实验八（一）

**实验目的**

了解数据的全生命周期

了解基本的数据处理过程和数据可视化

**实验内容**

**1、matplotlib**

Matplotlib 是一个 Python 的 2D绘图库，它以各种硬拷贝格式和跨平台的交互式环境生成出版质量级别的图形。通过 Matplotlib，开发者可以仅需要几行代码，便可以生成绘图，直方图，功率谱，条形图，错误图，散点图等。本次试验需要在本地机器运行，kfcoding平台目前不支持可视化显示。

​

参考学习地址：

https://matplotlib.org/gallery/index.html



安装matplotlib，如果你使用Anaconda，直接使用下面的命令进行安装

conda install matplotlib



或者使用pip进行安装

pip install matplotlib



安装之后，导入matplotlib模块就可以使用，下面是一些简单的例子：

#导入matplotlib

import matplotlib.pyplot as plt

import numpy as np

​

n = 1024

X = np.random.normal(0,1,n)

Y = np.random.normal(0,1,n)

​

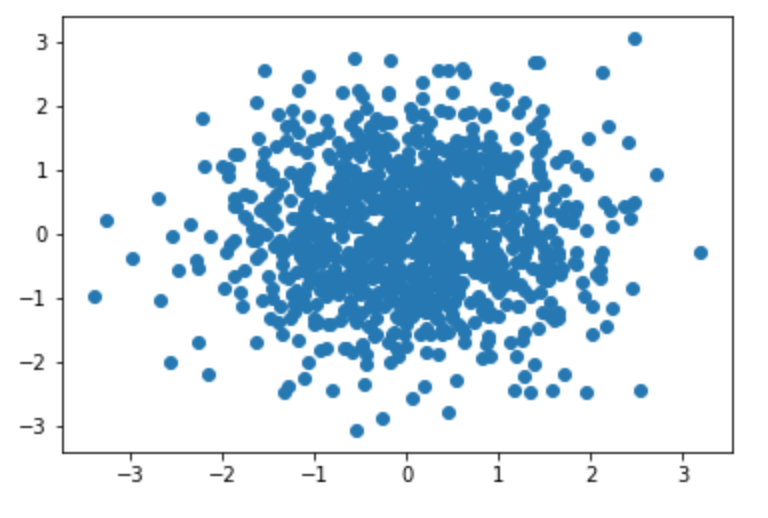
plt.scatter(X,Y) #绘制散点图

plt.show() #显示图像

​



运行结果：

​​

​

import matplotlib.pyplot as plt

import numpy as np

X = np.linspace(-np.pi, np.pi, 256,endpoint=True)

C,S = np.cos(X), np.sin(X)

​

plt.plot(X,C)

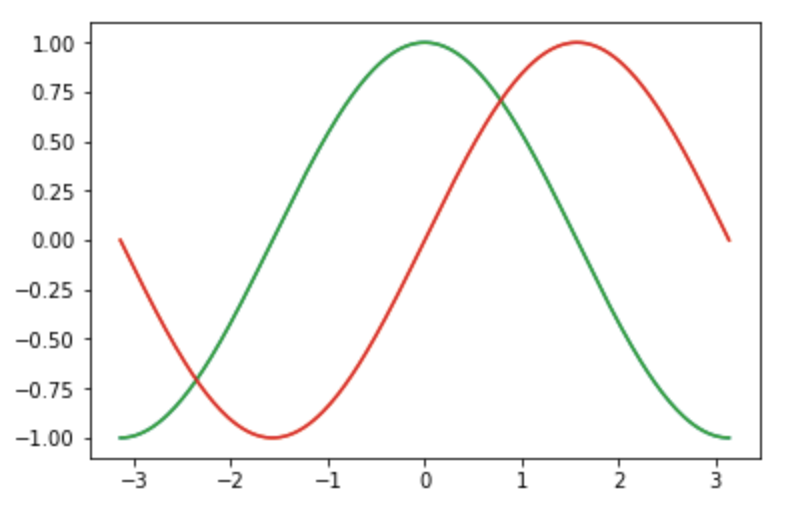
plt.plot(X,S)

​

plt.show()



运行结果：

​​

import matplotlib.pyplot as plt

import numpy as np

​

def f(x,y): return (1-x/2+x\*\*5+y\*\*3)\*np.exp(-x\*\*2-y\*\*2)

​

n = 256

x = np.linspace(-3,3,n)

y = np.linspace(-3,3,n)

