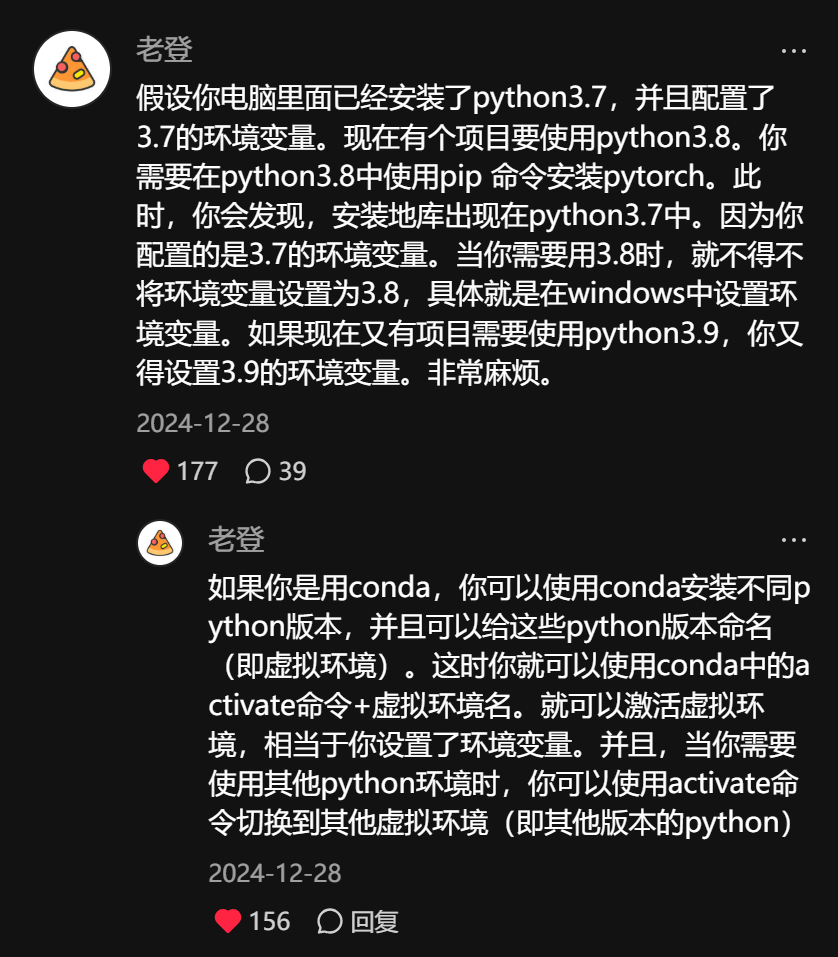
**一、配置环境**

我是用的anaconda，能帮你管理环境，快速切换虚拟环境，也能通过anaconda powershell prompt快速下载一些机器学习要用到的库和TensorFlow、Pytorch这些深度学习框架，非常方便。

下图是小红书上一位老哥解释的“anaconda是干嘛的”：



下载后进入anaconda powershell prompt，将下图代码一次输入一行并回车，就新建了一个环境，并在这个环境里安装了python和tensorflow (pytorch的操作类似，gpu版建议找ai要安装指令)。

这里我们就可以直观地看到anaconda的好处，就是可以帮你管理多个互相不干扰的环境，你可以在不同环境里装不同版本的python、框架、库，以适配不同的任务，环境之间不会互相污染。



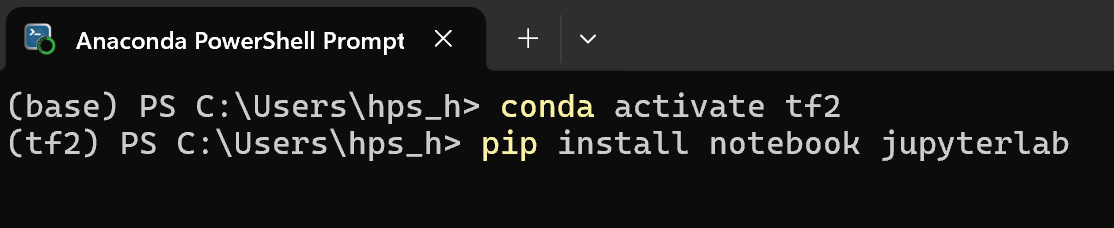


只需简单输入conda activate xxx(你设的环境名称)就可激活一个环境。

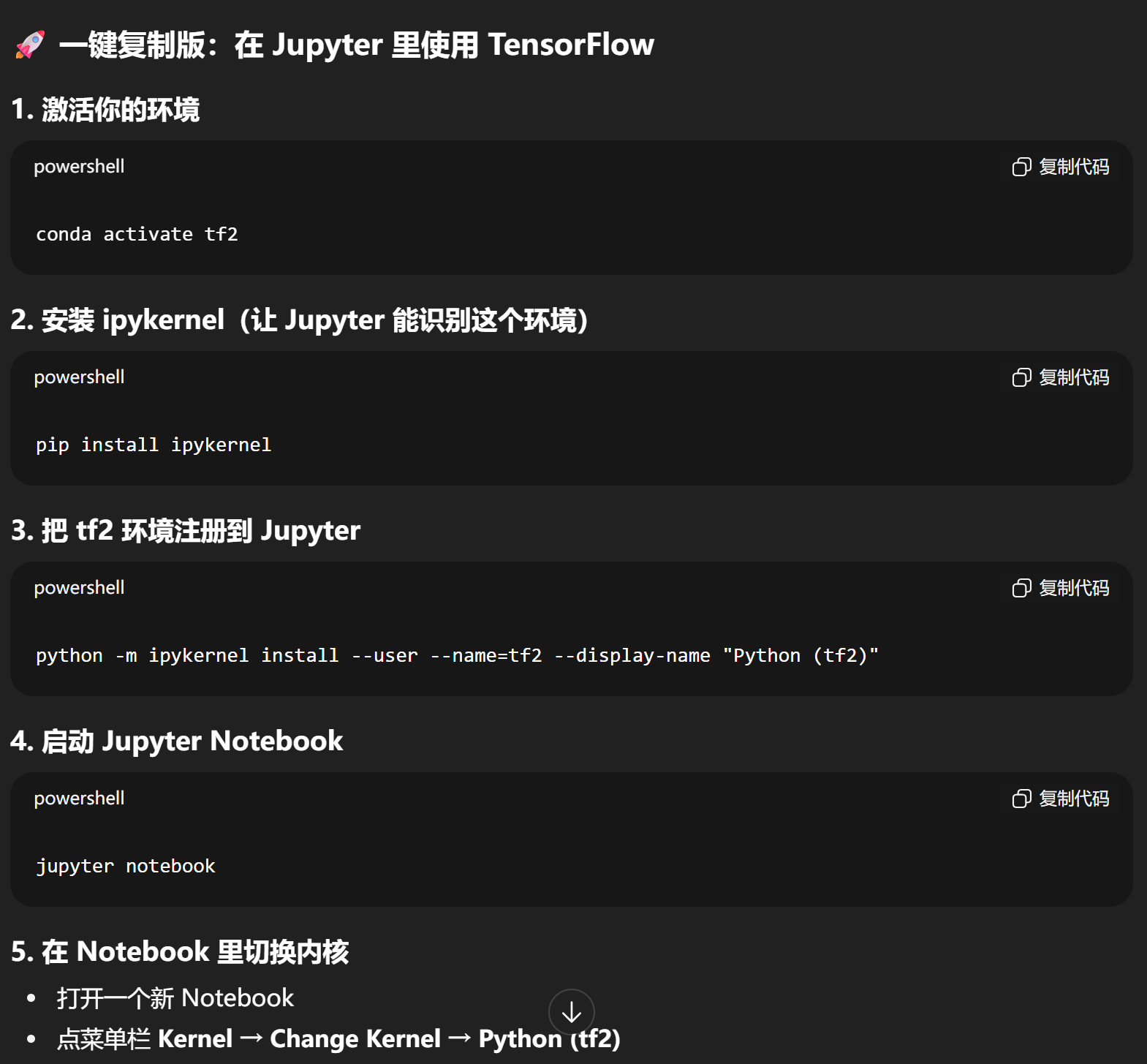


下载anaconda后里面就自带了jupyter notebook、pycharm这些用于python语言的编辑器了。我用的是gpt给我推荐的jupyter lab：

