## 数据选取

## 数据来源

## （二）指标选取

## 

在得到的数据中，我们选取了“UPDRS\_ON，UPDRS\_OFF，Yrs Since Diagnosis”三个变量做为贝叶斯公式统计建模的变量，以便于更好分析结果"UPDRS\_ON"、"UPDRS\_OFF"和"Yrs Since Diagnosis"是与帕金森疾病相关的三个变量，它们在评估和诊断帕金森疾病时经常被使用。

## UPDRS\_ON：UPDRS（Unified Parkinson's Disease Rating Scale）是一种用于评估帕金森病患者运动功能障碍程度的量表。UPDRS\_ON表示在药物治疗下，患者在"ON"状态时的UPDRS评分。

## UPDRS\_OFF：UPDRS\_OFF表示在患者未服药或药物作用减弱时的UPDRS评分。UPDRS\_OFF评分较高意味着患者在未经药物干预时的运动功能障碍较严重。UPDRS\_OFF评分通常用于评估帕金森病的基础症状严重程度。

## Yrs Since Diagnosis：Yrs Since Diagnosis表示自帕金森病诊断以来的年数。这个变量用于衡量帕金森病的持续时间。随着疾病的发展，帕金森病的症状可能会逐渐加重，因此Yrs Since Diagnosis可用作评估帕金森病的进展程度的指标之一。

## 这三个变量都是在帕金森疾病的评估和监测过程中被广泛使用的指标。通过对这些变量的观察和分析，可以了解患者的运动功能障碍程度、药物疗效以及疾病的持续时间，从而帮助医生做出诊断、评估疾病进展，并制定相应的治疗计划

## 一、pycharm下载安装

pycharm下载地址：  
http://www.jetbrains.com/pycharm/download/#section=windows  
下载详细步骤：

