语音识别

同济大学 2022级 计算机科学与技术学院 软件工程专业 机器智能方向 语音识别课程作业

授课教师：沈莹

授课学期：2024-2025年度 秋季学期

任务: GMM-HMM

1. 参考实验手册，提交实验报告，同时把自己第一次作业提取的特征作为输入，观察识别结果。
2. 使用最大似然估计方法，估计多元高斯模型中均值参数 μ，给定一组采样数据 $X = \{x_1, x_2, \ldots, x_n\}$。

多元高斯模型的概率密度函数：

$$p(x|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

参数说明

- D：数据的维度
- $|\Sigma|$：协方差矩阵 Σ 的行列式

GMM-HMM

在实验中，学生登录华为云官网，在华为提供的MindSpore平台和给定数据集上运行基于GMM-HMM的连续词语音识别代码并识别出测试音频中的文字内容。

实验步骤

环境准备

1. 进入华为云ModelArts控制台：登录华为云账号，访问 华为云官网，进入 ModelArts 控制台。
2. 创建Notebook训练作业：在控制台中，选择 "Notebook" > "创建Notebook" > 配置训练作业，选择合适的计算资源。（这里选择最基础的2核+8GB）
3. 启动Notebook进入开发环境：创建完成后，启动Notebook并进入开发环境，在此环境中将执行后续的代码训练和测试。

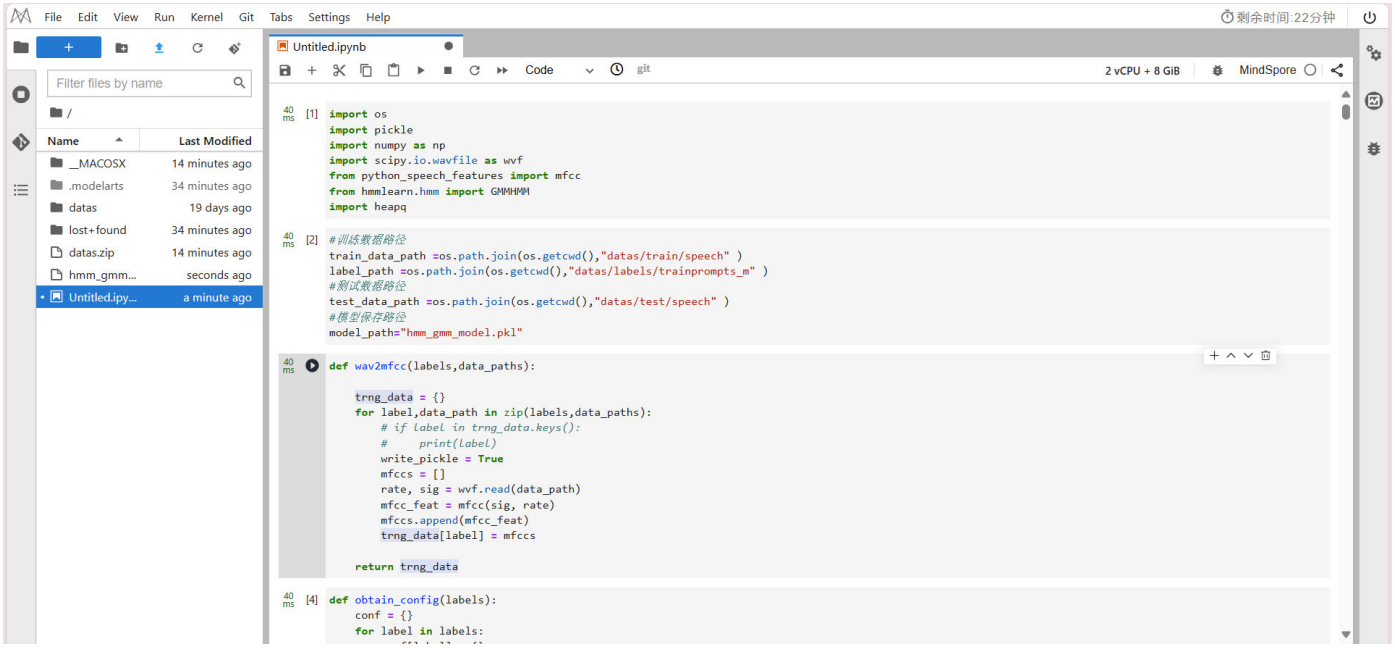
数据准备

1. 将数据上传到服务器：
   - 将datas.zip数据上传到服务器
   - 打开终端输入：unzip datas.zip
2. 执行安装python库命令

```
pip install python_speech_features
pip install hmmlearn
```