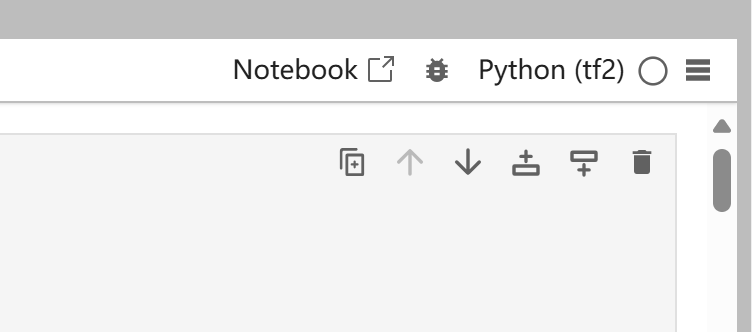




安装也很简单，在anaconda powershell prompt里输入activate指令，回车激活环境，再输入安装指令pip install notebook jupyterlab再回车就可以了。



这样再进入jupyter notebook或jupyter lab右上角就有tf2的kernel了：



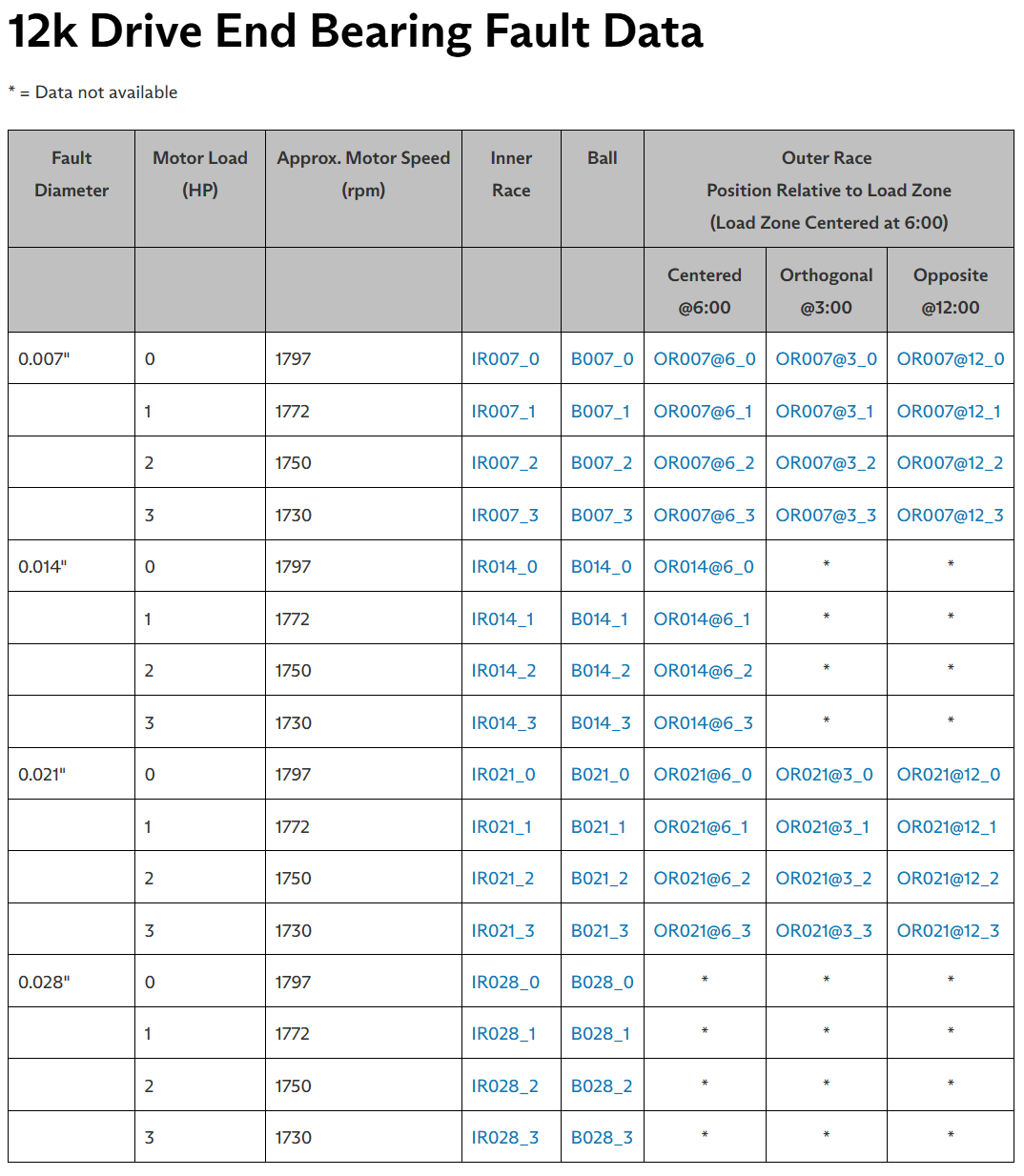
这只是我的配置过程，如果你是从下载anaconda开始的话，建议找个ai一步步跟着走，很多地方我这里没有写那么详细。当然也可以不用anaconda，这只是我的推荐。

**二、代码实现**

我喂给gpt的指令：

我装好了anaconda和jupyter notebook、jupyter lab、Tensorflow以及一些常用科学计算库，现在要开始用CWRU数据集训练一个ResNet1D模型了。我们下载的CWRU数据集有4个大文件夹：正常数据（Normal Baseline Data）、12kHz采样率下驱动端（12kDE）故障数据、48kHz采样率下驱动端（48kDE）故障数据、12kHz采样率下风扇端故障数据。我在网上看人说大家一般都用12kDE 0hp的数据做验证。如图，12kDE里分0.007、0.014、0.021、0.028这4种故障尺寸，又有内圈（inner race）、滚动体（ball）和外圈（outer race）这3种故障位置，但因为0.028的只有内圈数据，所以一般只用前三种故障尺寸。外圈故障中也分不同的方向，但由于0.014尺寸下缺了两组数据，所以也就只用外圈故障里的centered@6:00的数据了。这样3\*3就是把12kDE里的数据分成9类，再加上Normal Baseline Data文件夹里的正常数据，就是10类，我看也是很多人用CWRU是去做十分类。总的来说，就是要在0hp的工况下，模型要能够将数据分成 0.007内圈、0.007滚珠、0.007外圈、0.014内圈、0.014滚珠、0.014外圈、0.021内圈、0.021滚珠、0.021外圈、正常 这10类，然后显示一个分类准确率。我这个想法是否正确？如果正确要怎么通过代码实现呢？我知道可能要先读取数据、分割数据集，再送进模型分类，但具体的流程和代码我不会

并附图：



gpt的反馈：



我认为把情况说清楚，gpt才好写代码。毕竟CWRU轴承故障的分类，有四分类的，有十分类的，有用12kDE的，有用48kDE的，还有DE、FE都用的，不说清楚gpt怎么知道你想怎么分类呢？就算它随便写一个，你不知道自己分的是啥你怎么写文件路径呢？所以无论用哪种还是跟ai确认好。