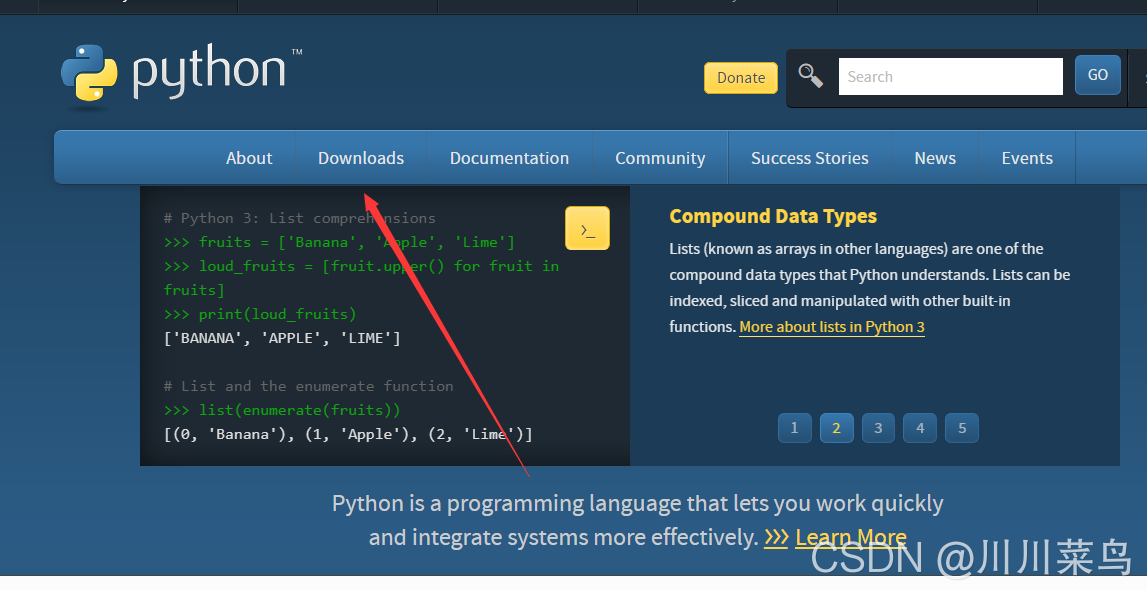


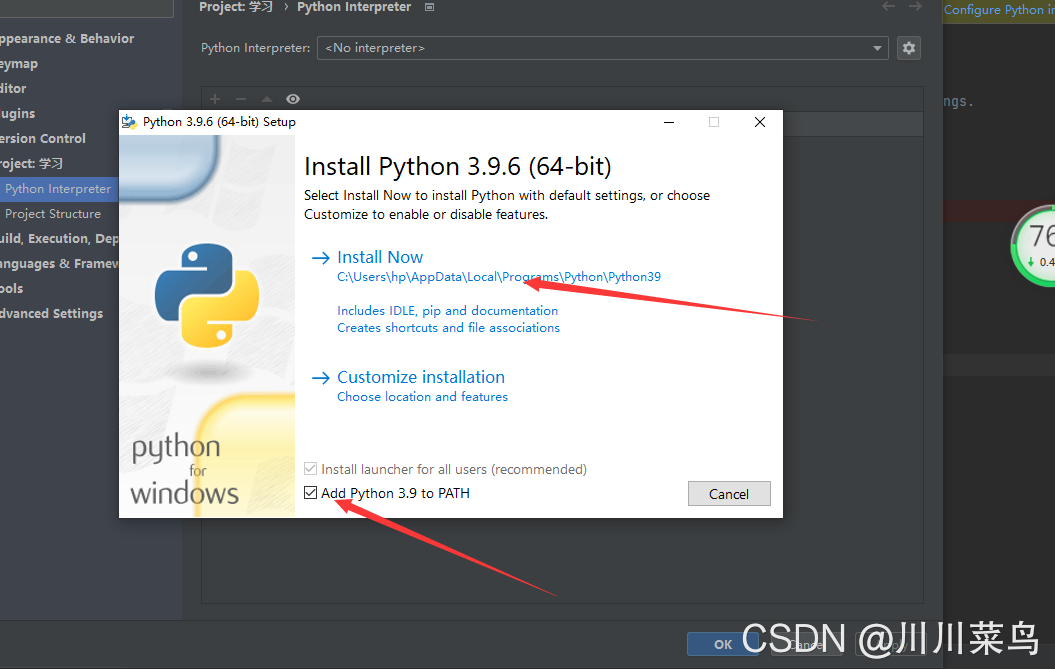
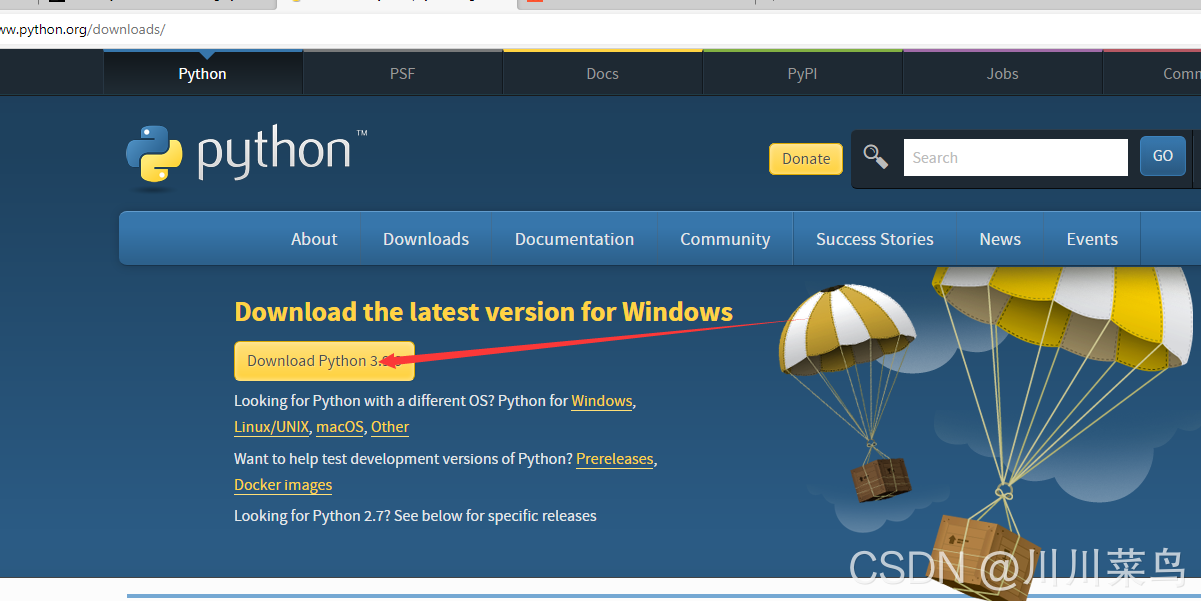
有专业版和免费版，我们在这使用的是专业版

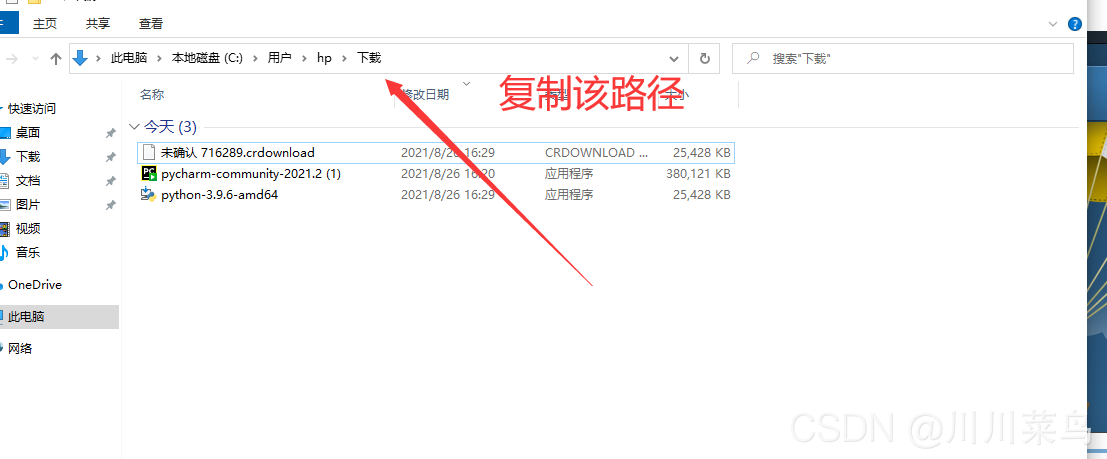
**二.**[python](https://so.csdn.net/so/search?q=python&spm=1001.2101.3001.7020)**下载安装**

