

Python机器学习——数学建模与分析

初识Python与机器学习

微信公众号 / B站号：数学建模老哥



数学建模老哥

目录

CONTENTS

01. 初识Python

02. 初识机器学习

03. Numpy快速入门

初识Python

- 数学建模为什么逐渐开始流行Python？
- Matlab逐渐被Python淘汰吗？
- 数学建模用Matlab好还是Python好（当然是两种编程都是小白的情况下）？
- “Matlab创新奖” 暗示更倾向于Matlab？



数学建模为什么逐渐开始流行Python？

- Python + Github = 编程军火库

Python数值运算： Numpy

Python数据分析： Pandas

Python时序分析： FbProphet

Python可视化工具： matplotlib

Python地图分析： folium

Python web开发： Flask

Python机器学习： Scikit-Learn XGBoost

Python深度学习： TensorFlow PyTorch

人生苦味百般味，Python编程不怕累



数学建模为什么逐渐开始流行Python？

- 入手快：Python相较其他编程语言比较容易学习
- 颜值高：Python语言追求代码美观、简洁
- 效率高：Python有强大的类库
- 环境好：新兴产业的支持：人工智能、目标检测.....
- 应用广：数据分析，绘制图片，人工智能，机器学习，网络爬虫，Web开发.....



Matlab逐渐被Python淘汰吗？

- 首先，说代替是不存在的，matlab强大的工科支持，是不可能被替代的，比较没有任何意义！

比如：Simulink仿真，硬件支持（摄像头，激光雷达……），可靠性（Matlab会为它的算法负责任）

- 但是，Matlab市场被挤压得很惨

比较应该相同功能的比较：

Python：免费

Matlab：付费（部分学校被限制，但盗版完全可以应付数模）

Python：开源，可随意修改算法细节

Python：第三方生态比Matlab强

数学建模用Matlab好还是Python好（当然是两种编程都是小白的情况下）？

- 1. 建模时间短，只有3天，代码修改和借鉴的能力很重要（Python都是开源）
- 2. 如果涉及到数据清洗、数据挖掘、预测、图像处理、机器学习、深度学习相关问题，使用python会更好。
- 3. 如果涉及到仿真、有限元可以考虑Matlab

“Matlab创新奖”暗示更倾向于Matlab？

- 数学建模比赛并不会有所谓“官方指定”的软件！
- 国赛中设有“Matlab创新奖”，让很多人误认为数学建模比赛必须使用Matlab。
- 这其实是误解，迈斯沃克公司（Matlab母公司）是国赛赞助商，为了推广其软件，所以才在比赛中设立奖项，奖励一篇使用Matlab的优秀论文。
- 但是，如果你通篇没有使用Matlab，除了不能参评这个奖项外，对你论文获奖没有任何影响。
(Matlab创新奖每年数万个队伍中仅有一项，获奖概率可以忽略，故影响不大。)

Python安装（Anaconda安装）

- Anaconda包括Conda、Python以及一大堆安装好的工具包，比如：numpy、pandas等
- 1）包含conda：conda是一个环境管理器，其功能依靠conda包来实现
- 2）安装大量工具包：Anaconda会自动安装一个基本的python
- 3）可以创建使用和管理多个不同的Python版本