具体分类的方法我是参考的咱们群里的两篇公众号文章：

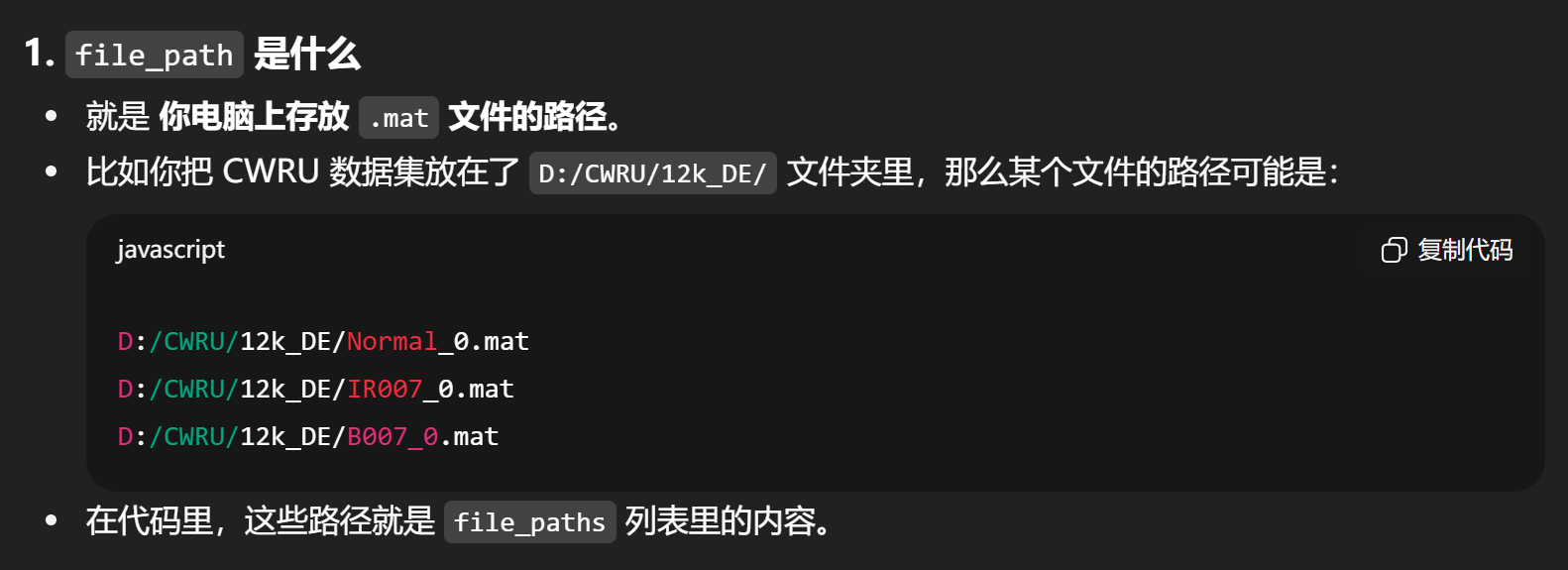


上篇链接：

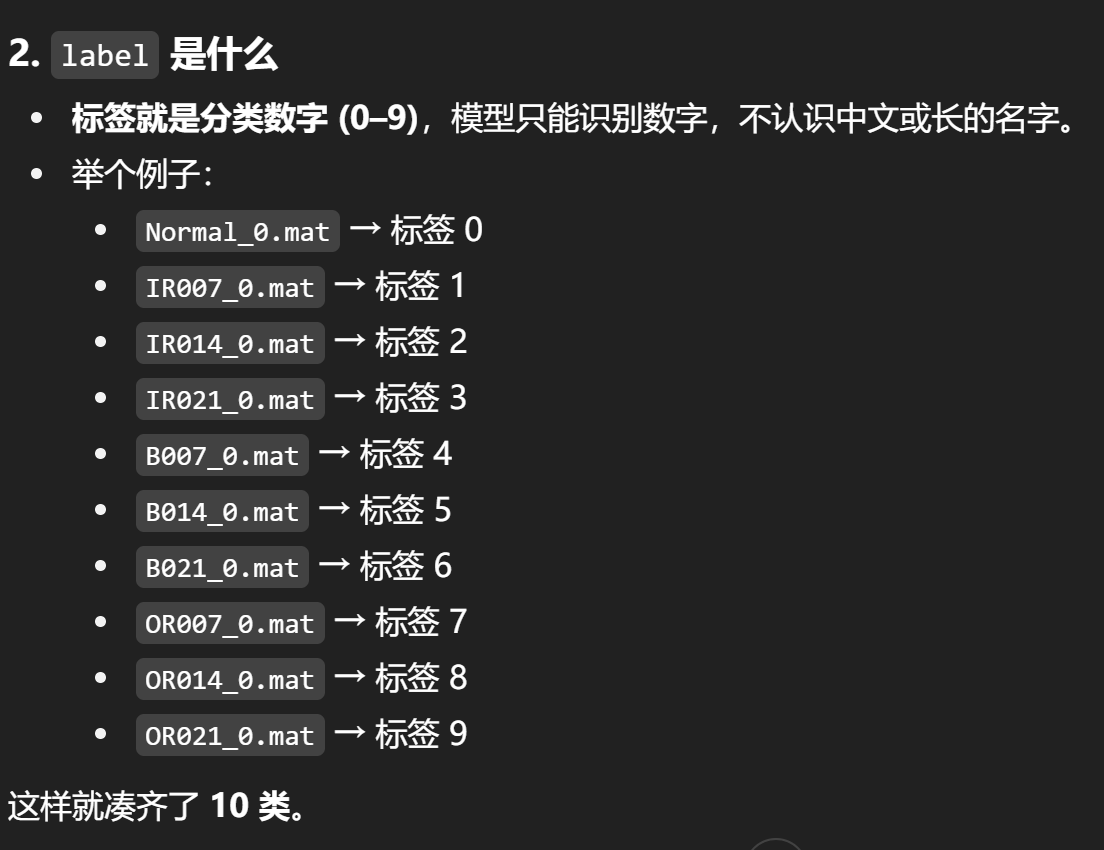
<https://mp.weixin.qq.com/s/1JWxvsvwi_TEJ9I7IpvOTQ?from=groupmessage&scene=1&subscene=318&sessionid=1756534474&clicktime=1756534556&enterid=1756534556&ascene=1&realreporttime=1756534556654&forceh5=1>