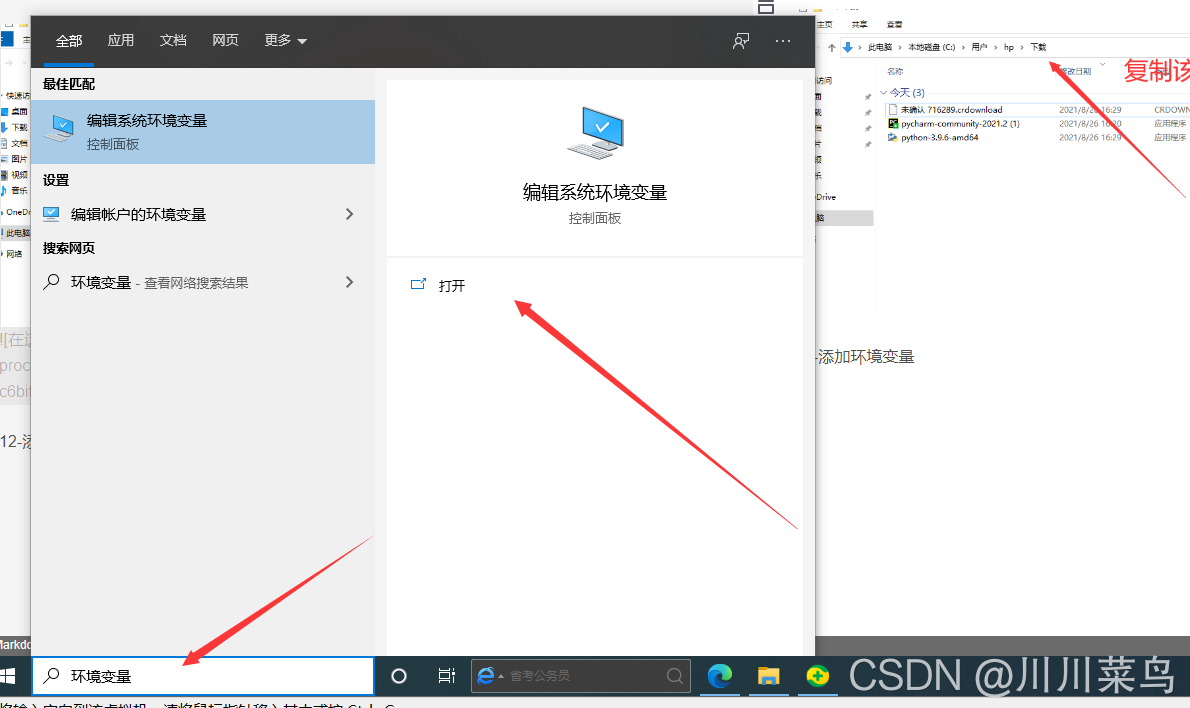
9-python官网：

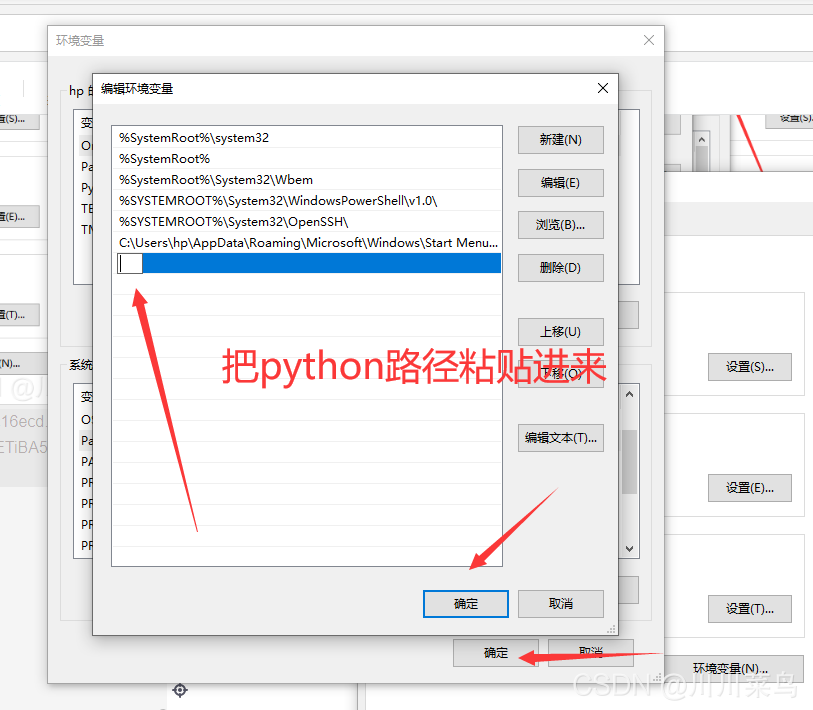
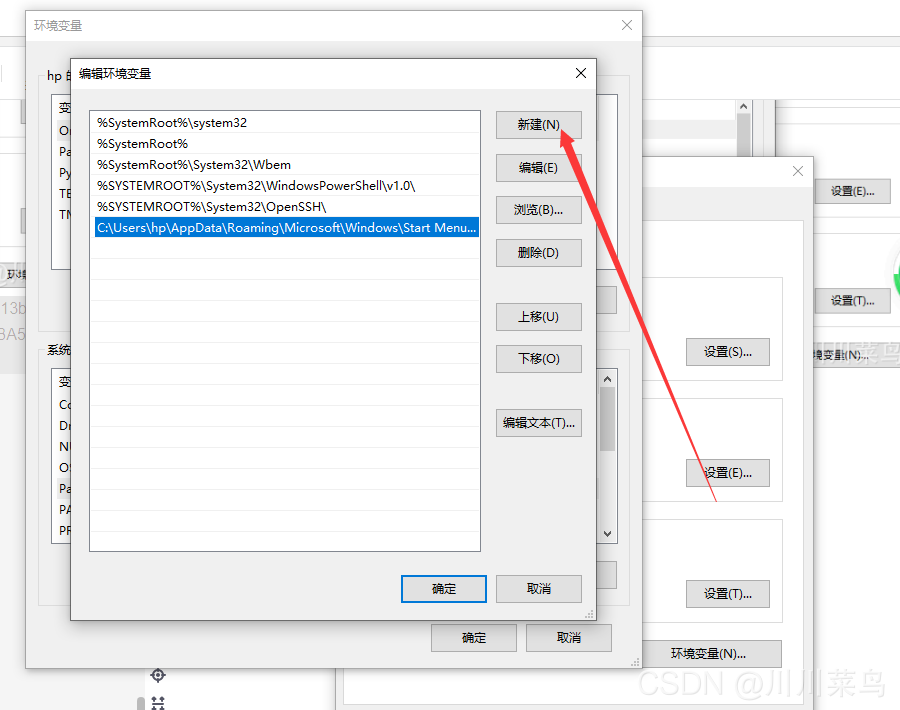