Python安装（Anaconda安装）

第一步：[Free Download | Anaconda](#)（官网）

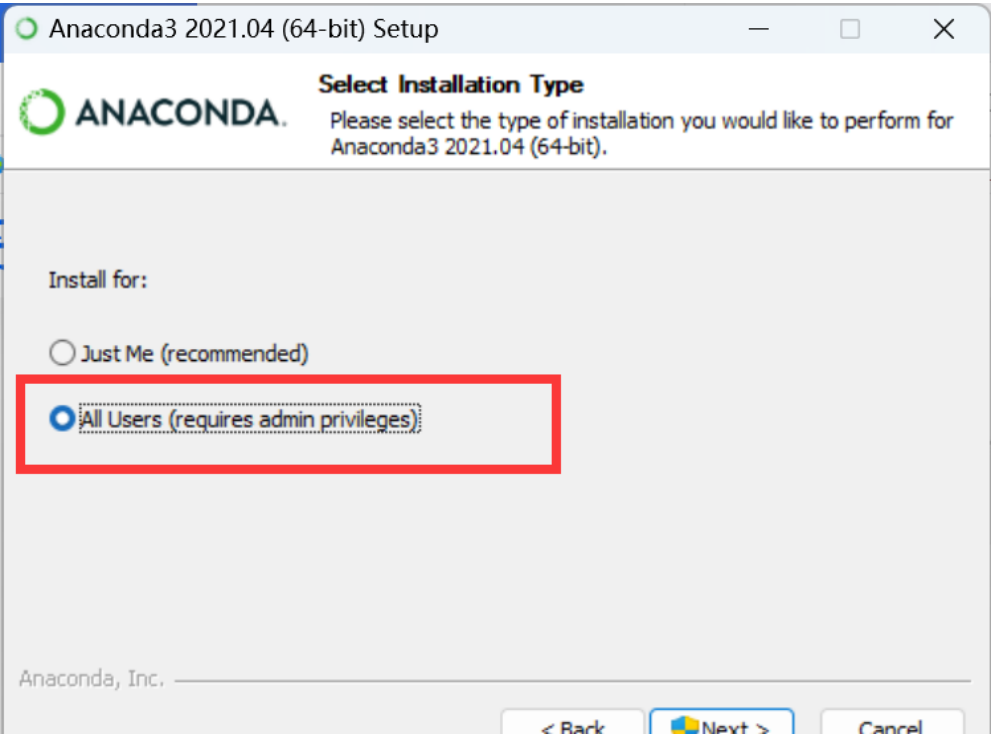


Python安装（Anaconda安装）

第二步：

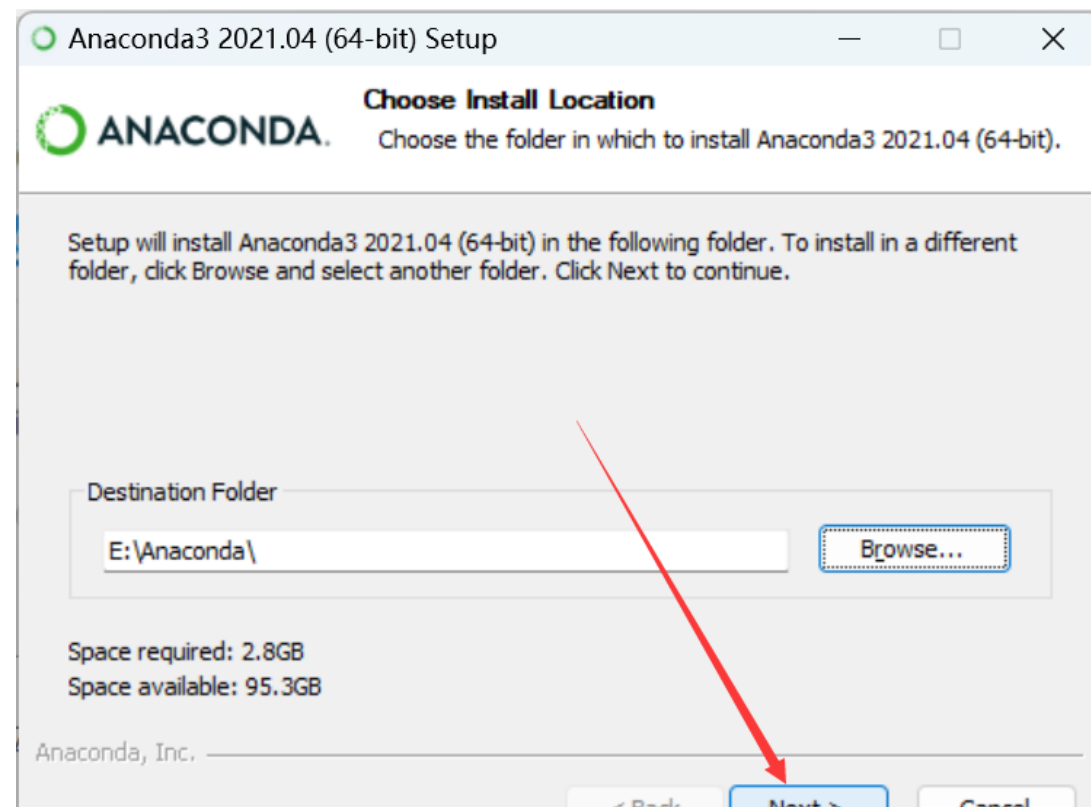