然后再看代码之前，我认为还需要搞懂的：

1.file path 这个很好懂

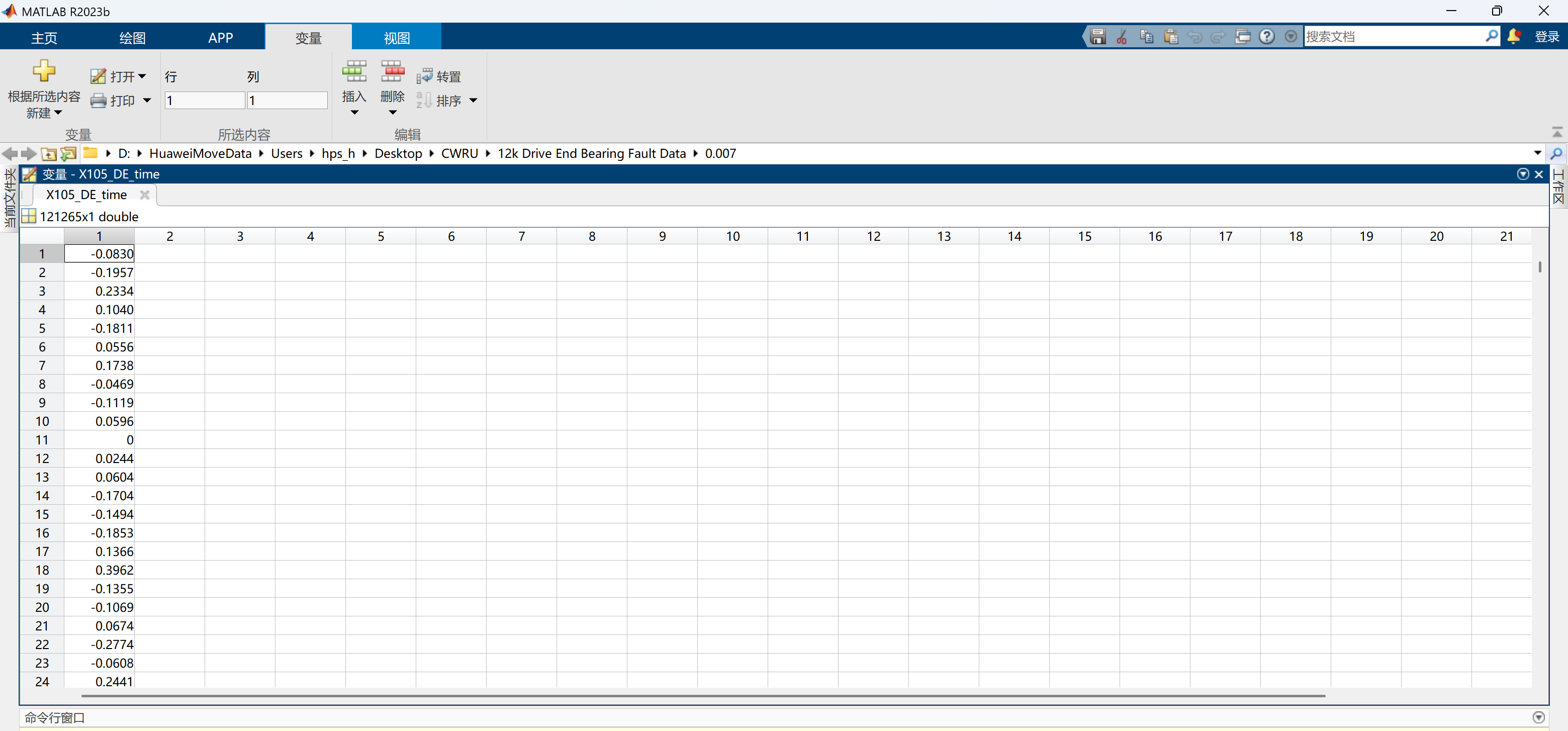


2.lable 这个如果后面看混淆矩阵（一种判断模型效果的方式，贺文斌学长大论文里也用到过）要用到，你得知道设的每个数字对应的是哪种故障类别

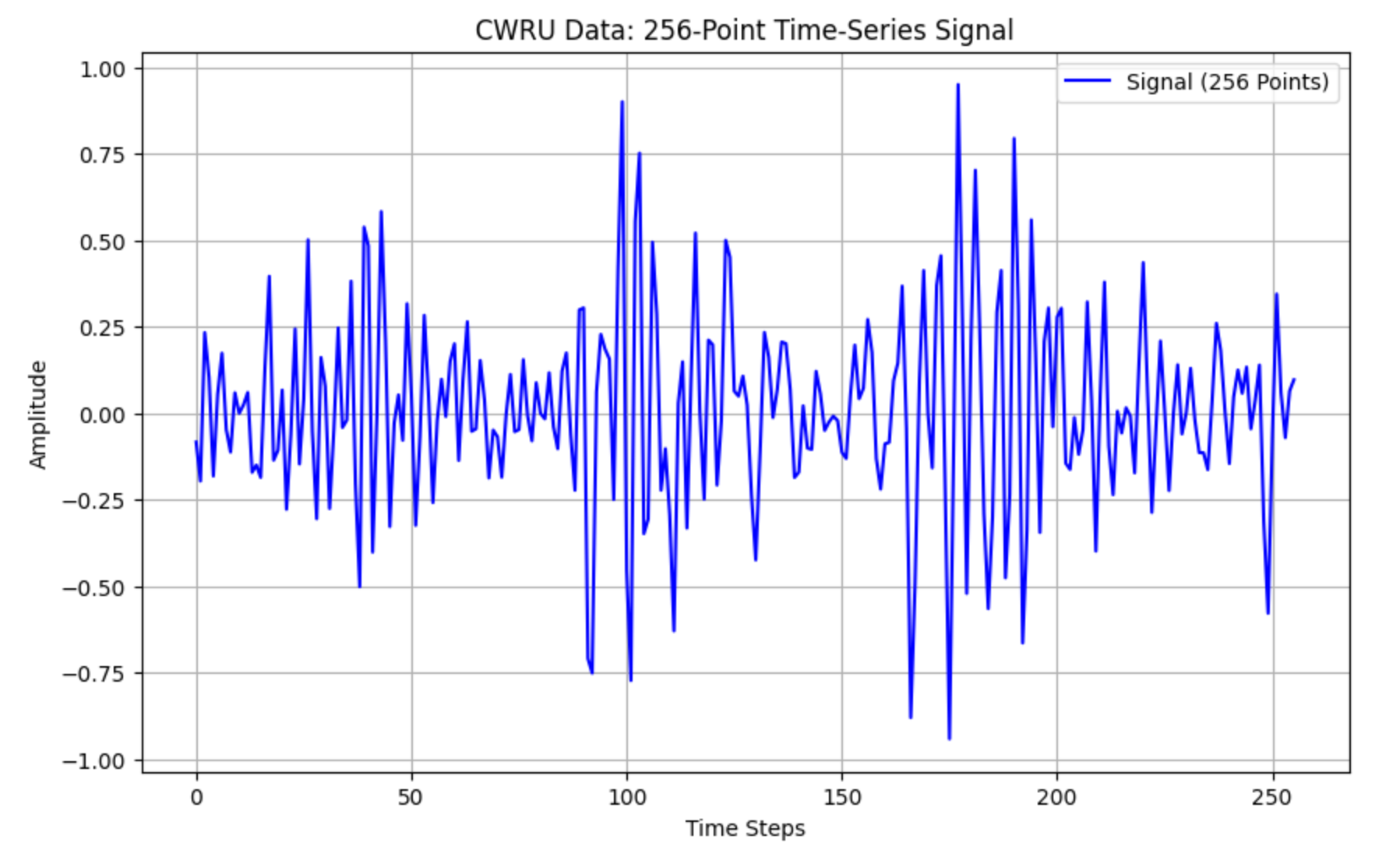


1. 切割样本

比如我们在matlab里打开一个.mat文件，它是这个样子：

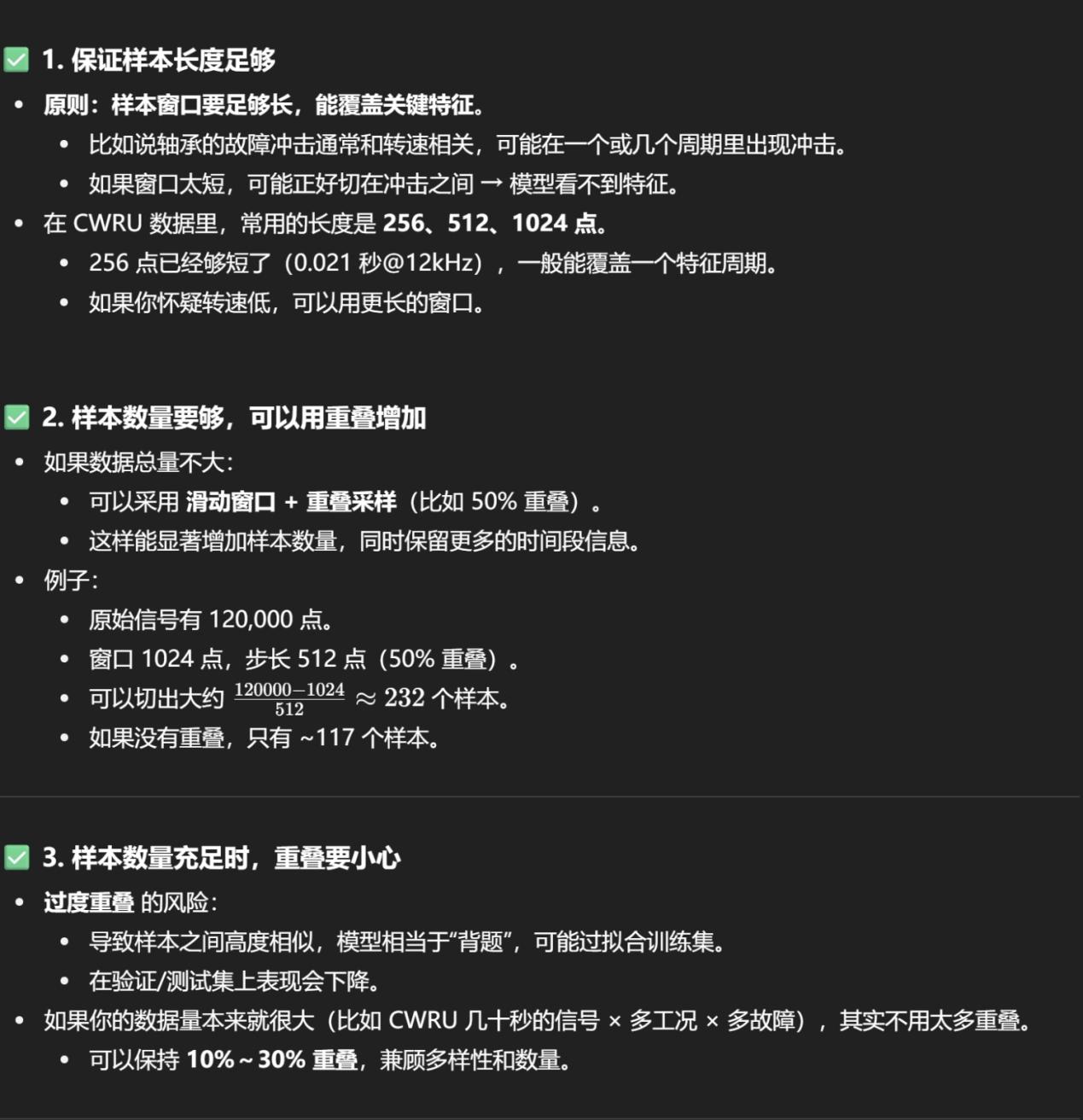


里面记载的是时序信息，这个是12kHz的采样频率，即每秒采集12000次数据，那每个数据就是这一个单元格里的一个小数，即该时刻的加速度值。但我们不能把一个单独的加速度值当作一个样本去给模型训练或测试，因为单个加速度值没有任何模式信息，就像一首歌里取出一个音符，完全无法代表整首歌的风格。