X,Y = np.meshgrid(x,y)

​

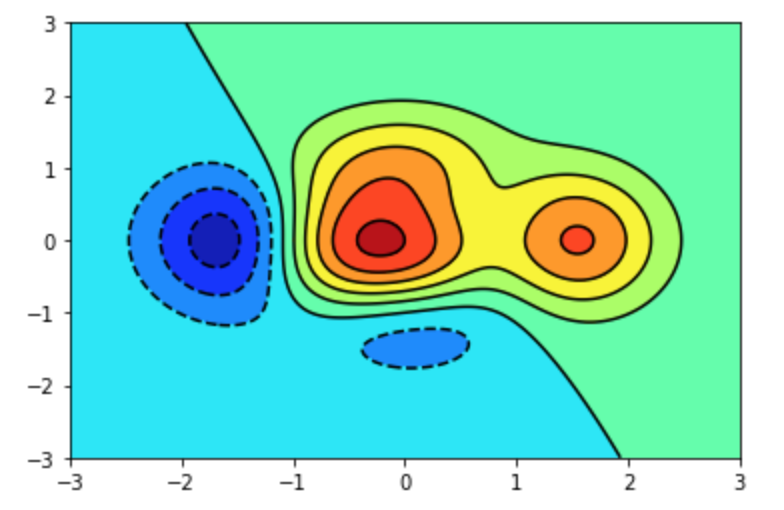
plt.contourf(X, Y, f(X,Y), 8, alpha=.75, cmap='jet')

C = plt.contour(X, Y, f(X,Y), 8, colors='black', linewidth=.5)

plt.show()



运行结果:

​​

**2、数据可视化**

本次试验提供了一个数据集test.csv，本次试验将基于这一数据集，了解和熟悉数据分析、数据可视化和分析的整个过程。

读取数据集并使用matplotlib进行可视化分析，数据展示效果如下图所示。我们可以很清晰的辨别出来整个数据集大致分为两个不同的类别。

def loadDataSet():

dataMat = []

labelMat = []

fr = open('testset.csv')

for line in fr.readlines():

lineArr = line.strip().split()

dataMat.append([1.0,float(lineArr[0]),float(lineArr[1])]) #生成一个包含三个元素的列表,且第一个全是1

labelMat.append(int (lineArr[2]))

fr.close()

return dataMat,labelMat



def show\_data(data,label):

data\_plus = []

data\_minus = []

for i in range(len(data)):

if label[i] ==1 :

data\_plus.append(data[i])

else:

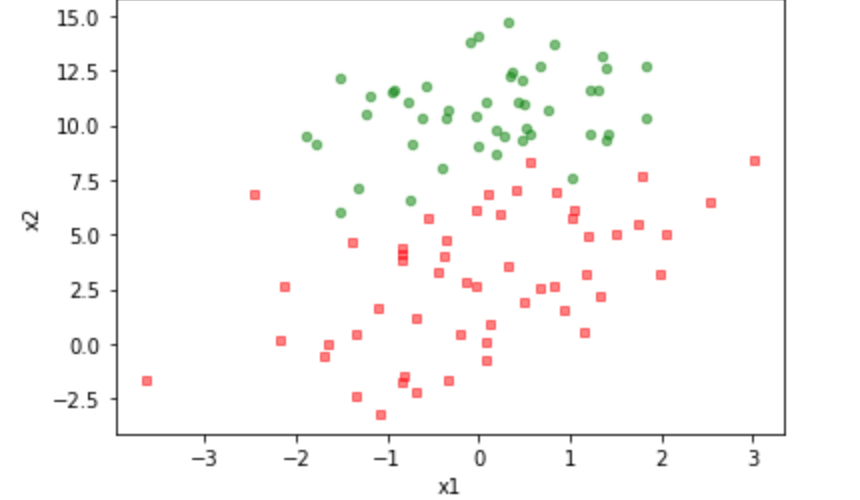
data\_minus.append(data[i])

plt.scatter(np.transpose(data\_plus)[1],np.transpose(data\_plus)[2],s=30,alpha=0.7)

plt.scatter(np.transpose(data\_minus)[1],np.transpose(data\_minus)[2],s=30,alpha=0.7)

plt.show()



​​

那么，基于我们已知的数据信息，如何设计一个分类器来划分这两类点呢？下次实验，我们将会介绍如何设计一个算法自动的判别输入数据的类别。

​