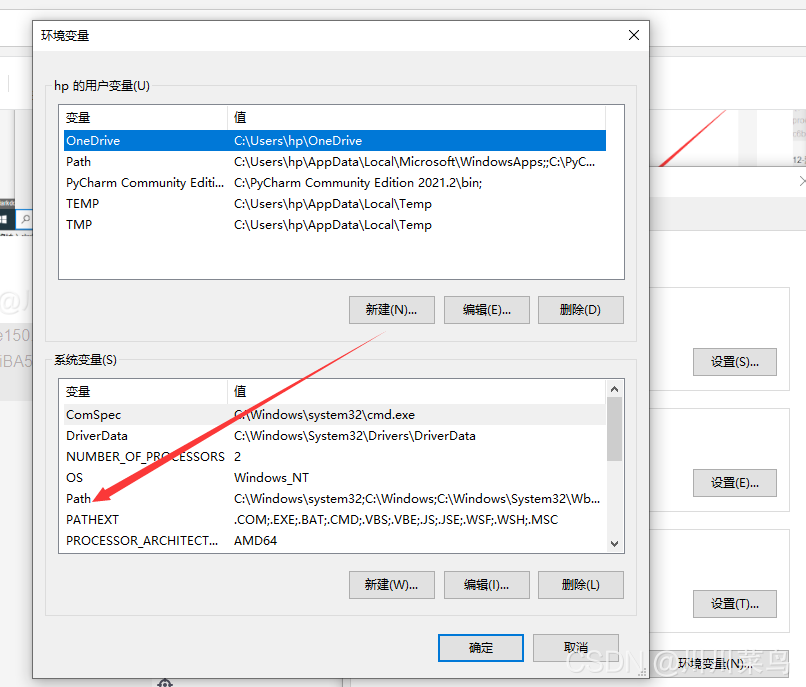
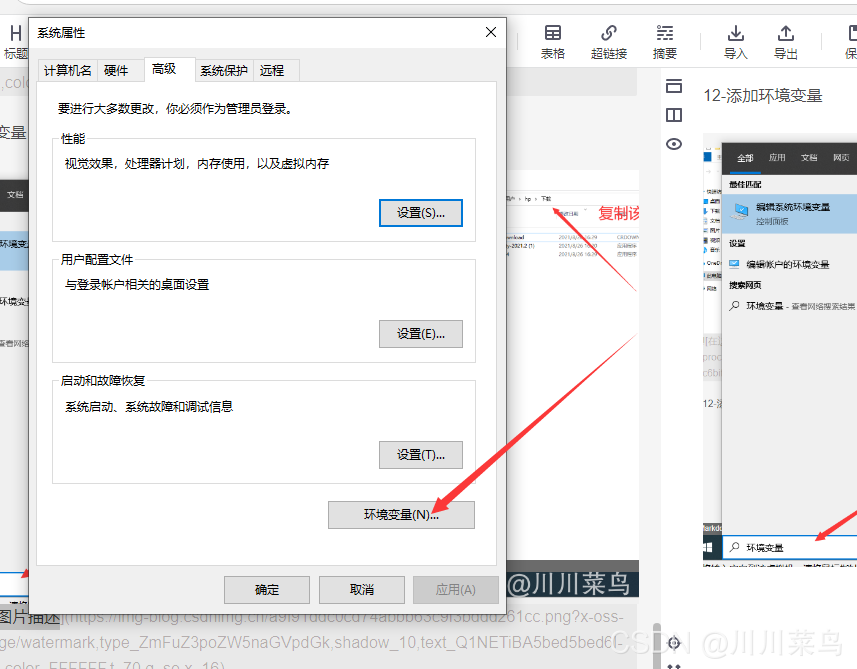
https://www.python.org/