故障特征体现在振动的变化规律（随时间的周期性、冲击性、幅值大小等），而不是某个单点。**所以我们必须取连续的一段信号，才能保留时序特征。**如图：



上图是取了IR007\_0.mat中X105\_DE\_time信号的前256个数据点绘制成的波形图。这样一个波形图是有特征的，而单一的某个数据点只反应当时的加速度值，没有明显特征。

那么我们在切割时到底要选多少个数据点作为一段样本合适呢？这是我在改进了几次使用ResNet1D的训练（test accuracy从0到99.61%）后总结出来的，想法在gpt这里也得到了验证。



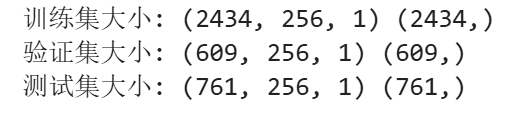
下面是用ResNet1D模型和1D-CNN模型进行训练的代码及结果：

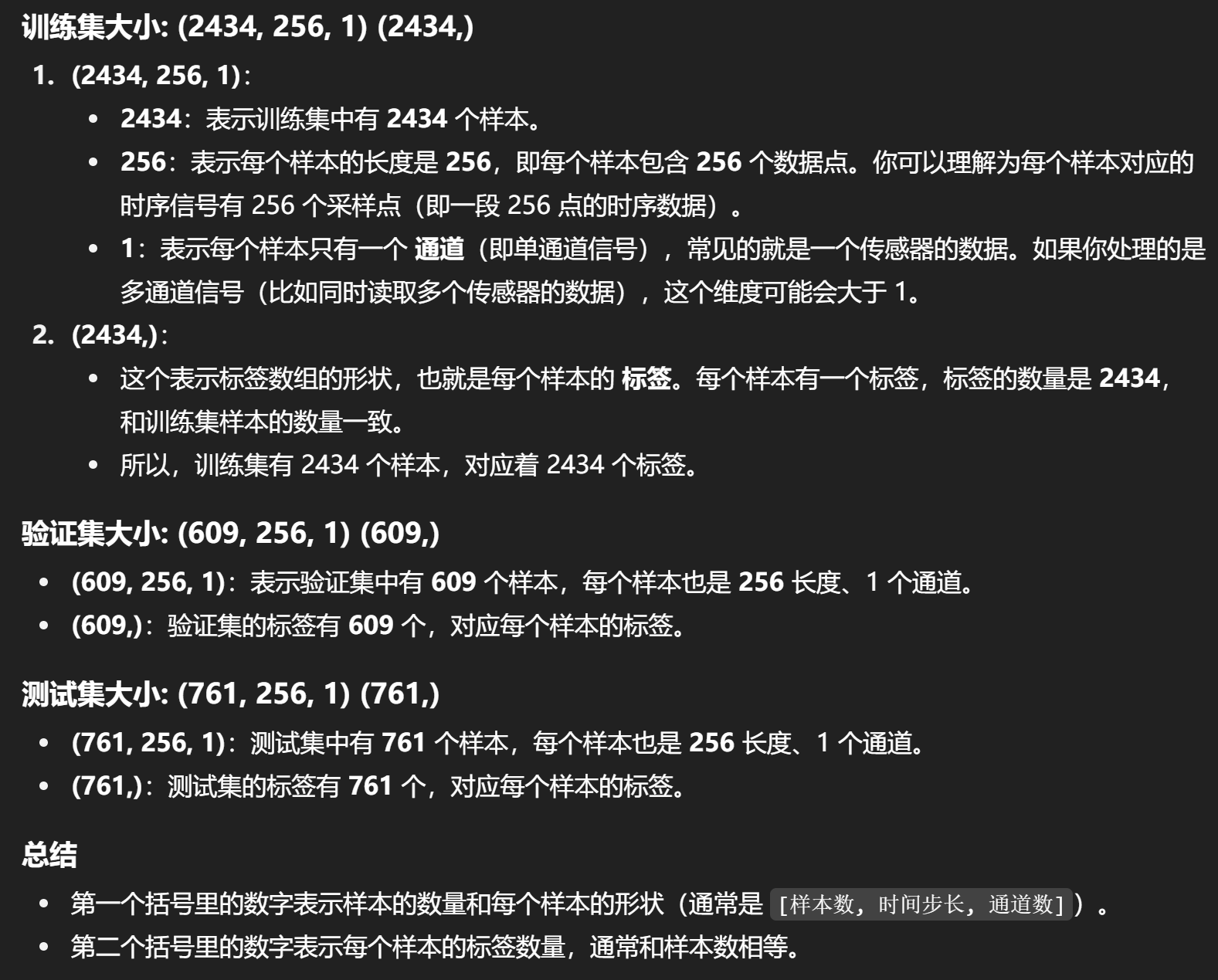
<https://github.com/RiverFloodCity/lalalahyc>