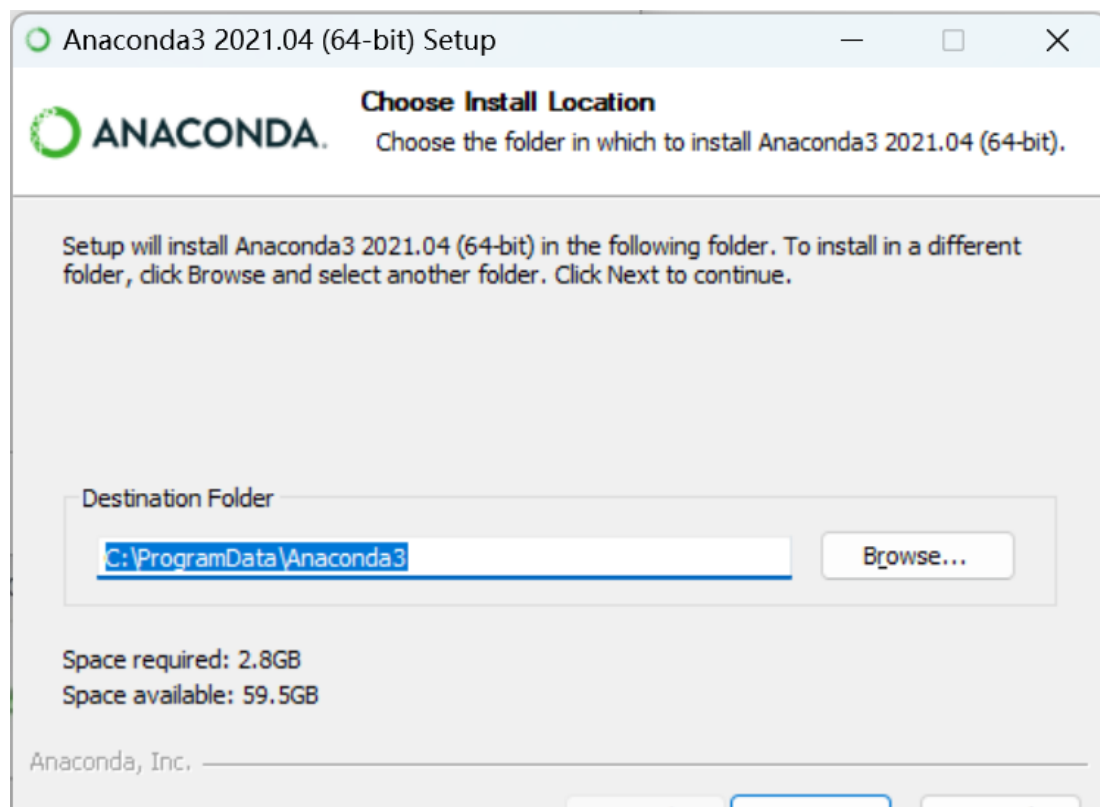
名称	修改日期	类型	大小
今天 (1)			
Anaconda3-2023.03-Windows-x86_64.exe	2023/3/26 15:36	应用程序	804,860 KB