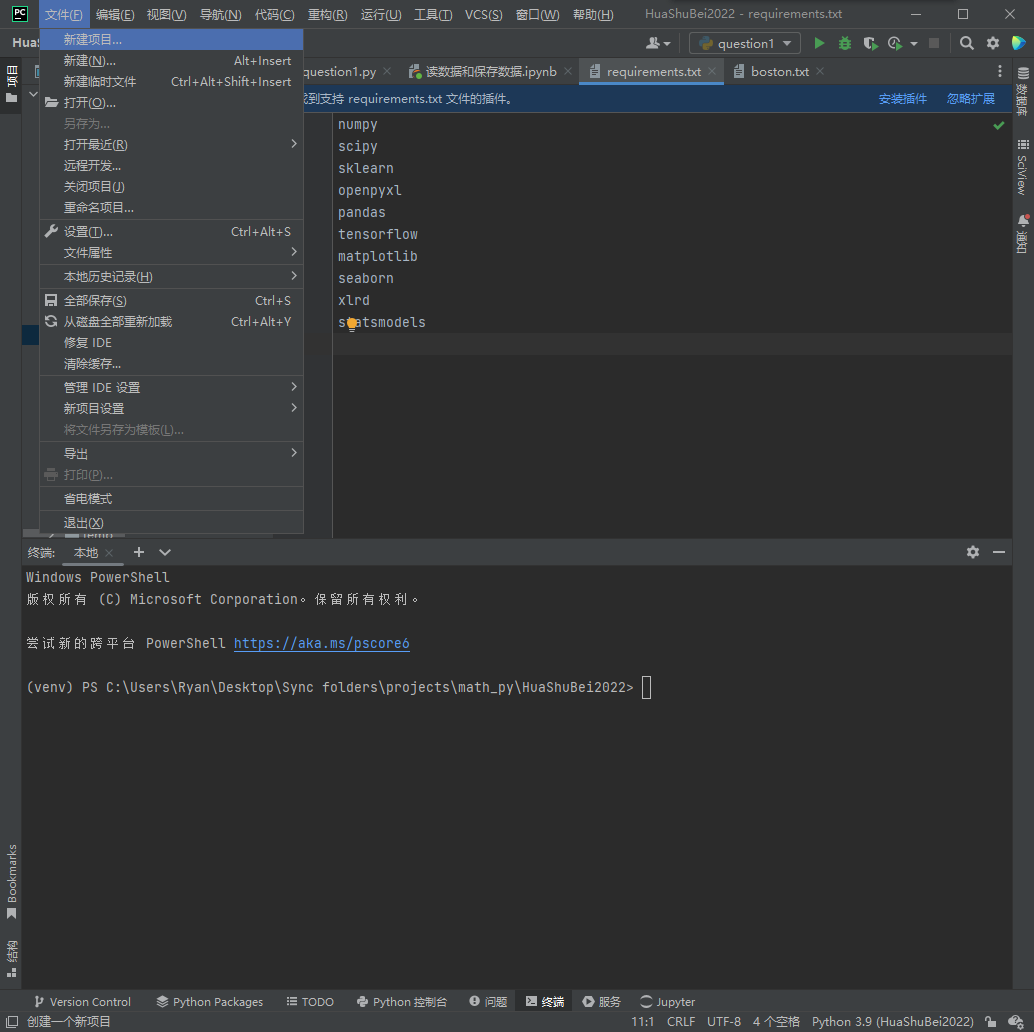


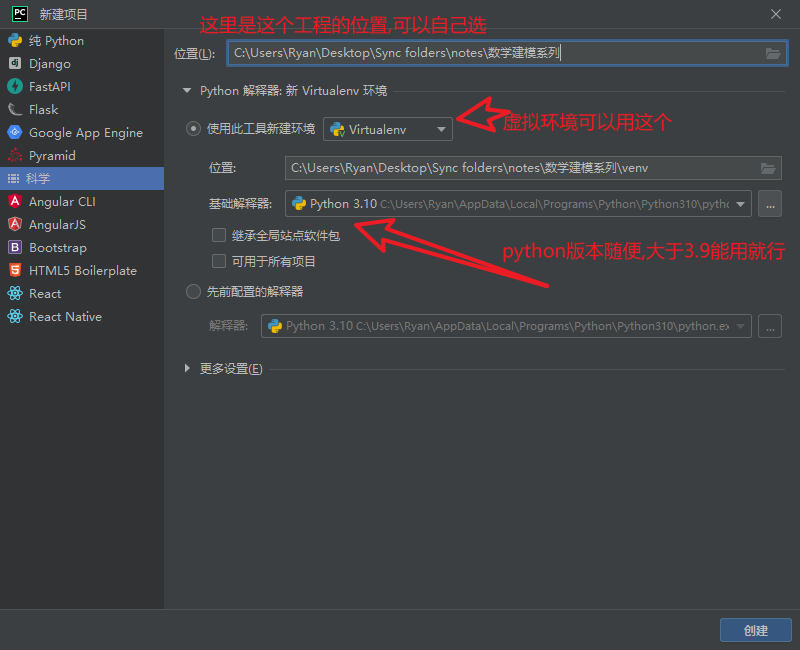




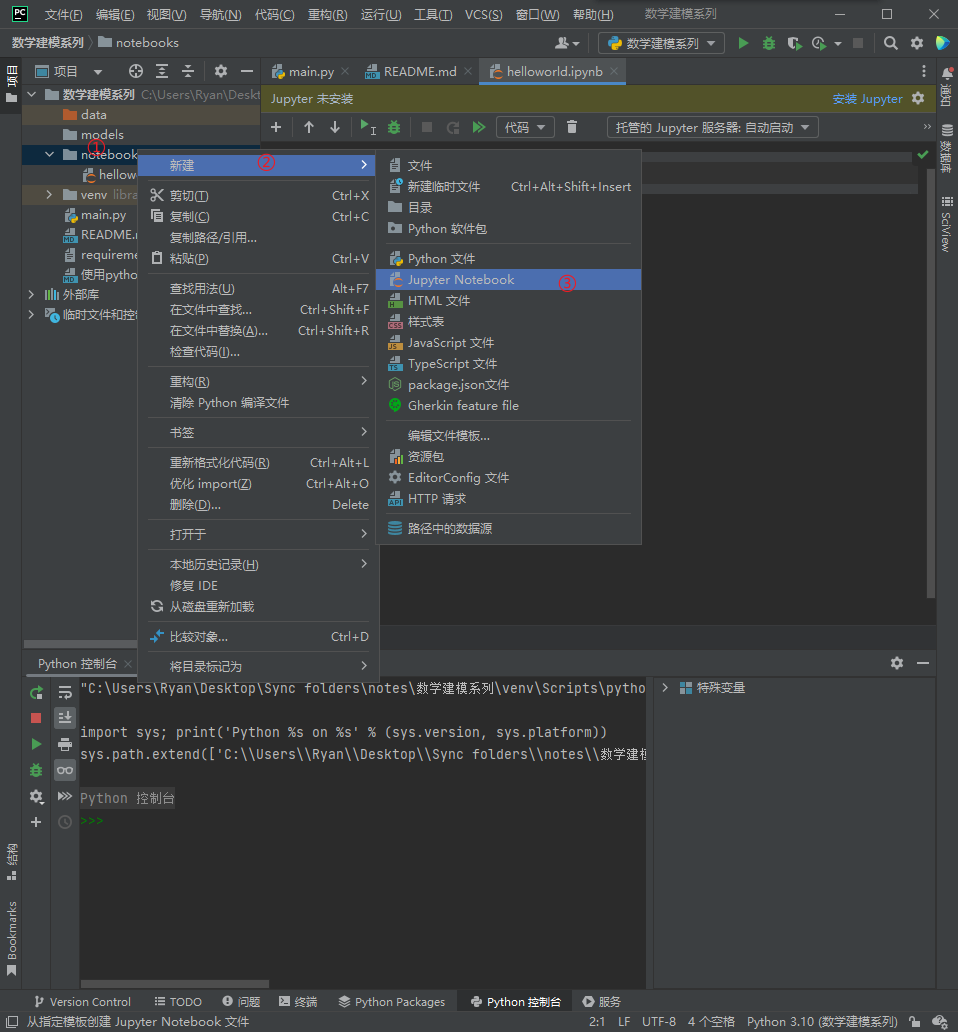
## 三、进行环境部署

1. 使用pycharm新建一个工程