其中ResNet1D更新了四版，所以应该能看到四组代码和运行结果。

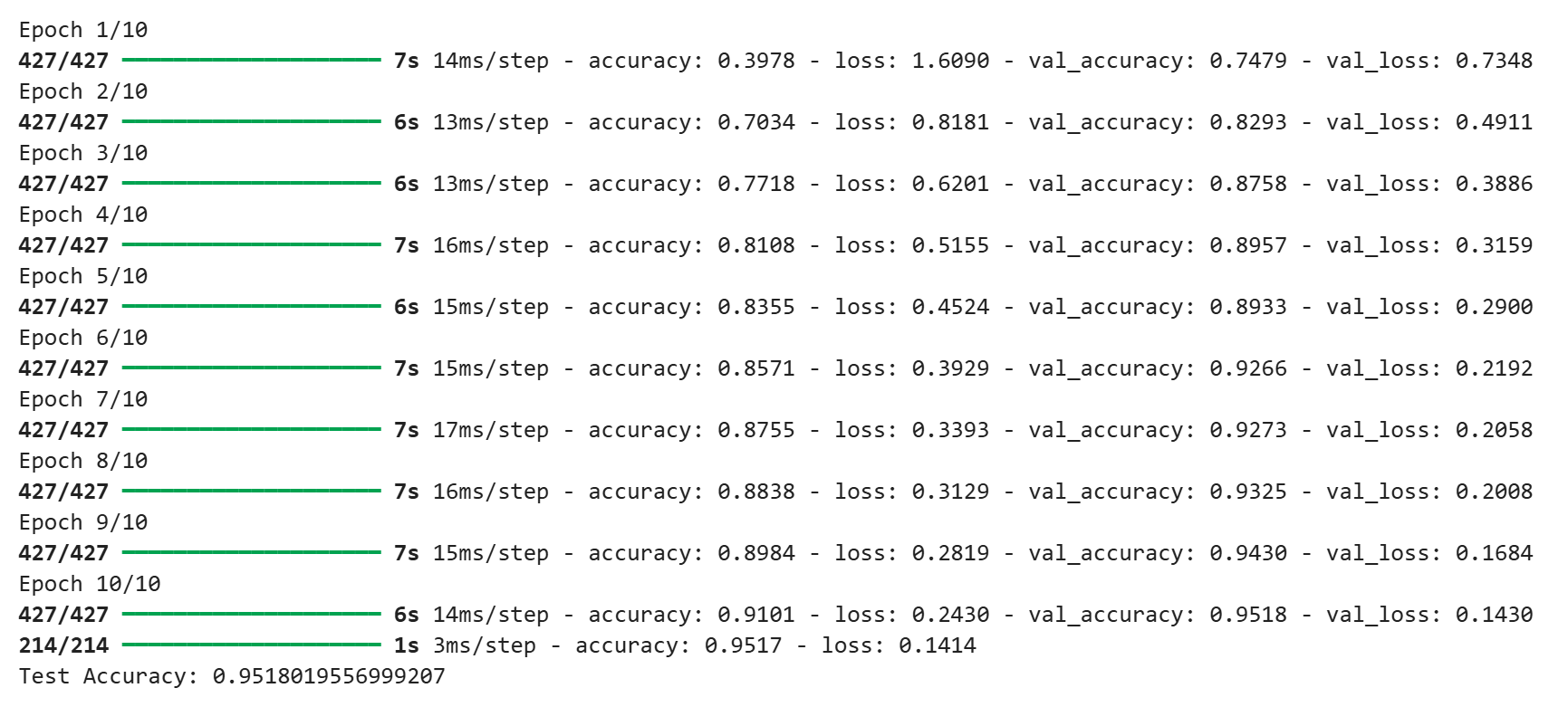
下面是一些帮助理解训练结果的：

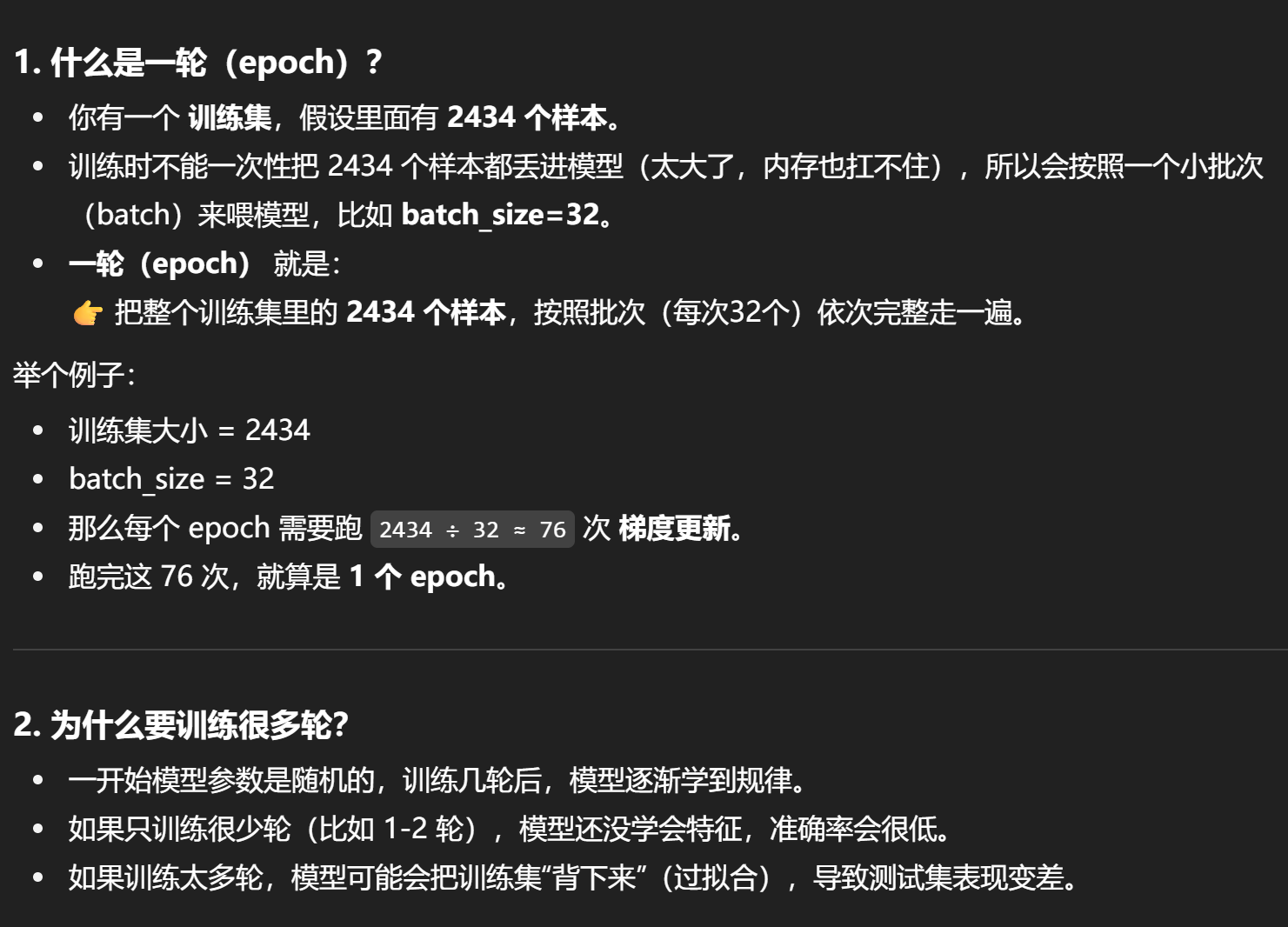
1. 集的大小（形状）：