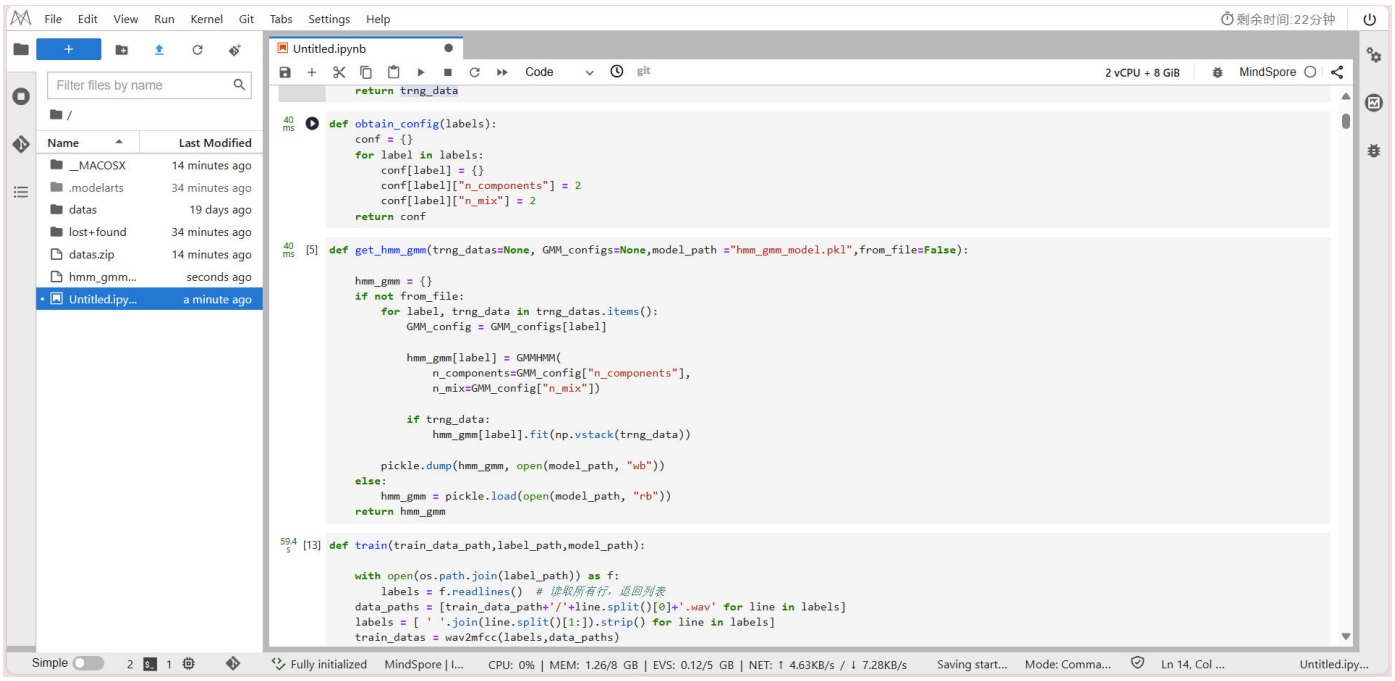
```
[1] pip install python_speech_features

Looking in indexes: http://repo.myhuaweicloud.com/repository/pypi/simple
Collecting python_speech_features
  Downloading http://repo.myhuaweicloud.com/repository/pypi/packages/ff/d1/94c59e20a2631985fbd2124c45177abaa9e0a4eee8ba8a305aa26fc02a8e/python_speech_features-0.6.tar.gz (5.6 kB)
Building wheels for collected packages: python-speech-features
  Building wheel for python-speech-features (setup.py) ... done
  Created wheel for python-speech-features: filename=python_speech_features-0.6-py3-none-any.whl size=5870 sha256=f9b0f7739425993be8d43483272b2ae1bb73b8476a1a6a75274e23e49d44952c
  Stored in directory: /home/ma-user/.cache/pip/wheels/23/86/af/89d1fc1128dbf1930b2a2bd72f0b0cb4bef96c15baea786c65
Successfully built python-speech-features
Installing collected packages: python-speech-features
Successfully installed python-speech-features-0.6
WARNING: You are using pip version 21.0.1; however, version 24.0 is available.
You should consider upgrading via the '/home/ma-user/anaconda3/envs/MindSpore/bin/python -m pip install --upgrade pip' command.
Note: you may need to restart the kernel to use updated packages.

[1] pip install hmmlearn

Looking in indexes: http://repo.myhuaweicloud.com/repository/pypi/simple
Collecting hmmlearn