第三， 四步



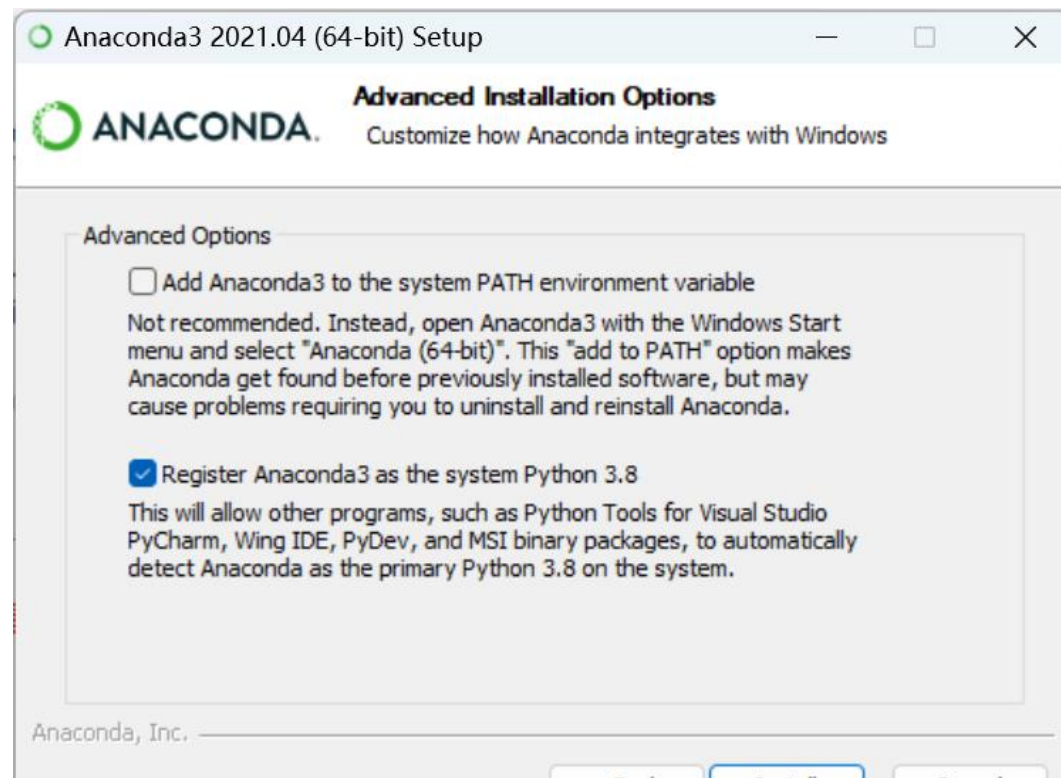
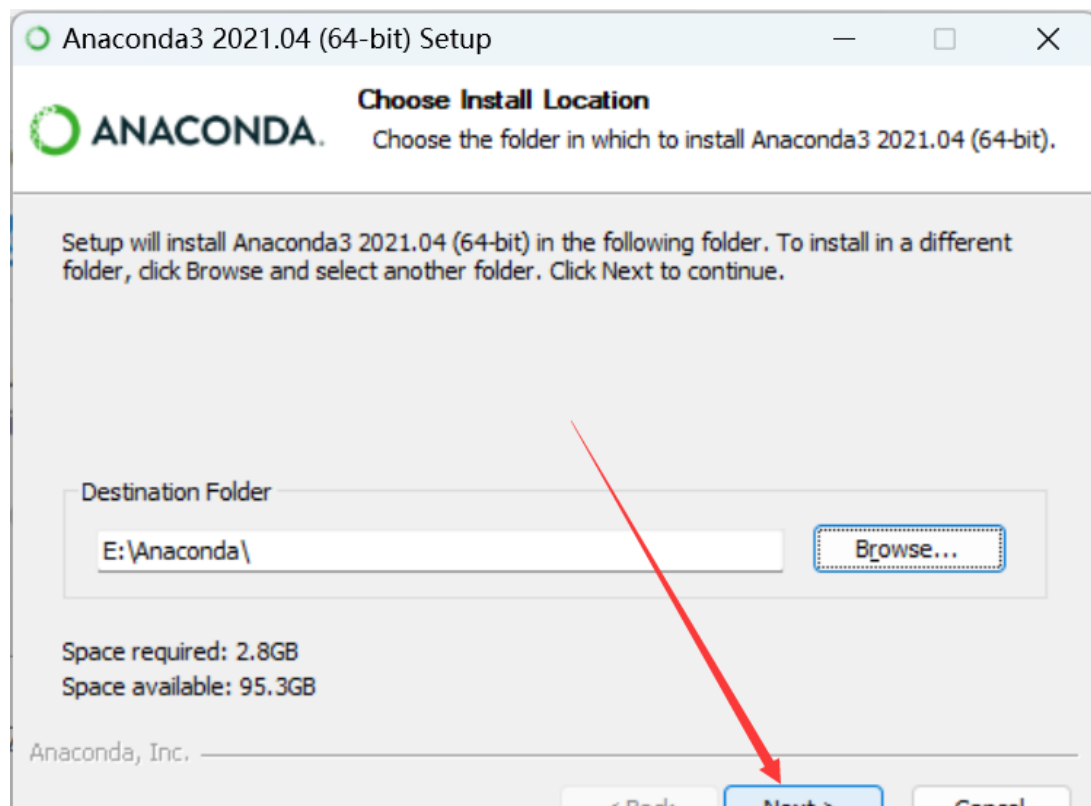
Python安装（Anaconda安装）