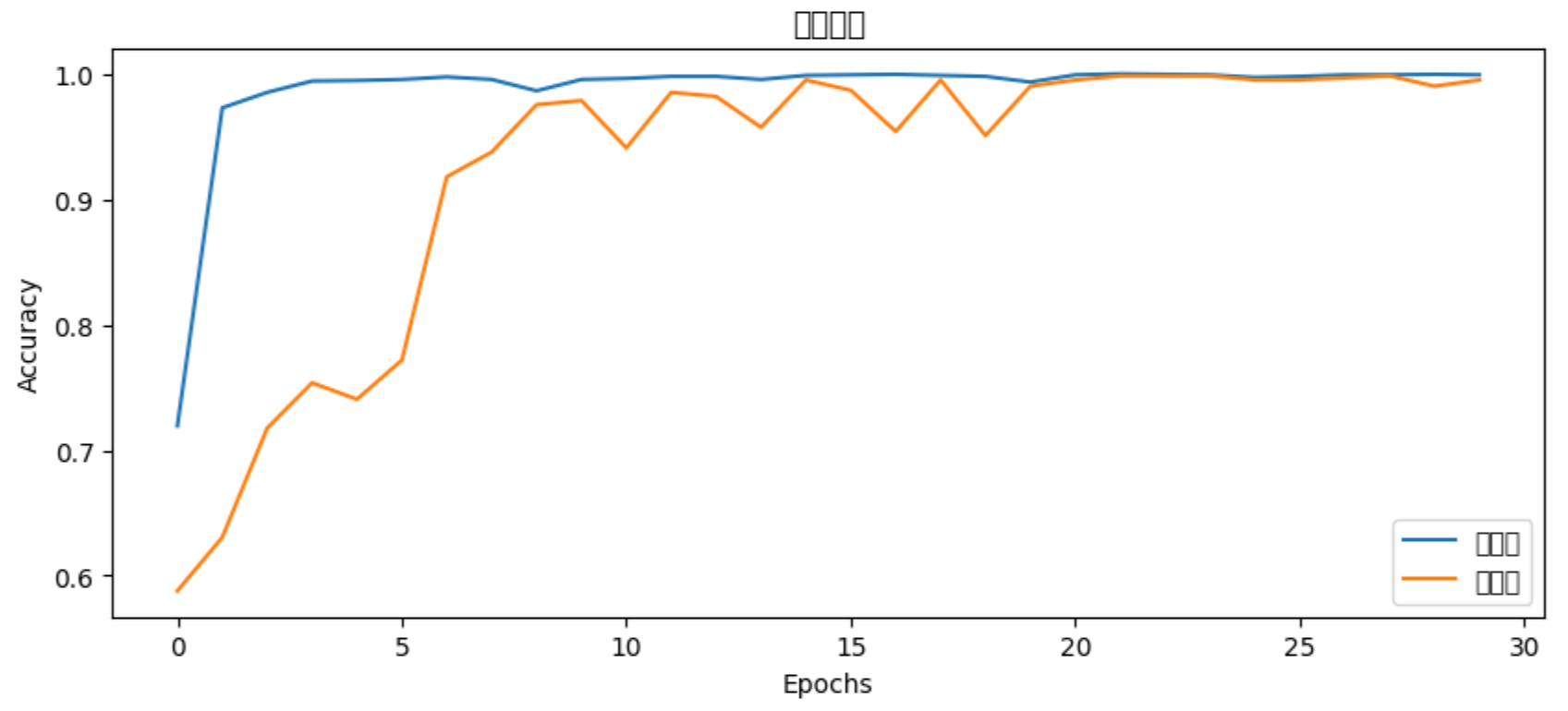
1. Epoch、accuracy、loss：

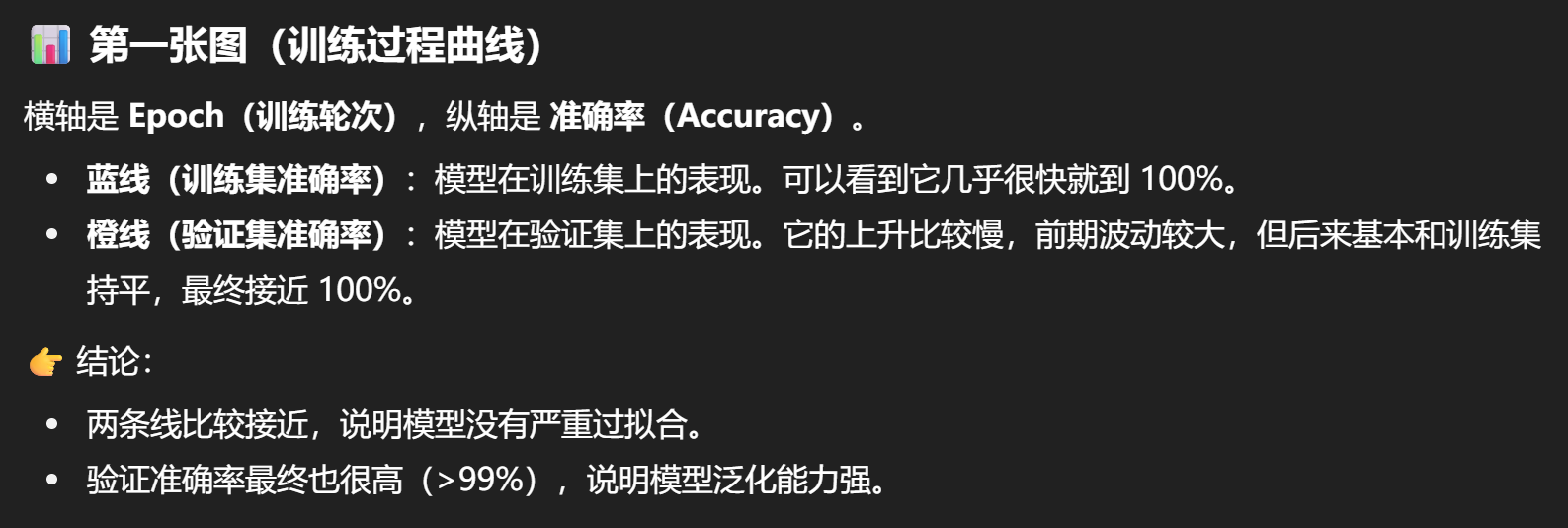




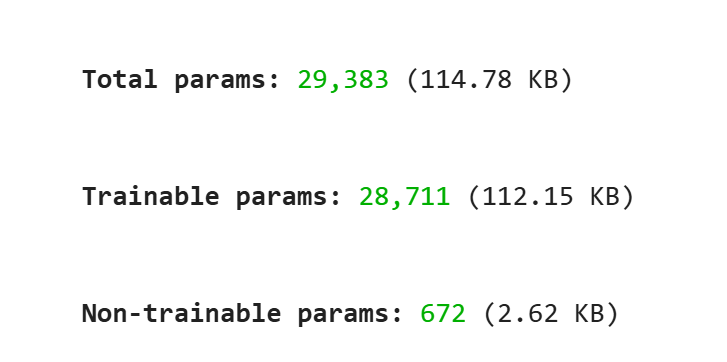