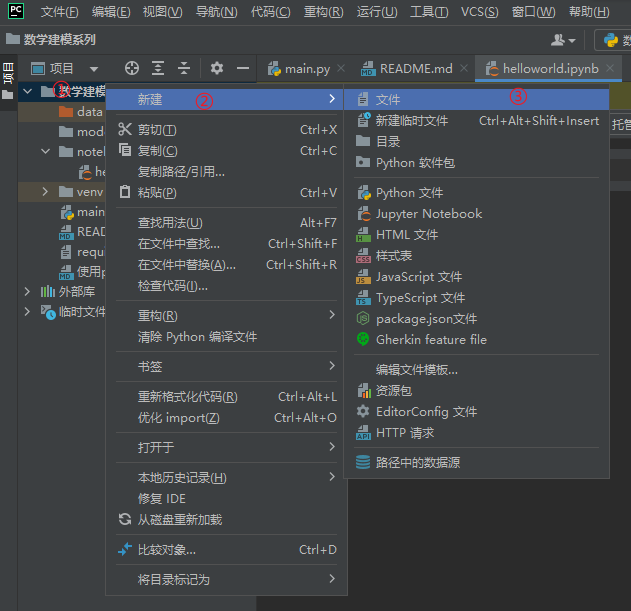


新建一个jupyter notebook文件,名字随便起



文件新建好了之后会看到这么个提示: jupyter没安装

新建一个requirements.txt 文件,在requirements.txt 文件里





里边的内容如下(复制粘贴保存即可).