第五、六步：



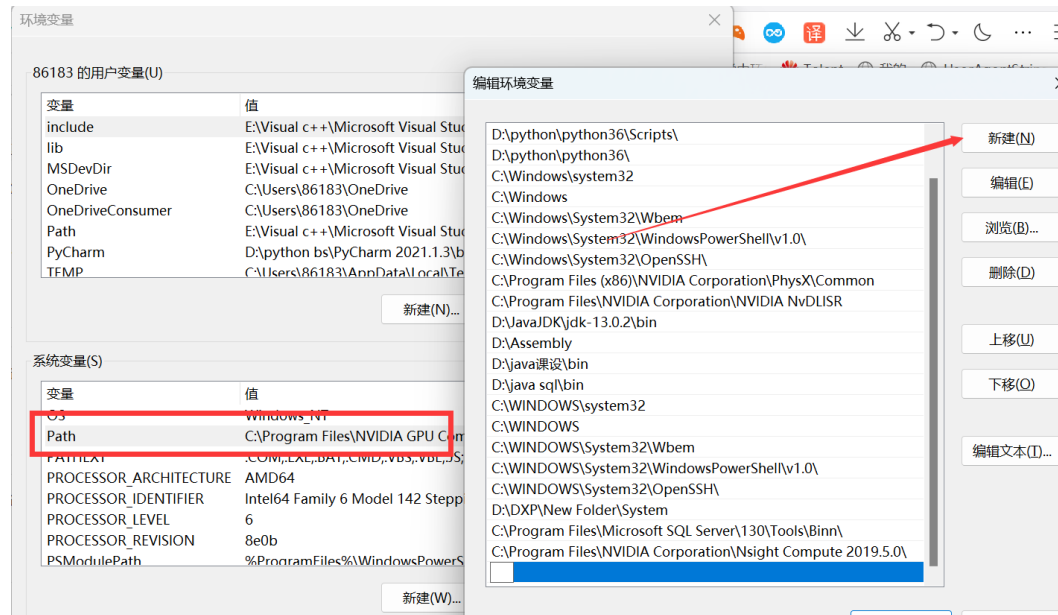
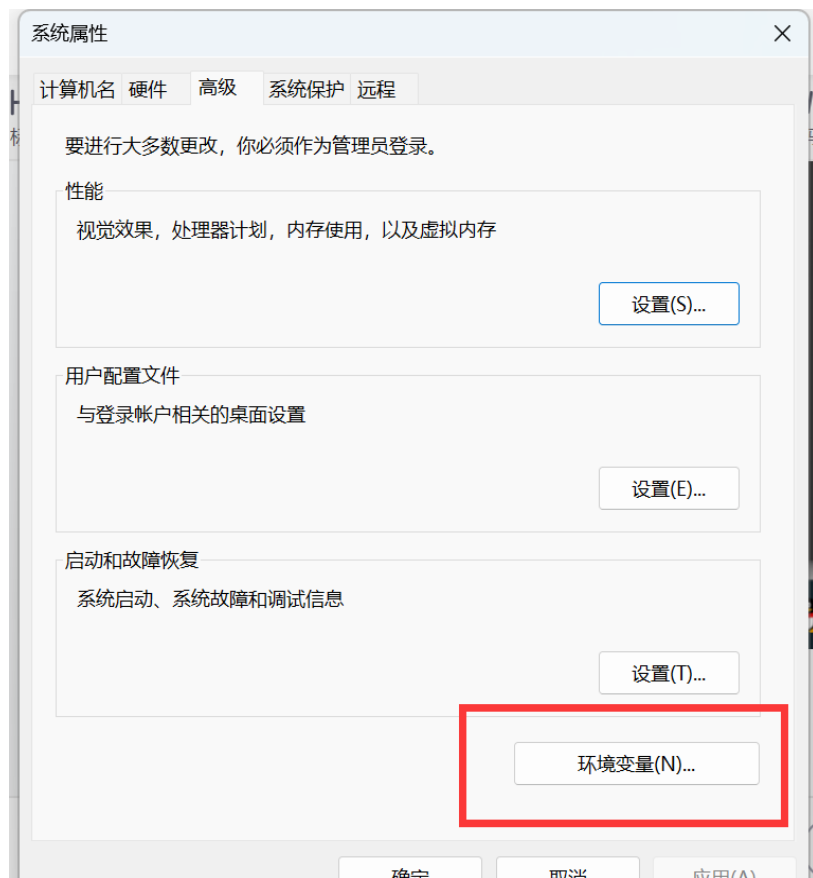
Python安装（Anaconda安装）