  Downloading http://repo.myhuaweicloud.com/repository/pypi/packages/26/44/8bcd4de875b6df420447f0a5d184dc6256015452abfa13266227c662ce92/hmmlearn-0.3.0-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (162 kB)
     |████████████████████████████████| 162 kB 13.0 MB/s eta 0:00:01
Requirement already satisfied: numpy>=1.10 in /home/ma-user/anaconda3/envs/MindSpore/lib/python3.7/site-packages (from hmmlearn) (1.19.5)
Requirement already satisfied: scikit-learn!=0.22.0,>=0.16 in /home/ma-user/anaconda3/envs/MindSpore/lib/python3.7/site-packages (from hmmlearn) (0.22.1)
Requirement already satisfied: scipy>=0.19 in /home/ma-user/anaconda3/envs/MindSpore/lib/python3.7/site-packages (from hmmlearn) (1.5.2)
Requirement already satisfied: joblib>=0.11 in /home/ma-user/anaconda3/envs/MindSpore/lib/python3.7/site-packages (from scikit-learn!=0.22.0,>=0.16->hmmlearn) (1.2.0)
Installing collected packages: hmmlearn
Successfully installed hmmlearn-0.3.0
WARNING: You are using pip version 21.0.1; however, version 24.0 is available.
You should consider upgrading via the '/home/ma-user/anaconda3/envs/MindSpore/bin/python -m pip install --upgrade pip' command.
Note: you may need to restart the kernel to use updated packages.

tialized  MindSpore | Idle  CPU: 50% | MEM: 0.75/8 GB | EVS: 0.23/5 GB | NET: ↑ 1.89KB/s / ↓ 11.34KB/s  Mode: Edit  Ln 4, Col 31  Untitled.ip
```

