

3ilINGENIEURS

Beyond the curves seen previously, the aim of this lecture is to discover the essential graphs for scientific computation (in particular for Data Science) and learn

</les sciences informatiques> **Data Science Lecture** By Raymond Houé Ngouna

Python and scientific computation: Advanced Matplotlib

(raymond.houe-ngouna@enit.fr)

Lectures objectives

how to use them.

1. Scatter plots.

2. 3D graphs.

3. Histograms.

[Outline]

In []: import numpy as np

In []: %matplotlib inline

plt.show()

plt.show()

In []: import pandas as pd

4. Contour plots

Scatter plots

In []: x = x = np.linspace(0,7,1000)

Examples of scatter plots

plt.figure(figsize=(10,5))

plt.scatter(x,y_raw)

plt.ylabel('y=f(x)')

In []: plt.figure(figsize=(10,5))

plt.xlabel('x')

plt.plot(x,y,c='r',lw=3)

plt.ylabel('y=f(x)')

Using real-world data

In []: plt.figure(figsize=(10,5))

In []: plt.figure(figsize=(10,5))

plt.xlabel('Age')

plt.show()

plt.show()

[Outline]

3D Graphs

3D scatter plots

In []: # allow interactive plots #%matplotlib widget

In []: | %matplotlib inline

plt.show()

Surfaces

In []: x = np.linspace(0,7,100)

In []: print('Dim. x:',x.shape)

In []: x[:3]

In []: | y[:3]

In []: | xx[:3,:3]

In []: yy[:3,:3]

In []: | %matplotlib inline

plt.show()

In []: | %matplotlib inline

plt.show()

[Outline]

Histograms

1D Histogram

In []: plt.figure(figsize=(10,5))

In []: plt.figure(figsize=(10,5))

In []: plt.figure(figsize=(10,5))

plt.ylabel('Age')

In []: plt.figure(figsize=(10,5))

In []: plt.figure(figsize=(10,5))

plt.show()

plt.show()

plt.show()

plt.show()

plt.legend()

plt.legend()

In []: plt.figure(figsize=(10,5))

plt.ylabel('Age')

In []: plt.figure(figsize=(10,5))

plt.ylabel('Area Income')

plt.ylabel('Daily Internet Usage')

plt.xlabel('Income of the area')

plt.xlabel('Daily Internet Usage')

plt.xlabel('Daily Internet Usage')

In []: np.corrcoef(df['Daily Internet Usage'], df['Age'])

In []: # based on mesgrid xx and yy previously created

plt.contour(xx,yy,zz,30,colors='black')

plt.contourf(x,y,zz,cmap='terrain')

fig = plt.figure(figsize=(15,10)) ax = plt.axes(projection='3d')

ax.set xlim3d(-np.pi, 3*np.pi);

ax.set_zlim3d(-np.pi, 2*np.pi);

ax.set_ylim3d(0, 3*np.pi);

Image visualization

face = misc.face(gray=True)

plt.imshow(face, cmap='gray')

plt.hist(face.ravel(), bins=255)

bact_image = mpimg.imread('../figures/bacteria.png')

ax.plot_surface(xx,yy,zz,cmap='terrain')

Scatter plots | 3D graphs | Histograms | Contour plots

cset = ax.contour(xx, yy, zz, 20, zdir='z', offset=-np.pi, cmap='terrain') cset = ax.contour(xx, yy, zz, 20, zdir='x', offset=-np.pi, cmap='terrain') cset = ax.contour(xx, yy, zz, 20, zdir='y', offset=3*np.pi, cmap='terrain')

Scatter plots | 3D graphs | Histograms | Image visualization

plt.ylabel('Daily Time Spent on Site')

plt.ylabel('Daily Time Spent on Site')

In []: np.corrcoef(df['Daily Time Spent on Site'], df['Area Income'])

In []: np.corrcoef(df['Daily Internet Usage'], df['Daily Time Spent on Site'])

plt.hist2d(df['Daily Internet Usage'], df['Age'], cmap='terrain')

plt.legend()

plt.legend()

plt.legend()

2D Histogram

In []: plt.figure(figsize=(10,5))

plt.colorbar()

In []: plt.figure(figsize=(10,5))

plt.colorbar()

In []: plt.figure(figsize=(10,5))

plt.ylabel('Age') plt.colorbar()

Contour plots

plt.figure(figsize=(7,7))

plt.contour(xx,yy,zz,30)

plt.contour(xx,yy,zz)

plt.colorbar()

In []: plt.figure(figsize=(10,7))

In []: plt.figure(figsize=(7,7))

In []: plt.figure(figsize=(10,7))

ax.set_xlabel('x') ax.set_ylabel('y') ax.set_zlabel('z')

plt.colorbar()

plt.colorbar()

plt.show()

plt.show()

plt.show()

plt.show()

In []: %matplotlib inline

plt.show()

[Outline]

In []: from scipy import misc

plt.figure()

In []: plt.figure(figsize=(10,5))

In []: import matplotlib.image as mpimg

plt.imshow(bact image)

In []: bact_image_gray = bact_image[:,:,0]

plt.imshow(bact_image_gray, cmap='gray')

plt.hist(bact_image_gray.ravel(), bins=255)

plt.imshow(bact_image_gray, cmap='gray')

plt.imshow(bact_image_gray<0.22, cmap='gray')</pre>

plt.imshow(bact_image_gray<0.3, cmap='gray')</pre>

plt.show()

plt.show()

plt.figure()

bact_image.shape

plt.figure()

In []: bact image gray.shape

In []: plt.figure(figsize=(10,5))

In []: plt.figure(figsize=(10,5)) plt.subplot(1,3,1)

plt.subplot(1,3,2)

plt.subplot(1,3,3)