第五、六步：



Python安装（Anaconda安装）

第七步：

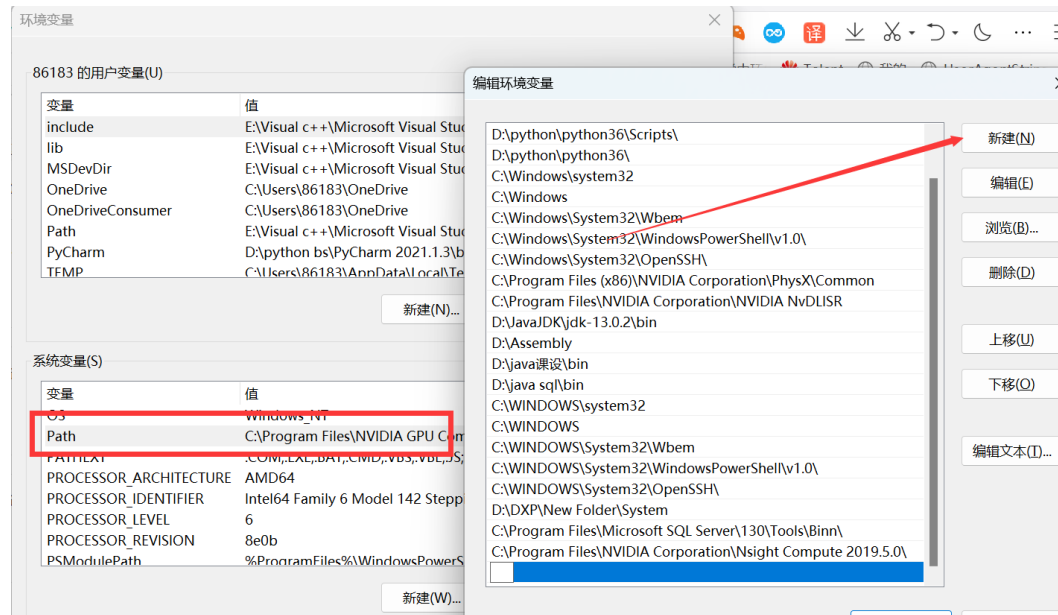
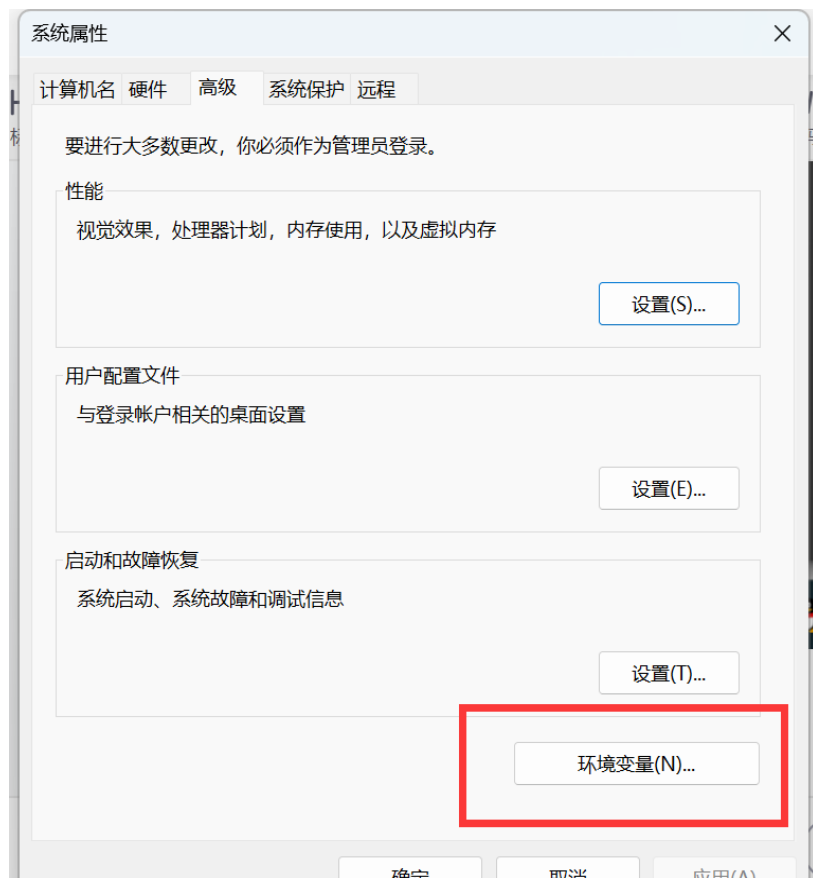


E:\Anaconda
E:\Anaconda\Scripts
E:\Anaconda\Library\mingw-w64\bin
E:\Anaconda\Library\usr\bin
E:\Anaconda\Library\bin



Python安装（Anaconda安装）

第七步：



E:\Anaconda
E:\Anaconda\Scripts
E:\Anaconda\Library\mingw-w64\bin
E:\Anaconda\Library\usr\bin
E:\Anaconda\Library\bin



初识机器学习——什么是机器学习？

- 机器学习（Machine Learning, ML）是一门多领域的交叉学科，涉及概率论、统计学、线性代数、算法等多门学科。它专门研究计算机如何模拟和学习人的行为，以获取新的知识和技能，重新组织已有的知识结构使之不断完善自身的性能。
- 机器学习已经有了十分广泛的应用，例如：数据挖掘、计算机视觉、自然语言处理、生物特征识别、医学诊断、检查信用卡欺诈、证券市场分析、DNA测序……

如何解释数据，处理数据，从中抽取价值，展示和交流数据结果，在未来十年将是最重要的职业技能，甚至是大学，中学，小学的学生也必须具备的技能，因为我们每时每刻都在接触大量的免费信息，如何理解数据，从中抽取有价值的信息才是其中的关键。