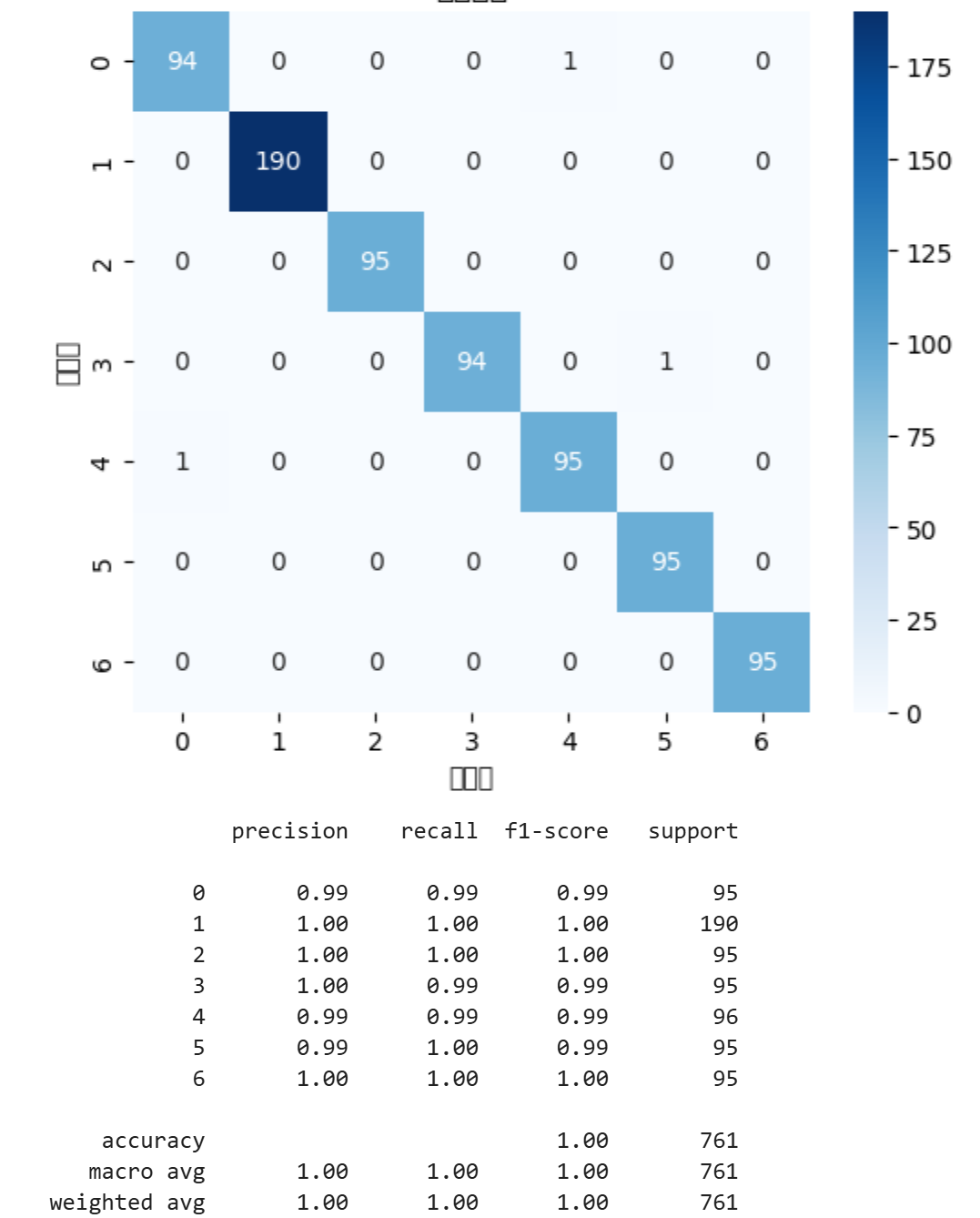


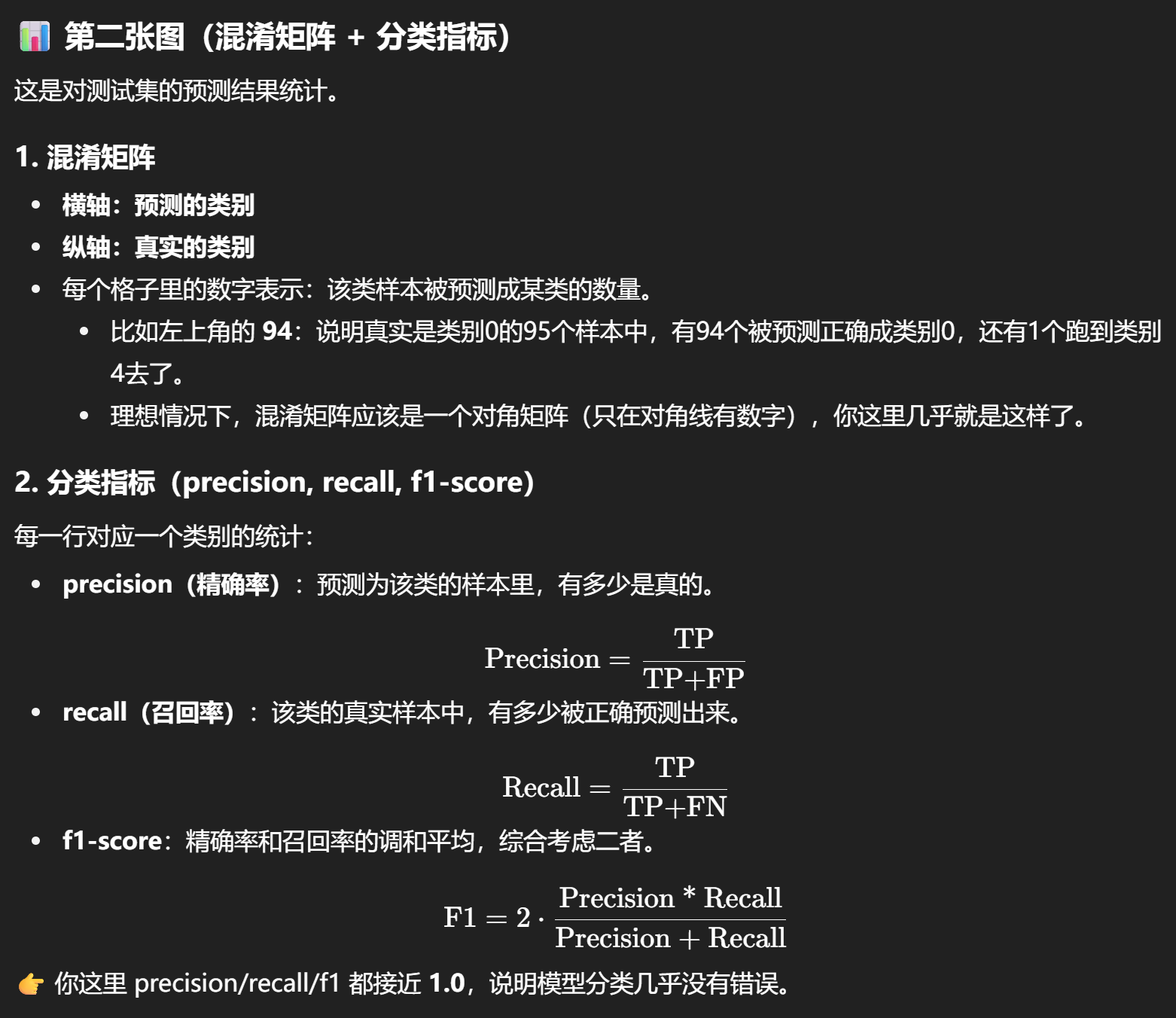
1. Params：





1. 混淆矩阵：





至于具体模型的结构就等后面再研究吧。