Jupyter

numpy

scipy

sklearn

openpyxl

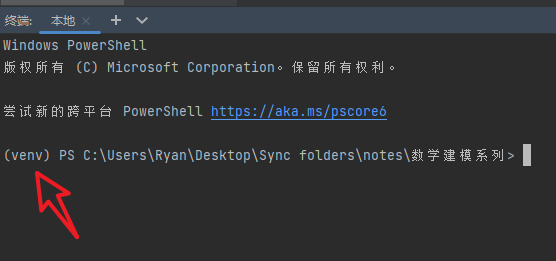
pandas

matplotlib

seaborn

xlrd

点击左下角终端查看是否有（venv）字样，如果有说明已经进入虚拟环境



**打开终端,输入以下内容即可安装所需要的主流的python的包**

**pip install -r requirements.txt**

## 四、运行代码

## 复制代码

import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
  
# Prior probability values  
prior\_probs = [0.3758, 0.3758, 0.2484]  
  
# 读取文件数据  
filename = 'C:/Users/金煜/PycharmProjects/scientificProject2/file.xlsx'  
data = pd.read\_excel(filename, usecols=[11, 12, 13, 14])  
  
# 检查数据列名  
if 'Handedness' in data.columns:  
 data = data.drop('Handedness', axis=1)  
  
# 将空值替换为0  
data = data.fillna(0)  
  
# 使用贝叶斯公式  
counts = np.zeros((data.shape[0], 3))  
for i in range(data.shape[0]):  
 if data.iloc[i, 0] != 0 and data.iloc[i, 1] != 0 and data.iloc[i, 2] != 0 and data.iloc[i, 3] != 0:  
 counts[i, 0] = (data.iloc[i, 0] / data.iloc[i, 3]) \* prior\_probs[0]  
 counts[i, 1] = (data.iloc[i, 1] / data.iloc[i, 3]) \* prior\_probs[1]  
 counts[i, 2] = (data.iloc[i, 2] / data.iloc[i, 3]) \* prior\_probs[2]  
# 删除所有特征值都是0的行  
counts = counts[~np.all(counts == 0, axis=1)]  
# 计算后验概率  
posterior\_probs = counts / counts.sum(axis=1)[:, None]  
  
# 输出先验概率值  
print('先验概率值:\n', prior\_probs)  
  
# 输出后验概率值  
print('贝叶斯公式输出的值:\n', posterior\_probs)  
  
# 输出特征的概率值  
print('特征的概率值:\n', counts)  
  
# 绘制柱状图  
fig, ax = plt.subplots()  
index = np.arange(3)  
bar\_width = 0.35  
opacity = 0.8  
  
rects1 = ax.bar(index, counts.mean(axis=0), bar\_width,  
 alpha=opacity, color='b',  
 label='Mean probability')  
  
rects2 = ax.bar(index + bar\_width, counts.max(axis=0), bar\_width,  
 alpha=opacity, color='g',  
 label='Max probability')  
  
# 添加先验概率柱状图  
rects3 = ax.bar(index, prior\_probs, bar\_width,  
 alpha=opacity, color='r',  
 label='Prior probability')  
  
ax.set\_xlabel('Feature')  
ax.set\_ylabel('Probability')  
ax.set\_title('Probability of Each Feature')  
ax.set\_xticks(index + bar\_width / 2)  
ax.set\_xticklabels(('UPDRS\_ON', 'UPDRS\_OFF', 'Yrs Since Diagnosis'))  
ax.legend()  
  
plt.tight\_layout()  
plt.savefig('概率柱状图.png') # 保存图像为名为 概率柱状图.png 的文件  
plt.show()  
  
# 绘制折线图  
plt.plot(counts[:, 0], label='UPDRS\_ON')  
plt.plot(counts[:, 1], label='UPDRS\_OFF')  
plt.plot(counts[:, 2], label='Yrs Since Diagnosis')  
  
# 添加先验概率折线  
# plt.plot(prior\_probs, label='Prior probability', linestyle='--')  
  
plt.legend()  
plt.title('Bayesian Formula Output')  
plt.savefig('贝叶斯公式.png') # 保存图像为名为贝叶斯公式.png 的文件  
plt.show()

plt.show()