——谷歌公司的首席经济学家Hal Varian



初识机器学习——什么是机器学习？

- 下面通过以构建“西瓜数据集”为例，我们将接触机器学习涉及的常用术语。

西瓜数据集				
编号	色泽	根蒂	敲声	好瓜
1	青绿	蜷缩	浊响	是
2	乌黑	蜷缩	浊响	是
3	青绿	硬挺	清脆	否
4	乌黑	稍蜷	沉闷	否

数据集：这组记录的集合

样本：其中每条记录是关于一个事件或对象（这里是一个西瓜）的描述

特征：反映对象在某方面的性质的事项，例如“**色泽**”“**根蒂**”“**敲声**”

特征值：属性上的取值，例如“**青绿**”“**乌黑**”

样本空间：由特征张成的空间，例如我们把“**色泽**”“**根蒂**”“**敲声**”作为三个坐标轴，每个西瓜都可在这个空间中找到自己的坐标位置。

特征向量：空间中的每个点都对应一个坐标向量

初识机器学习

● 1. 模型

模型这一词语将会贯穿整个教程的始末，它是机器学习中的**核心概念**。你可以把它看做一个“**魔法盒**”，你向它**许愿**（输入数据），它就会帮你**实现愿望**（输出预测结果）。整个机器学习的过程都将围绕模型展开，训练出一个最优质的“魔法盒”，它可以尽量精准的实现你许的“愿望”，这就是机器学习的目标。

● 2. 数据集

数据集，从字面意思很容易理解，它表示一个**承载数据**的**集合**，如果说“模型”是“魔法盒”的话，那么数据集就是负责给它充能的“能量电池”，简单地说，如果缺少了数据集，那么模型就没有存在的意义了。数据集可划分为“**训练集**”和“**测试集**”，它们分别在机器学习的“**训练阶段**”和“**预测输出阶段**”起着重要的作用。



初识机器学习

● 3. 训练集&假设

从数据中学得的模型，称为“学习”或“训练”。这个过程通过执行某个学习算法完成，训练过程中使用的数据称为“**训练数据**”，其中每一个样本称为“**训练样本**”，由“训练样本”组成的集合称为“训练集”。学得模型对应了关于数据的某种潜在的规律，因此亦称“假设”，假设可以理解成“模型”；这种潜在的规律自身，则称为“真相”。学习的过程就是为了找出或逼近真相。

● 4. 测试集

在获得“训练模型”后，我们还需要知道用该模型来预测其他情况的结果的**效果**好不好，所以需要引入“测试集”，如果该模型也能够很好的预测出“测试集”的结果，那么我们可以认为“训练模型”非常接近“真相”。



● 5. 标记信息

如上表，例如，其中一个样本为“（（色泽=青绿；根蒂=蜷缩；敲声=浊响），好瓜）”，这里关于**样本结果**的信息，例如“**好瓜**”，称为“**标记**”；拥有标记信息的示例，则称为“**样例**”。注意：分类和回归问题需要使用带“标记”的数据，聚类使用的数据集一般不拥有标记信息。

● 6. 分类&回归

若我们欲预测的结果是离散值（即标记是离散的），例如“**好瓜**”“**坏瓜**”，此类学习任务称为“**分类**”；若欲预测的是连续值（即标记是连续的），例如西瓜的成熟度0.95，0.37. 此类学习任务称为“**回归**”。

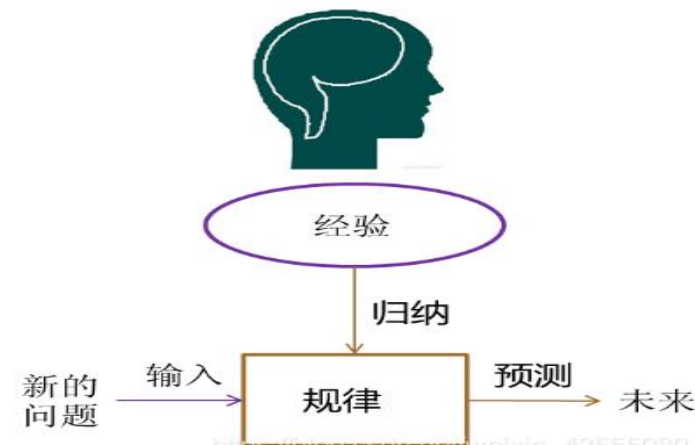
7. 聚类

我们还可以对西瓜做“**聚类**”，相关算法将自动将训练集中的西瓜分成若干组，每组称为一个“簇”。这些自动形成的簇可能对应一些潜在的划分，比如“本地瓜”“外地瓜”。需注意的是，在聚类学习中，像“本地瓜”“外地瓜”这些概念我们事先是不知道的，“簇”是计算机自动识别数据进行的划分，而且学习过程中使用的训练集样本不拥有标记信息。

