)
10 ax.set_xlabel(iris.feature_names[x_index])
11 ax.legend(loc='upper right')
12 plt.show()
```

- Hãy vẽ đồ thị cho hai đặc trưng đầu tiên của các loại hoa

```
1 #plot all features
2 import matplotlib.pyplot as plt
3 n = len(iris.feature_names)
4 fig, ax = plt.subplots(n, n, figsize=(16, 16))
5 colors = ['blue', 'red', 'green']
6 for x in range(n):
7     for y in range(n):
8         xname = iris.feature_names[x]
9         yname = iris.feature_names[y]
10        for color_ind in range(len(iris.target_names)):
11            ax[x, y].scatter(iris.data[iris.target==color_ind, x
12                               ], iris.data[iris.target==color_ind, y],
13                               label=iris.target_names[color_ind],
14                               c=colors[color_ind])
15        ax[x, y].set_xlabel(xname)
16        ax[x, y].set_ylabel(yname)
17        ax[x, y].legend(loc='upper left')
18 plt.show()
```

```
1 import matplotlib.pyplot as plt
2 from sklearn.datasets import load_iris
3 from mpl_toolkits.mplot3d import Axes3D
4 iris = load_iris()
5 X = []
6 for iclass in range(3):
7     X.append([[], [], []])
8     for i in range(len(iris.data)):
9         if iris.target[i] == iclass:
10             X[iclass][0].append(iris.data[i][0])
11             X[iclass][1].append(iris.data[i][1])
12             X[iclass][2].append(sum(iris.data[i][2:]))
13
14 colours = ("r", "g", "y")
15 fig = plt.figure()
16 ax = fig.add_subplot(111, projection='3d')
17
18 for iclass in range(3):
19     ax.scatter(X[iclass][0], X[iclass][1], X[iclass][2], c=
20               colours[iclass])
21 plt.show()
```

Hoặc tham khảo tại https://scikit-learn.org/stable/auto_examples/datasets/plot_iris_dataset.html#sphx-glr-download-auto-examples-datasets-plot-iris-dataset

Tài liệu tham khảo



Trung tâm tin học, Đại Học KHTN Tp.HCM
Lập trình Python nâng cao. 03/2017.



Bernd Klein
Data Analysis: Numpy, Matplotlib and Pandas.
bernd.klein@python-course.eu, 2021.



Luciano Ramalho
Fluent Python (2nd Edition). *O'Reilly Media, Inc, 2021.*



Python Software Foundation
<https://docs.python.org/3/tutorial/>