## 注意：此行代码是读取数据来源excel文件，打开途径请自己改动

## 注意：此行代码是读取数据来源excel文件，打开途径请自己改动

## 运行此代码

## 

## 得出利用贝叶斯公式计算3个变量的二维图像

## 和一张带有先验概率和特征概率的柱形图和贝叶斯公式输出的值，特征的概率值:

## 

三、模型有效性验证

具体的模型有效性验证包括以下几个步骤：

1. 先验概率值：输出先验概率值的语句 **print('先验概率值:\n', prior\_probs)** 可以用于验证先验概率是否正确设置。先验概率是在没有观测数据的情况下对事件发生概率的主观估计，可以通过领域知识或经验来确定。
2. 贝叶斯公式输出的值：通过输出 **print('贝叶斯公式输出的值:\n', posterior\_probs)** 可以得到贝叶斯公式计算得到的后验概率值。这些后验概率值反映了在给定观测数据的情况下，每个特征发生的概率。可以验证计算过程是否正确，并进一步分析后验概率的结果。
3. 特征的概率值：通过输出 **print('特征的概率值:\n', counts)** 可以获得特征的概率值。这些概率值是在贝叶斯公式计算过程中基于观测数据和先验概率计算得出的。可以分析特征的概率分布，观察不同特征的值对应的概率大小，进一步理解特征与结果之间的关系。
4. 概率柱状图：代码中的柱状图绘制部分可以通过 **plt.show()** 来展示。这个图可以用于可视化先验概率、特征的概率值等信息，直观地呈现概率分布情况。可以通过观察柱状图来比较不同特征的概率大小、不同特征的均值概率和最大概率等。
5. 贝叶斯公式折线图：通过代码中的折线图绘制部分，可以将特征的概率值可视化为折线图。可以观察特征的概率值随着观测数据的变化而变化的趋势，进一步分析特征与结果之间的关系。

四、结果分析

一．统计建模与分析

(- )先验概率.  
先验概率是指不考虑任何因素时根据历史资料获得的患某疾病的可能性，即未考虑其他任何因素时各人群前驱期帕金森病的患病率。我国65岁以上人群帕金森病患病率为1 700/100 000[6],与欧美人群无差异[6,7,8], 因此，可直接采用MDS的先验概率。根据各年龄段帕金森病患病率,以帕金森病平均前驱期10年计算，可得各年龄段帕金森病前驱期的先验概率:以5年为一个年龄组，50~ 54岁年龄段的先验概率为0.40%，55 ~ 59岁先验概率为0.75%，60 ~ 64岁先验概率为1.25%, 65 ~ 69岁先验概率为2.00%，70~ 74岁先验概率为2.50%, 75 ~ 79岁先验概率为3.50%，80岁及以上先验概率为4.00%。

根据《中国第七次人口普查得知》

65岁到69岁之间人口比例为：5.46%

70岁到79岁之间人口比例为：5.5%

80岁及80岁以上的人口比例为：2.54%

我们需要将每个年龄段的人口比例与其对应的先验概率相乘，并将它们相加。

首先，我们计算65岁到69岁之间的人口比例和先验概率的乘积： 5.46% \* 2.00% = 0.1092%

接下来，计算70岁到79岁之间的人口比例和先验概率的乘积： 5.5% \* 3.00% = 0.165%

然后，计算80岁及80岁以上的人口比例和先验概率的乘积： 2.54% \* 4.00% = 0.1016%

最后，将这三个结果相加： 0.1092% + 0.165% + 0.1016% = 0.3758%

因此，65岁以上年龄段的先验概率为0.3758%。

二、数值结果分析

第一个值是UPDRS\_ON变量的后验概率。帕金森疾病患者的UPDRS\_ON评分通常用于衡量他们在药物治疗下的症状严重程度。这个后验概率值表示在给定先验概率和观测数据的情况下，UPDRS\_ON评分处于某个特定程度的概率分布。较高的后验概率值可能表示患者在药物治疗下症状更严重。

第二个值是UPDRS\_OFF变量的后验概率。UPDRS\_OFF评分用于衡量帕金森病患者在没有药物治疗下的症状严重程度。这个后验概率值表示在给定先验概率和观测数据的情况下，UPDRS\_OFF评分处于某个特定程度的概率分布。较高的后验概率值可能表示患者在没有药物治疗下症状更严重。

第三个值是Yrs Since Diagnosis变量的后验概率。Yrs Since Diagnosis表示自帕金森病诊断以来的年限。这个后验概率值表示在给定先验概率和观测数据的情况下，患者被诊断为帕金森病已经过了某个特定的年限的概率分布。较高的后验概率值可能表示患者已经被诊断为帕金森病已经较长时间。

综合起来，这些后验概率值可以提供关于患者症状严重程度和帕金森病发展时间的概率信息。通过进一步分析和解释这些概率值，我们可以更好地理解帕金森疾病在不同变量上的表现和特征，并为疾病管理和治疗提供一定的参考依据。

根据贝叶斯建模输出的图像和数值我们可以得出以下结论：

1. 药物治疗效果：UPDRS\_ON评分通常是在帕金森病人在药物治疗下的症状严重程度评分，而UPDRS\_OFF评分是在没有药物治疗下的评分。UPDRS\_ON评分普遍高于UPDRS\_OFF评分，表示药物治疗对症状的改善起到了一定的效果，使得患者在药物治疗下的症状更轻。
2. 变量差异：UPDRS\_ON和UPDRS\_OFF评分反映了帕金森病人在不同状态下的症状表现。UPDRS\_ON的值普遍大于UPDRS\_OFF的值，这表示在药物治疗下，患者的症状得到了一定程度的缓解，表现为UPDRS\_ON评分较低。而在没有药物治疗下，症状可能会加重，导致UPDRS\_OFF评分较高。
3. 治疗效应：帕金森疾病的治疗可能会影响患者的症状严重程度。如果UPDRS\_ON评分普遍高于UPDRS\_OFF评分，可能暗示了治疗的有效性，药物可以改善患者的运动功能和症状。