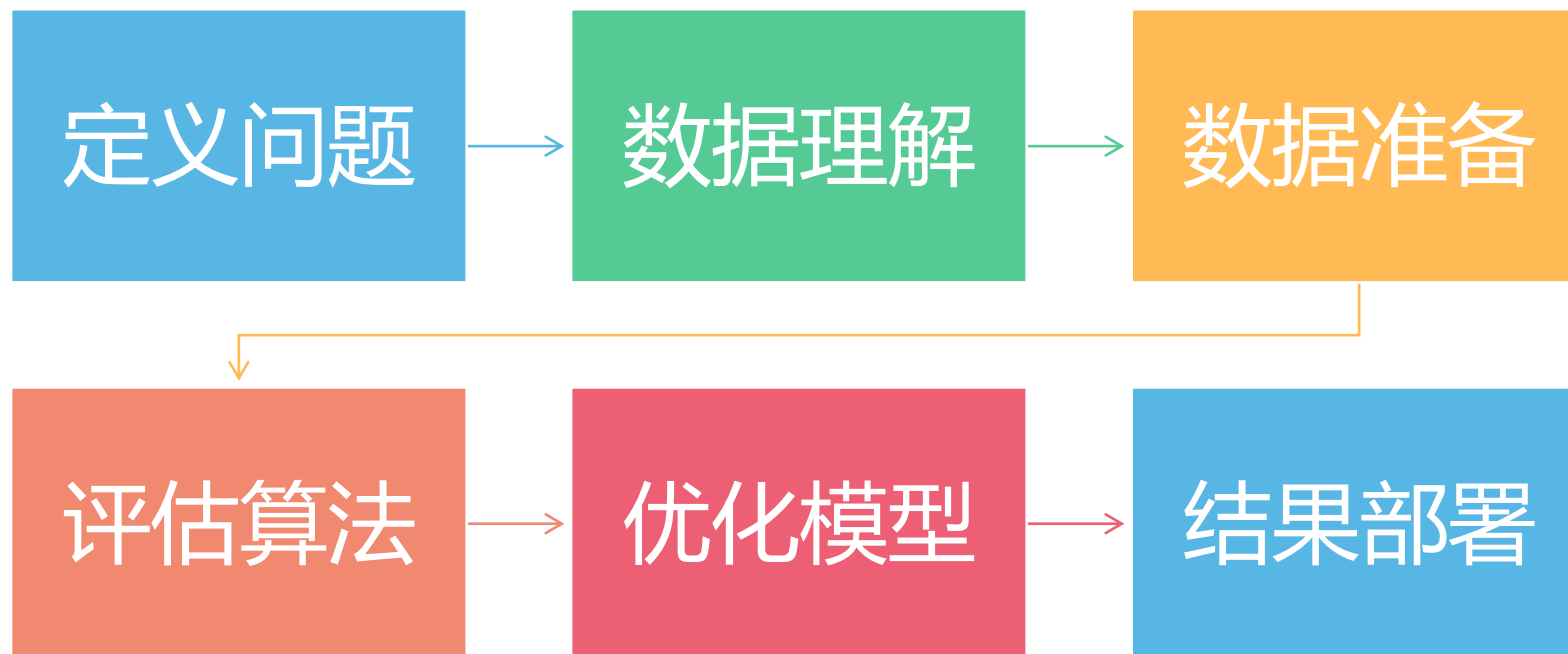


初识Python——机器学习分类

- 机器学习的算法分为两大类：**监督学习**、**无监督学习**
- 监督学习：在机器学习过程中提供对错指示。比如数据的结果部分为（0，1），通过算法让机器自己减少误差。这一类主要应用于**分类**和**回归**（Regression&Classify）。监督学习从给定的训练集中学习一个**目标函数**，当新的数据到来时，可以根据这个函数**预测结果**。监督学习要求包括输入和输出，也就是特征和目标。
- 非监督学习：归纳行学习，利用K方式建立中心，通过循环和递减运算减小误差达到分类的目的。



数学建模Python机器学习五大步骤



课件代码下载地址



关注公众号：“**数学建模老哥**” 回复“**课件**” 即可免费获取

更多模型、代码、优秀论文等请加QQ群：**1077734962**

更多资料请关注微信公众号：数学建模老哥



更多资料请关注公众号【老哥带你学数模】课件或代码请公众号回复“课件”，QQ群：720736320

THE END



数学建模老哥