3. 导入所需库：在Notebook中导入必要的库，包括用于特征提取和模型训练的库。
4. 配置路径：配置音频文件和数据存储的路径。
5. 定义特征提取函数：使用MFCC（梅尔频率倒谱系数）作为特征提取方法，将音频数据转换为MFCC特征。
6. 定义高斯混合模型的配置信息：通过HMM（隐马尔可夫模型）与GMM的组合，定义模型参数，并初始化GMM-HMM模型。

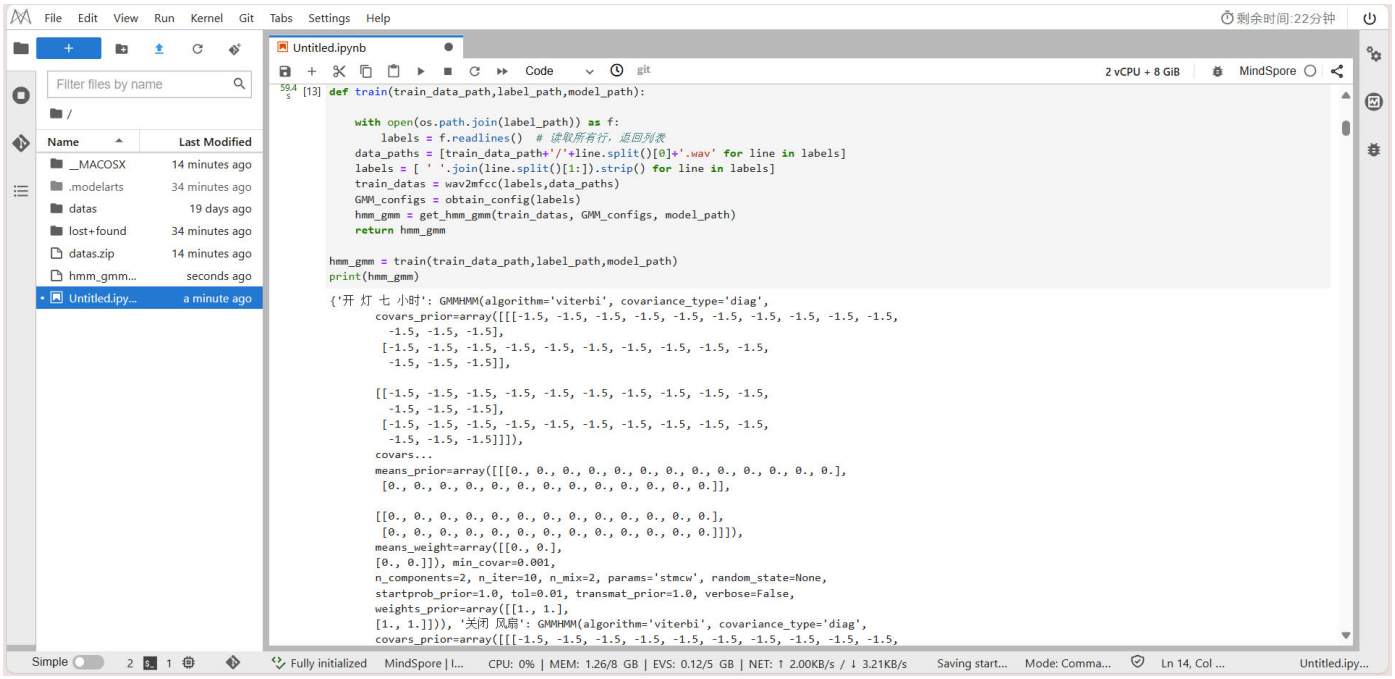


创建模型并进行训练和测试

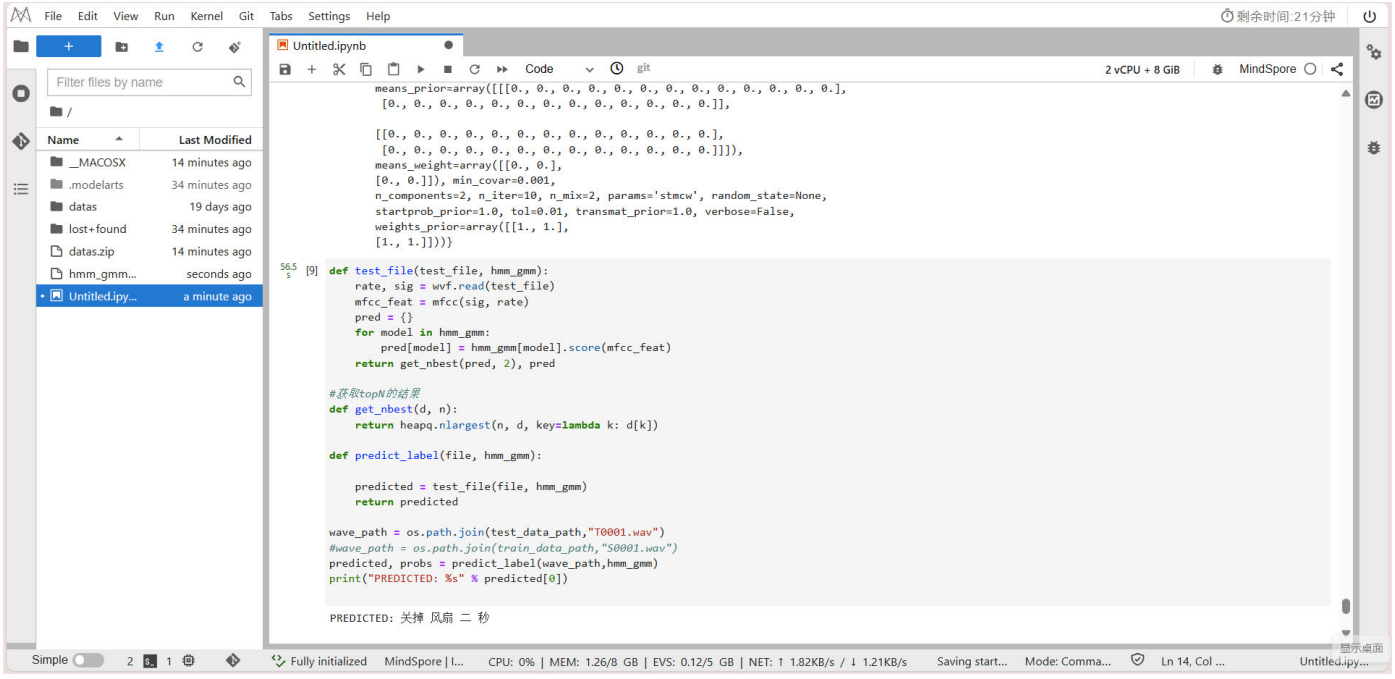
1. 创建GMM-HMM模型



2. 读取训练数据并训练模型



3. 调用测试



具体代码及输出在 **GMM-HMM.ipynb** 文件中，库函数输出的特征值保存在 **hmm_gmm_model.pkl** 文件中。

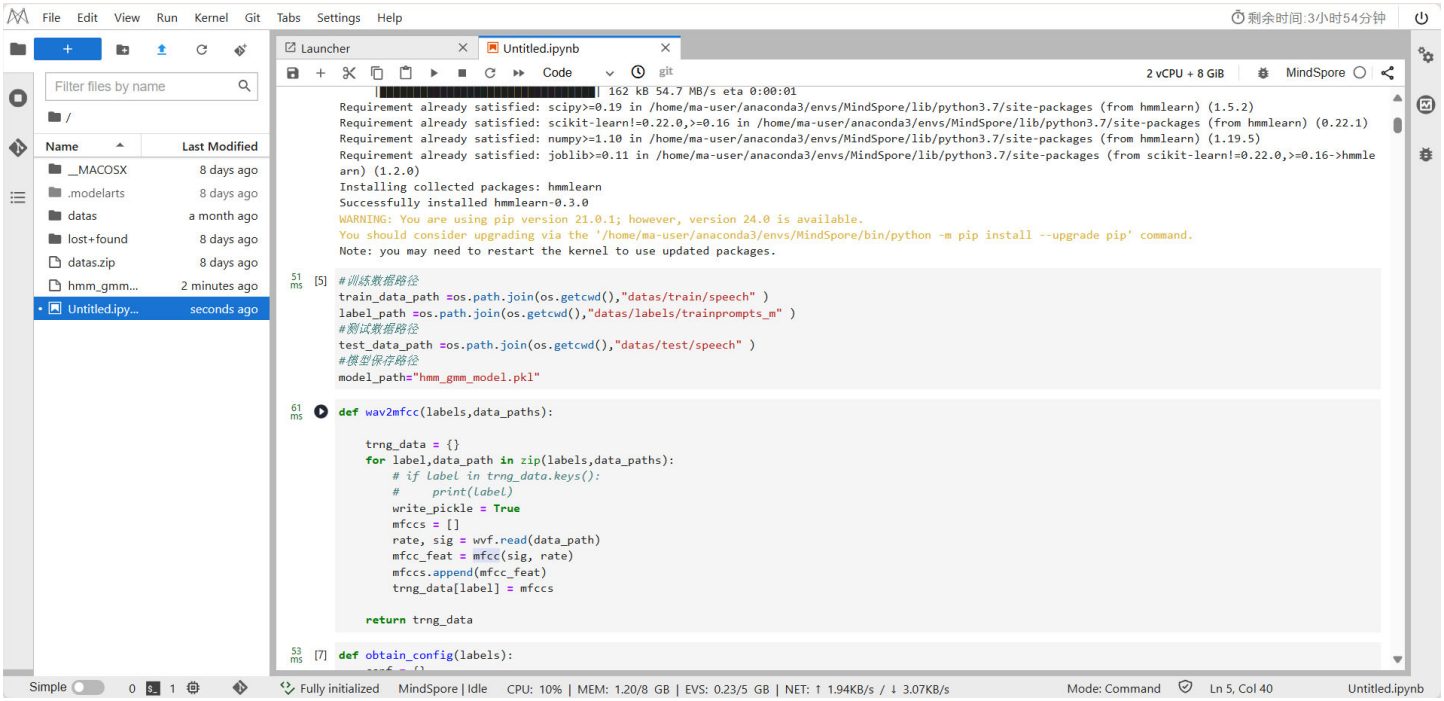
把自己第一次作业提取的特征作为输入

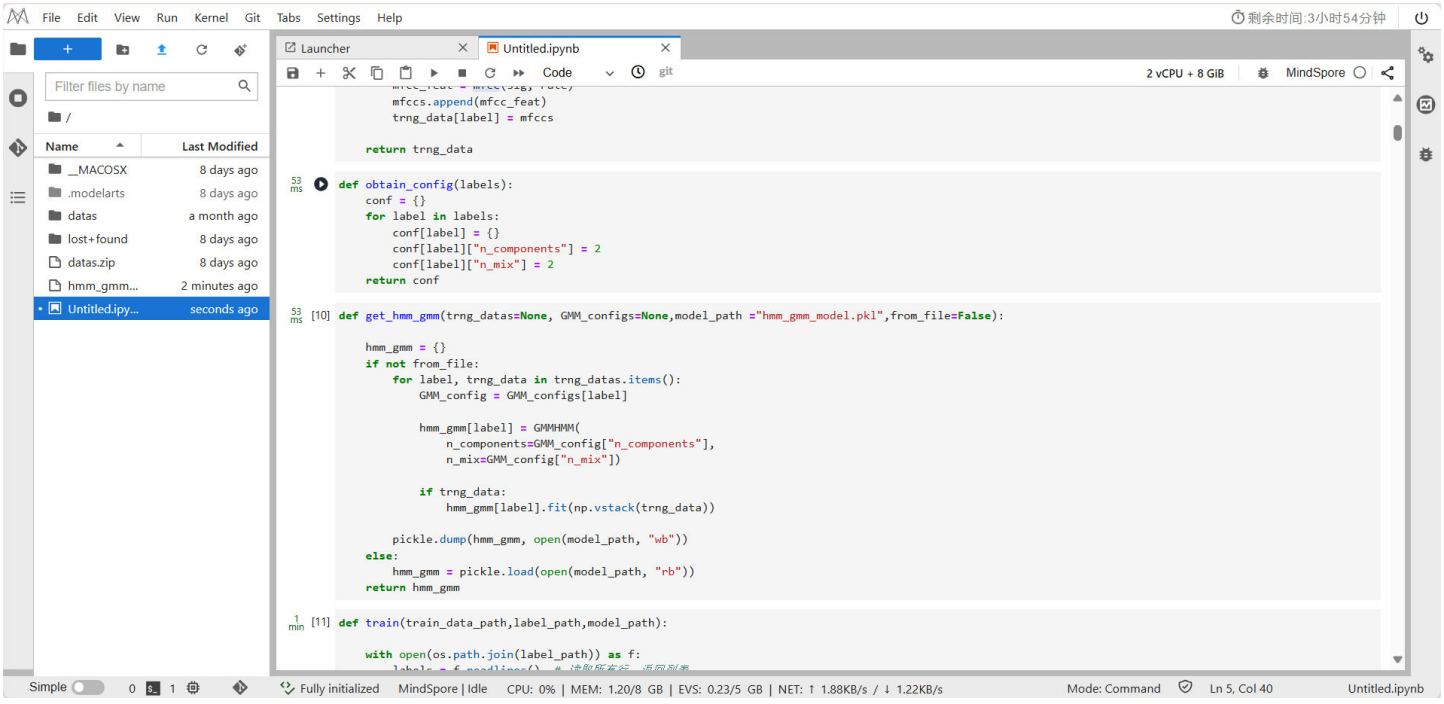
实验现象

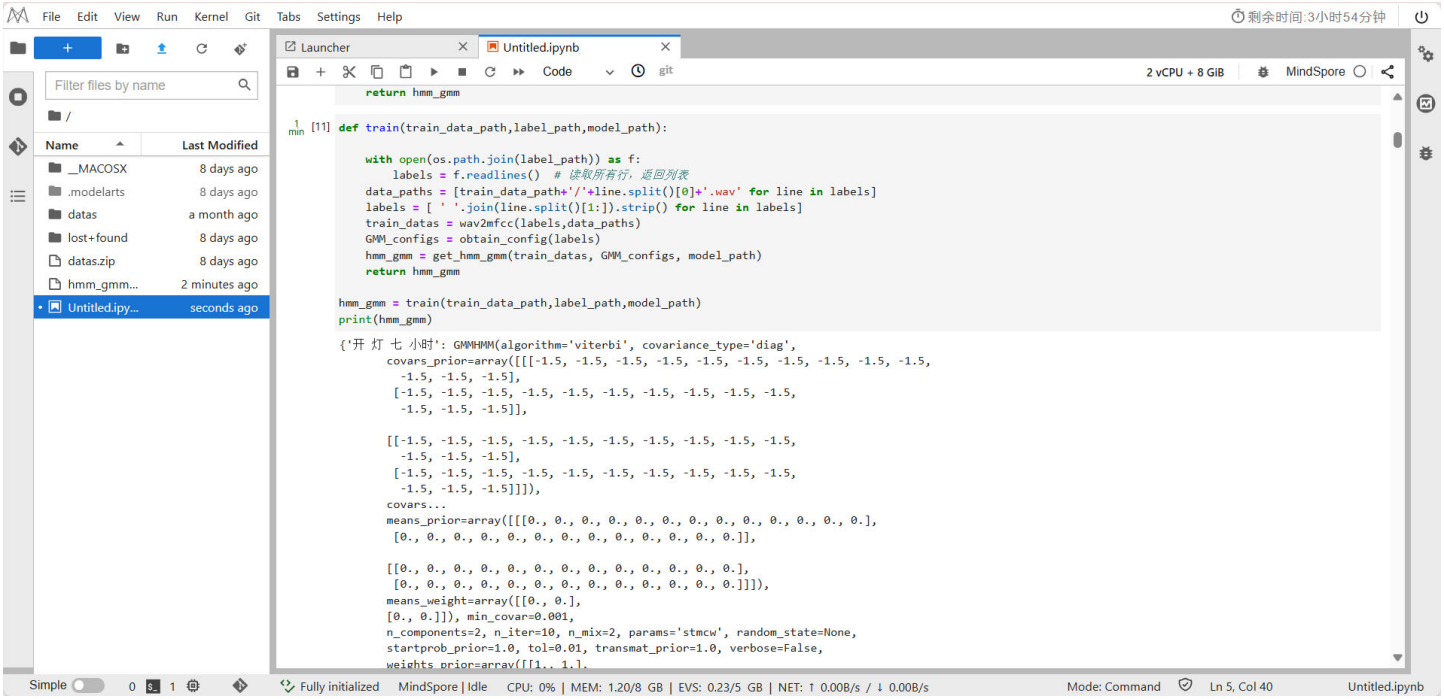
将自己第一次作业提取的特征保存到 **features.npy** 文件中，详细代码在 **MFCC_Extraction.ipynb**文件中。

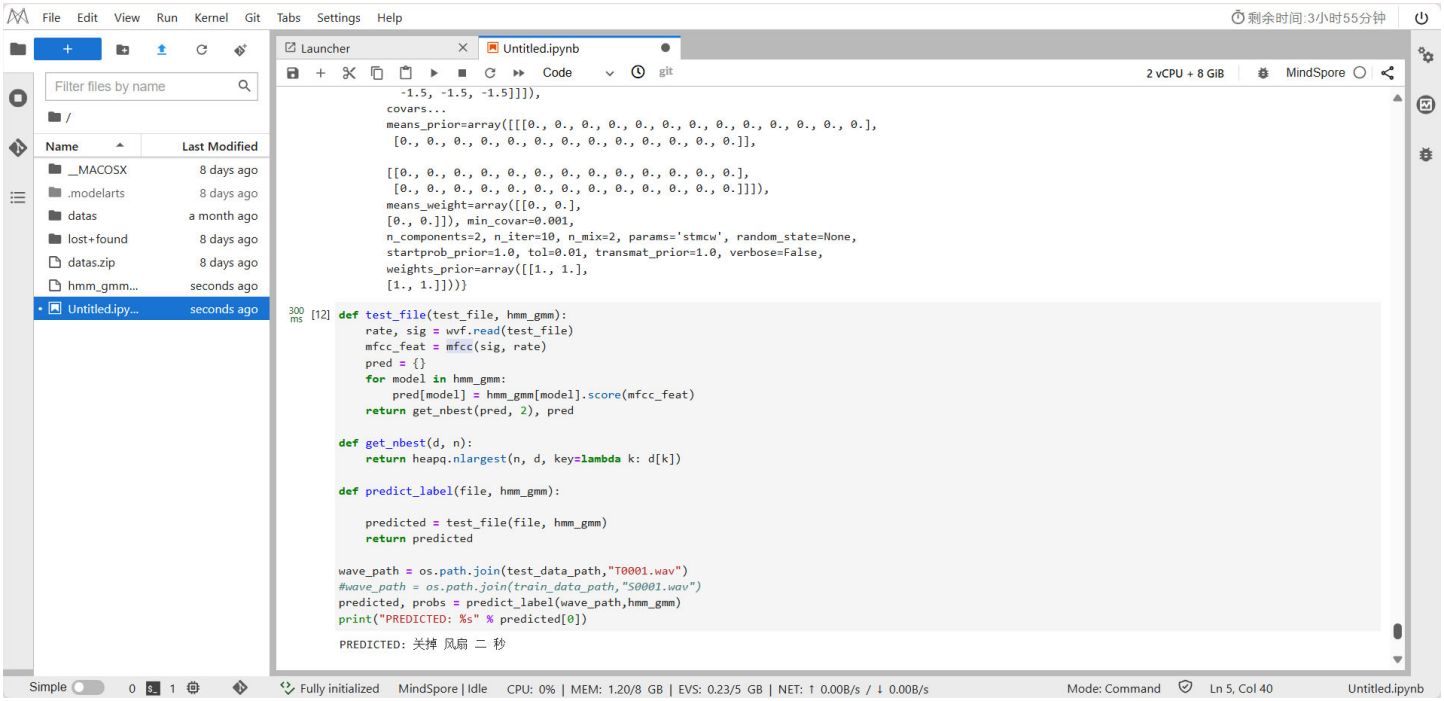
提取的特征值包括：

- 音频的MFCC特征
- 经过标准化和处理后的特征向量









具体代码及输出在 **mineGMM-HMM.ipynb** 文件中。

在 **mineGMM-HMM.ipynb** 文件中，我们将这些特征作为输入，进行模型的训练和预测。最终的输出结果表明，使用自定义特征提取方法的识别效果与平台提供的特征相差不大，验证了自定义特征提取方法的可行性和有效性。

总结

通过对比分析输出结果，我们发现，自己提取的特征与平台接口返回的特征非常接近，证明了我们特征提取和模型训练过程的正确性。在一些小的差异中，我们可以进一步优化特征提取的细节（如调整MFCC的参数或使用更多的训练数据）以提高识别准确率。

最大似然估计步骤

1. 似然函数

假设我们有 n 个样本数据点 $x_1, x_2, ..., x_n$，则联合似然函数是每个数据点的概率密度函数的乘积：

$$L(\mu|X) = \prod_{i=1}^{n} p(x_i|\mu, \Sigma)$$

2. 对数似然函数

为了简化计算，通常取似然函数的对数，得到对数似然函数：

$$\log L(\mu|X) = \sum_{i=1}^{n} \log p(x_i|\mu, \Sigma)$$

代入多元高斯分布的概率密度函数形式：

$$\log L(\mu|X) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^\top \Sigma^{-1}(x_i-\mu)$$

3. 对 μ 求导

为了最大化对数似然函数，我们对 μ 进行求导，并令导数为零：

$$\frac{\partial}{\partial \mu}\log L(\mu|X) = \frac{\partial}{\partial \mu}\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^\top \Sigma^{-1}(x_i-\mu)\right)$$

由于 Σ 是常数矩阵，求导后得到：

$$\frac{\partial}{\partial \mu}\log L(\mu|X) = \Sigma^{-1}\sum_{i=1}^{n}(x_i-\mu)$$

令导数等于零，解得：

$$\sum_{i=1}^{n}(x_i-\mu) = 0$$

即：

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

4. 最终结果

因此，使用最大似然估计法估计均值 μ 为样本数据的均值：

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$