Congratulations !!!

plt.show()

plt.show()

plt.show()

In []:

In []:

In []:

In []:

In []:

In []:

In []:

Out[1]:

In []:

In [1]: from IPython.core.display import HTML

return HTML(styles)

styles = open("../styles/custom.css", "r").read()

def css styling():

css_styling()

plt.show()

plt.show()

[Outline]

plt.show()

plt.show()

plt.show()

In []: plt.figure(figsize=(10,5))

plt.show()

plt.show()

plt.show()

In []: plt.figure(figsize=(10,5))

plt.show()

ax.set_xlabel('x') ax.set_ylabel('y') ax.set_zlabel('z')

ax.set_xlabel('x') ax.set_ylabel('y') ax.set_zlabel('z')

y = np.linspace(0,6,100)xx, yy = np.meshgrid(x,y)

print('Dim. y:',y.shape) print('Dim. xx:',xx.shape) print('Dim. yy:',yy.shape) print('Dim. zz:',zz.shape)

fig = plt.figure(figsize=(15,10))

fig = plt.figure(figsize=(15,10)) ax = plt.axes(projection='3d')

ax.plot_surface(xx,yy,zz,cmap='plasma')

Scatter plots | 3D Graphs | Contour plots | Image visualization

plt.hist(df['Daily Time Spent on Site']) plt.ylabel('Daily Time Spent on Site')

plt.ylabel('Daily Time Spent on Site')

plt.ylabel('Income of the area')

plt.ylabel('Daily Time Spent on Site')

plt.ylabel('Daily Time Spent on Site')

plt.hist(df['Age'][df['Clicked on Ad'] == 0], bins=30,

plt.hist(df['Age'][df['Clicked on Ad'] == 1], bins=30,

plt.hist(df['Area Income'][df['Clicked on Ad'] == 0], bins=30,

plt.hist(df['Area Income'][df['Clicked on Ad'] == 1], bins=30,

plt.hist(df['Age'], bins=30, edgecolor='black')

plt.hist(df['Area Income'], bins=30, edgecolor='black')

ax = plt.axes(projection='3d')

#%matplotlib notebook

ax.set_xlabel('Age')

In []: from mpl_toolkits.mplot3d import Axes3D

fig = plt.figure(figsize=(15,10)) ax = plt.axes(projection='3d')

ax.set_ylabel('Daily Time Spent on Site')

zz = np.sin(xx) + np.cos(xx+yy)*np.cos(xx)

ax.set_zlabel('Income of the area')

plt.xlabel('x')

Random data for the illustration

import matplotlib.pyplot as plt

y = np.sin(x) + np.cos(2*x) #- 0.03*x $y_raw = y + np.random.randn(len(x))$

plt.scatter(x,y_raw,s=50,alpha=0.5)

df = pd.read_csv('../DATA/advertising.csv')

plt.xlabel('Daily Internet Time Usage') plt.ylabel('Daily Time Spent on Site')

plt.ylabel('Daily Time Spent on Site')

['Daily Time Spent on Site', 'Age', 'Area Income', 'Daily Internet Usage', 'Ad Topic Line', 'City', 'Male', 'Country', 'Timestamp', 'Clicked on Ad']

s=df['Area Income']*0.002, alpha=0.7)

s=df['Area Income']*0.002, alpha=0.7)

Scatter plots | Histograms | Contour plots | Image visualization

plt.scatter(df['Age'], df['Daily Time Spent on Site'], c=df['Clicked on Ad'],

ax.scatter(df['Age'], df['Daily Time Spent on Site'], df['Area Income'],

ax.plot_surface(xx,yy,zz,cmap='terrain',rstride=1, cstride=1, linewidth=0)

plt.hist(df['Daily Time Spent on Site'], bins=30, edgecolor='black')

plt.hist(df['Daily Time Spent on Site'][df['Clicked on Ad'] == 0], bins=30,

plt.hist(df['Daily Time Spent on Site'][df['Clicked on Ad'] == 1], bins=30,

plt.hist(df['Daily Time Spent on Site'][df['Clicked on Ad'] == 0], bins=30,

plt.hist(df['Daily Time Spent on Site'][df['Clicked on Ad'] == 1], bins=30, alpha=0.5, edgecolor='black', label='Clicked', color='g')

alpha=0.5, edgecolor='black', label='Not Clicked', color='r')

alpha=0.5, edgecolor='black', label='Not Clicked', color='r')

alpha=0.5, edgecolor='black', label='Not Clicked', color='r')

alpha=0.5, edgecolor='black', label='Clicked', color='g')

plt.hist(df['Daily Internet Usage'][df['Clicked on Ad'] == 0], bins=30,

plt.hist(df['Daily Internet Usage'][df['Clicked on Ad'] == 1], bins=30,

alpha=0.5, edgecolor='black', label='Clicked', color='g')

plt.hist2d(df['Area Income'], df['Daily Time Spent on Site'], cmap='plasma')

plt.hist2d(df['Daily Internet Usage'], df['Daily Time Spent on Site'], cmap='plasma')

alpha=0.5, edgecolor='black', label='Not Clicked', color='r')

alpha=0.5, edgecolor='black', label='Clicked', color='g')

edgecolor='black', label='Not Clicked', color='r')

edgecolor='black', label='Clicked', color='g')

c=df['Clicked on Ad'], s=df['Area Income']*0.003, edgecolor='black', alpha=0.5)

plt.scatter(df['Daily Internet Usage'], df['Daily Time Spent on Site'], c=df['Clicked on Ad'],

5. Image visualization.

Scatter plots | 3D graphs | Histograms | Contour plots | Image visualization

Main graphs: