

Machine Learning

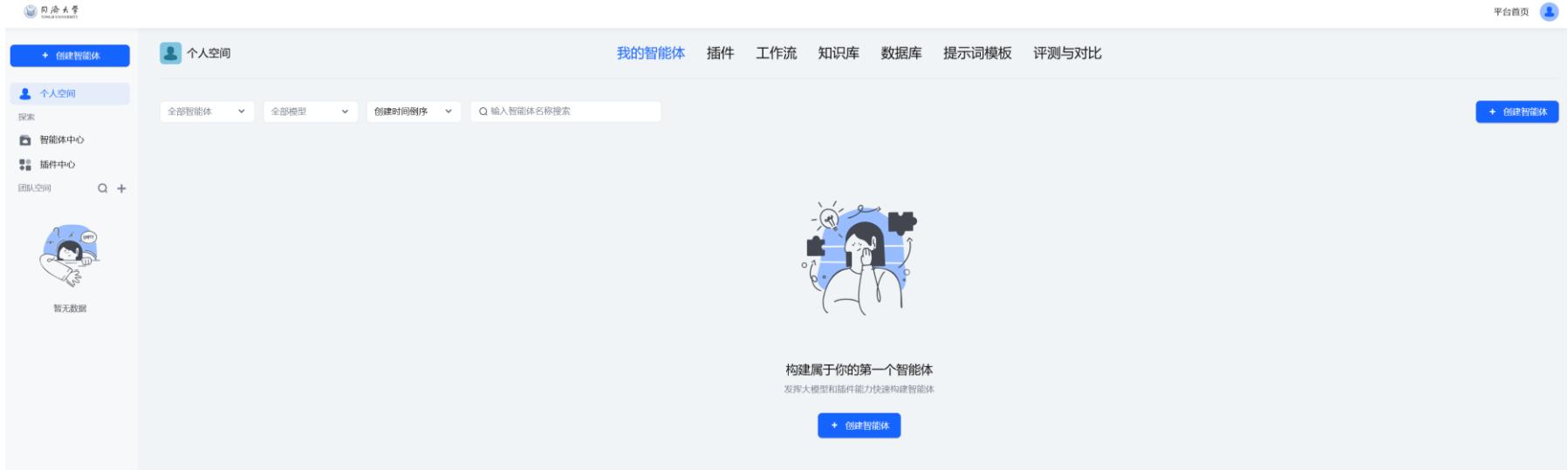
Lab 1: Evaluation & Performance Measure

Dr. Shuang LIANG

Basic lab instructions

- You may want to *bring your slides* to labs to look up syntax and examples.
- Have a question? *Ask a TA* for help, or look at the lecture slides.
- We encourage you to *talk to your classmates*; it's okay to share code and ideas during lab.
- Make good use of these tools: *search engines*, *API documentation* and *IDEs*.
- For the *basic sections*, you need to complete them and submit the results as required. Please submit *real experimental data*.
- For the *extension sections*, you don't have to finish all of the exercises. Just do as much as you can in the allotted time. You don't need to finish the rest after you leave the lab; there is no homework from lab.

Use plugins and knowledge base to build your own Tongji agent



The screenshot shows the Tongji Agent platform interface. At the top, there is a navigation bar with links for "我的智能体" (My Intelligent Body), "插件" (Plugins), "工作流" (Workflow), "知识库" (Knowledge Base), "数据库" (Database), "提示词模板" (Prompt Template), and "评测与对比" (Evaluation and Comparison). On the left side, there is a sidebar with sections for "个人空间" (Personal Space), "智能体中心" (Agent Center), "插件中心" (Plugin Center), and "团队空间" (Team Space). A search bar is located at the top center. In the center of the page, there is a large button labeled "+ 创建智能体" (Create Intelligent Body) and a cartoon illustration of a person working on a puzzle. Below the button, the text reads "构建属于你的第一个智能体" (Build your first intelligent body) and "发挥大模型和插件能力快速构建智能体" (Leverage the power of large models and plugins to quickly build intelligent bodies). The URL <https://agent.tongji.edu.cn> is displayed at the bottom of the screenshot.

<https://agent.tongji.edu.cn>

Document <https://dev.tongji.edu.cn/agent-document/#/>

Return to the home page to create an agent

个人空间 我的智能体 插件 工作流 知识库 数据库 提示词模板 评测与对比

全部智能体 全部模型 创建时间倒序 搜索框 + 创建智能体



构建属于你的第一个智能体
发挥大模型和插件能力快速构建智能体

+ 创建智能体

创建智能体

X

工作空间 *



Personal-2433268

智能体 logo * ⓘ



上传图片

智能体名称 *

7/128

我的校园小助手



智能体类型 *



对话型

建一个聊天场景智能体，采用一问一答模式持续对话。



流程编排型

构建流程编排类智能体，直接将构建的工作流发布使用。

智能体功能介绍 *

9/500

帮助我校园学习生活



取消

确定

我的校园小助手

个人空间 对话型 草稿最后保存于 2025-04-03 10...

编排 纪录 概览

AI 一键生成配置

提示词 ① 自动优化 选择模板

请输入提示词，输入 '{' 插入变量、输入 '}' 插入提示内容块

技能

变量 ^

变量 key * ① 变量描述 ① 操作

SYS_USERID 平台用户 ID

SYS_USERNAME 平台用户名

插件

插件允许智能体调用外部工具，例如搜索信息、浏览网页等，以此扩展智能体的功能。

工作流

工作流通过可视化界面支持插件、大语言模型、代码块和其他功能的组合，从而实现协调复杂而稳定的业

Select model

调试与预览 qwen-plus-latest

语言模型 qwen-plus-latest

随机性 temperature ① 0.50

核采样 top_p ① 0.50

单次回复 max_tokens ① 2000

对话轮数保留 ① 3

RAG 范围 ① 3

思维方式 react

迭代次数 ① 5

个人空间 我的智能体 插件 工作流 知识库 数据库 提示词模板 评测与对比

个人空间

探索 智能体中心 插件中心 团队空间

暂无数据

插件中心 自定义插件

插件分类 权限类别 输入关键词搜索 只看我的收藏

GitHub Toolkit GitHub 是一个开发者平台, 允许开发者创建、存... 8 个工具 工具效率 2 个智能体在使用 · 当前空间 0 个 来自 HiAgent | 2025-03-02 10:31:42 发布

快递鸟 国内快递查询, 根据快递单号查询快递状态。 1 个工具 生活助手 0 个智能体在使用 · 当前空间 0 个 来自 HiAgent | 2025-03-02 10:31:17 发布

正则表达式提取 一个用于正则表达式内容提取的工具 1 个工具 工具效率 8 个智能体在使用 · 当前空间 0 个 来自 HiAgent | 2025-03-02 10:30:57 发布

高德地图 高德地图插件, 该插件能够帮助用户规划出最优... 5 个工具 生活助手 2 个智能体在使用 · 当前空间 0 个 来自 HiAgent | 2025-03-02 10:15:29 发布

飞书 飞书群机器人 1 个工具 生活助手 0 个智能体在使用 · 当前空间 0 个

网页抓取 一个用于抓取网页的工具。支持 html, pdf 1 个工具 网页搜索 39 个智能体在使用 · 当前空间 0 个

The screenshot shows a user interface for AI configuration. At the top, there's a header with the text "AI 一键生成配置". Below it, the interface is divided into two main sections: "技能" (Skills) and "知识" (Knowledge).

技能 (Skills)

- 变量**: A section for defining variables. It includes fields for "变量 key" (e.g., "SYS_USERID") and "变量描述" (e.g., "平台用户ID"). There are also sections for "操作" (Operations) and "SYS_USERNAME" (e.g., "平台用户名").
- 插件**: A section for using external tools. It contains a red arrow pointing to the right, indicating a feature or link.
- 工作流**: A section for business workflow management.
- 触发器**: A section for triggers, described as "可以按照指定规则自动执行指定任务" (Can automatically execute tasks based on specified rules).

知识 (Knowledge)

- 知识库**: A section for knowledge bases. It shows a table with one item: "名称" (Name) "一卡通手册", "文件数" (File Count) "1", and "操作" (Operations). A "强制模式" (Force Mode) button is also present.
- 问答库**: A section for question-answer databases, described as "提供标准问答对给智能体参考, 如果命中问题则会直接返回对应答案" (Provides standard Q&A pairs for the intelligent agent to reference, if a question is命中 (hit), it will directly return the corresponding answer).
- 术语库**: A section for terminology databases, described as "通过术语定义别名、消除歧义、扩展描述、配置公式" (Through terminology definitions, aliasing, eliminating ambiguity, expanding descriptions, and configuring formulas).

Let's try using the plugin

添加插件

×

插件中心 自定义插件

插件分类



权限类别



Q arx



只看我的收藏

arXiv



帮助用户在arXiv中搜索论文，提供搜索关键字，返回论文信息。



1 个工具

网页搜索

2025-02-28 17:00:42 发布

arXiv

根据用户提供的搜索关键字，在arXiv进行搜索，并返回论文信息，包括论文...

query

参数

添加

编排配置 | 单 Agent 模式 ▾

* AI 一键生成配置

提示词 ①

自动优化 选择模板

角色(Role)

学术小助手，具备专业知识，能够高效准确地为用户提供学术方面的帮助

功能(Skills)

功能 1(Skill 1): 使用插件搜索用户提问的关键词
- 根据用户的问题，运用插件进行关键词搜索

限制(Constraint)

- 只能根据用户提问进行关键词搜索，不能进行其他操作

输出(Output)

- markdown
- 以列表形式呈现搜索到的关键词及相关信息，如：
- 论文标题
- 关键词
- 发表时间
- 总结
- 论文链接

检查(Check)

- 确保搜索结果的准确性和相关性
- 避免出现无关或错误的信息

要求(Claim)

- 按照指定格式输出结果
- 使用中文进行描述

Prompt words are used to guide the direction of the big model's answer, making the big model more inclined to answer what we want.

技能

变量 ^

+

变量 key * ①

变量描述 ①

操作

SYS_USERID

平台用户ID

设置

删除

SYS_USERN...

平台用户名

设置

删除

插件 ^

+

插件

工具

操作

X ar... ①

arXiv

设置

删除

工作流

+

工作流通过可视化界面支持插件、大语言模型、代码块和其他功能的组合，从而实现协调复杂而稳定的业务流程。

触发器

+

触发器可以按照指定规则自动执行指定任务。

知识

知识库 ^

强制模式

+

名称

文件数

操作

一卡通手册

1

设置

删除

调试与预览



qwen-plus-latest



证号码的倒数第七位至倒数第二位。例如，如果身份证号码为 36062219850330203X，则密码为 330203。

建议领取新卡后，应尽快到附近的圈存机上修改密码。修改密码时应对消费密码和查询密码区别设置，以增强安全性。

杨淳屹

最新关于视觉大模型的论文

我的校园小助手

已完成

以下是从arXiv搜索到的关于视觉大模型的最新论文信息：

- 论文标题：Review of Large Vision Models and Visual Prompt Engineering
发表时间：2023/07/03
关键词：视觉大模型、提示工程、零样本能力
总结：本文综述了计算机视觉领域中视觉大模型和提示工程的方法，探讨了最新的视觉提示工程技术进展。
论文链接
- 论文标题：Evolving Interpretable Visual Classifiers with Large Language Models
发表时间：2024/04/15
关键词：多模态预训练模型、可解释性、细粒度分类
总结：本文提出了一种新颖的方法，利用大型语言模型及其上下文学习能力，迭代变异用于分类的概念瓶颈属性，生成可解释且具有区分性的属性集。
论文链接
- 论文标题：Survey on Visual Signal Coding and Processing with Generative Models: Technologies, Standards and Optimization
发表时间：2024/05/23
关键词：生成模型、视觉信号编码、优化技术

Agent can use plugins to perform paper searches



Work flow

我的智能体 插件 工作流 知识库 数据库 提示词模板 评测与对比

Q 输入工作流名称搜索

↑ 导入工作流

+ 创建工作流

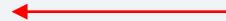


工作流为空

点击按钮创建工作流

↑ 导入工作流

+ 创建工作流



创建工作流

工作流名称 * ① 7/128
会议整理工作流

工作流描述 * 10/500
自动整理总结会议内容

工作流 logo * ①



+

上传图片

取消 确定



test

● 未发布

0个智能体引用

大模型 +

调用大语言模型，使用变量和提示词
生成回复

知识库 +

在选定的知识中，根据设置的召回分段数数量和要求的相似度设置值来召回最匹配的分段信息，并以List形式返回

术语库 +

在选定的术语库中，根据设置的召回分段数数量和要求的相似度设置值来召回最匹配的术语，并以List形式返回

智能体 +

使用已有的对话类型单智能体来编排
业务流程

Start

工作流的起始节点，用于设定启动工作流需要的信息

输入 ^

变量名 * 变量类型 * 描述 *

请输入

String

请输入

End

工作流的最终节点，用于返回工作流运行后的结果信息

回复模式 返回变量，由Bot生成回答

输出 ^

参数名 参数值

output

引用

请选择

成功 5 ms 展示结果

Start 工作流的起始节点，用于设定启动工作流需要的信息

输入 ①

变量名 *	变量类型 *	描述 *	必填
time	String	时间	<input checked="" type="checkbox"/>
room	String	地点	<input checked="" type="checkbox"/>
subject	String	主题	<input checked="" type="checkbox"/>
people	String	与会人员	<input checked="" type="checkbox"/>
content	String	内容	<input checked="" type="checkbox"/>

Add the required input parameters



成功 4752 ms 228 Tokens

大模型 01

调用大语言模型，使用变量和提示词生成回复

模型
qwen-plus-latest

输入参数

参数名	参数值
time	引用 Start/time
room	引用 Start/room
content	引用 Start/content
people	引用 Start/people
subject	引用 Start/subject

提示词
从模板导入

你是一个会议内容整理智能体，根据输入的时间{{time}}地点{{room}}主题{{subject}}人员{{people}}内容{{content}}，请你用一段文字来整理总结此次会议的情况

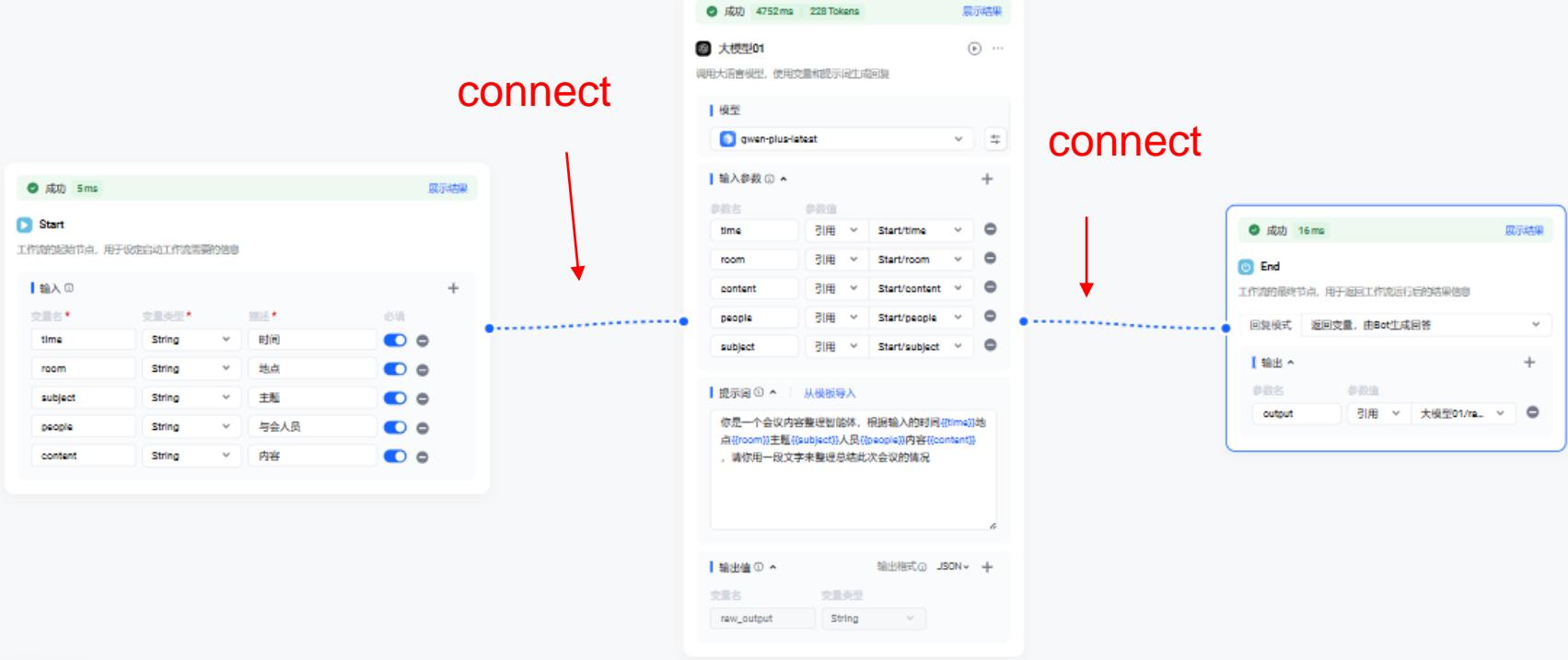
输出值
输出格式 JSON

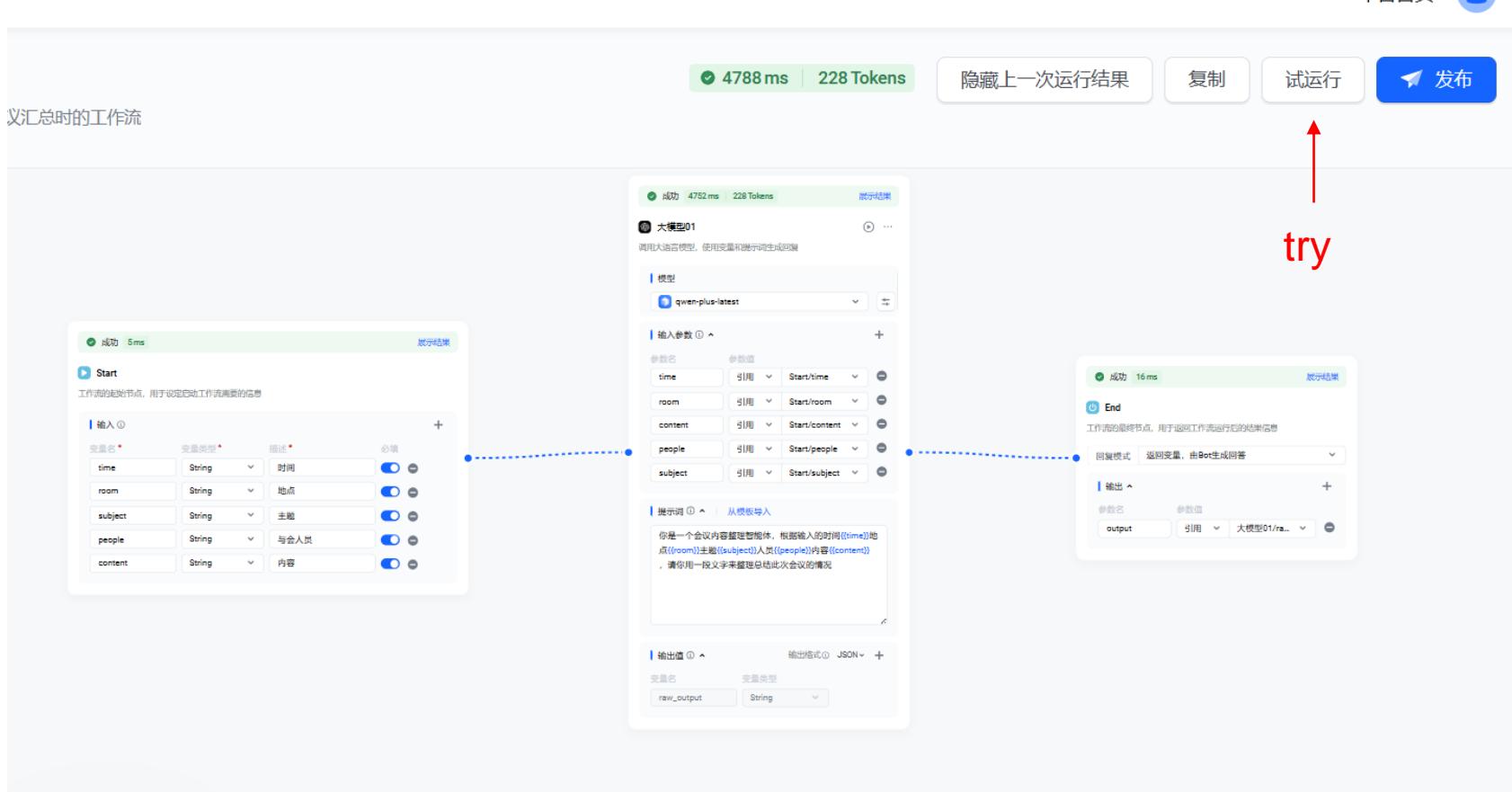
变量名 raw_output 变量类型 String

Add a new large model and then reference the start parameters and prompts



End output refer to the large model output





试运行

X

入参名称	入参类型	入参值	操作
time *	string	20240204	
room *	string	417	
subject *	string	CV	
people *	string	杨淳屹 闫书玮 梁老师	
content *	string	会议主要围绕计算机视觉...	

提交



成功 6974 ms | 200 Tokens

隐藏结果

大模型01

商用大语言模型，使用变量和提示词生成回复

模型

qwen-plus-latest

输入参数

参数名	参数值
time	引用 Start/time
room	引用 Start/room
content	引用 Start/content
people	引用 Start/people
subject	引用 Start/subject

提示词

你是一个会议内容整理智能体，根据输入的时间{{time}}地点{{room}}主题{{subject}}人员{{people}}内容{{content}}，请你用一段文字来整理总结此次会议的情况

Input

```
{ "content": "会议主要围绕计算机视觉领域的最新进展、技术应用及未来发展方向进行了深入讨论", "people": "杨淳屹 闫书玲 梁老师", "room": "417", "subject": "CV", "time": "20240204" }
```

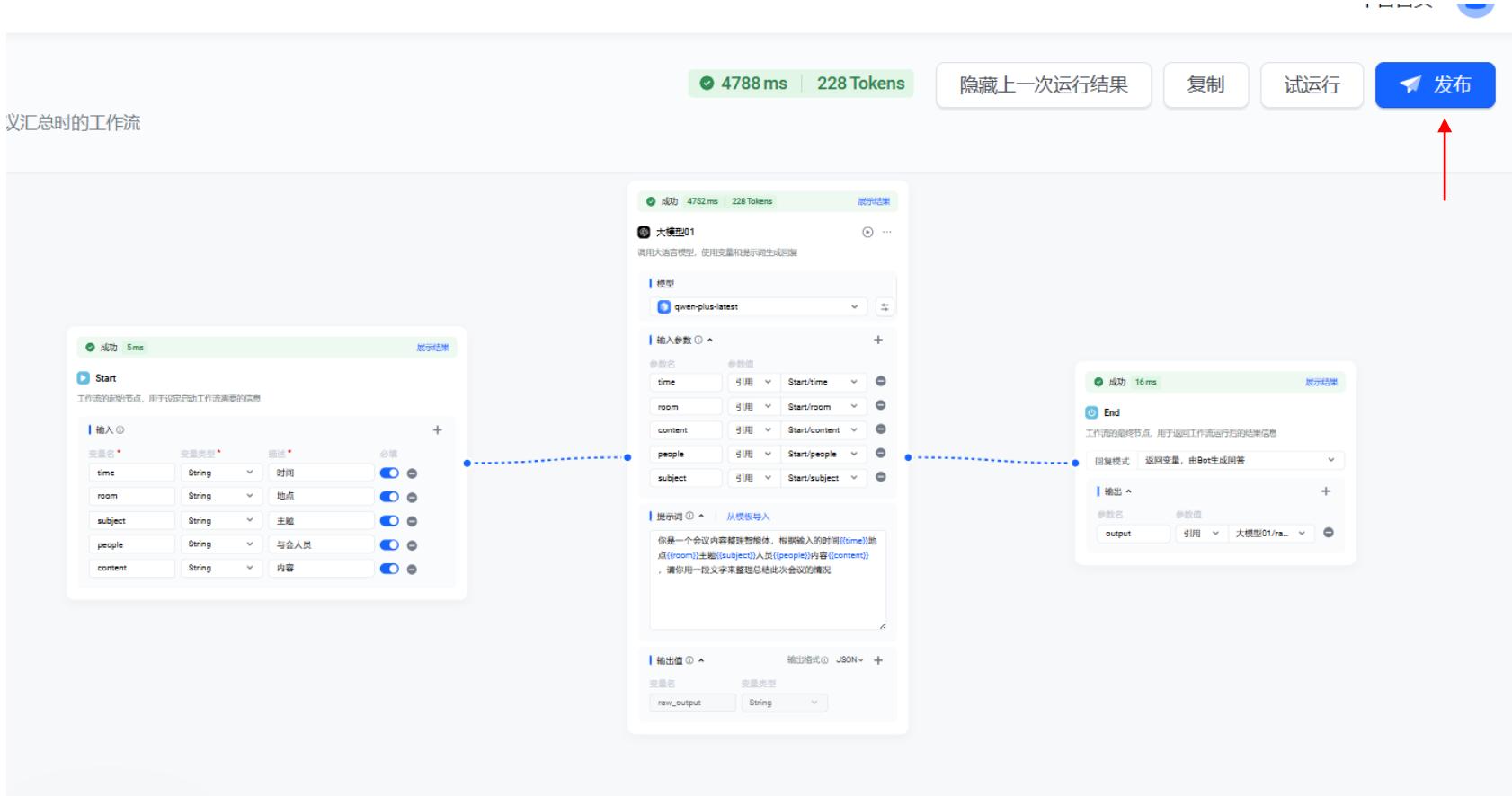
← input

Output

```
{ "raw_output": "[\"时间\": \"20240204\", \"地点\": \"417\", \"主题\": \"CV\", \"参与人员\": [\"杨淳屹\", \"闫书玲\", \"梁老师\"], \"会议内容总结\": \"本次会议主要围绕计算机视觉领域的最新进展、技术应用及未来发展方向进行了深入讨论。与会人员分享了当前计算机视觉技术的研究热点, 探讨了其在实际场景中的应用潜力, 并对未来的趋势和技术突破点提出了见解和展望。\"]" }
```

← output

The screenshot shows a user interface for a large language model. At the top, it displays a success message with performance metrics: 6974 ms and 200 Tokens. Below this is a section titled '大模型01' (Large Model 01) with a note: '商用大语言模型，使用变量和提示词生成回复' (Commercial large language model, generate responses using variables and prompts). The '模型' (Model) dropdown is set to 'qwen-plus-latest'. The '输入参数' (Input Parameters) section contains five entries: 'time' (引用 Start/time), 'room' (引用 Start/room), 'content' (引用 Start/content), 'people' (引用 Start/people), and 'subject' (引用 Start/subject). The '提示词' (Prompt) section contains a template for summarizing a meeting based on provided variables. The right side of the interface is divided into 'Input' and 'Output' sections. The 'Input' section shows a JSON object with the specified variables filled in. The 'Output' section shows the generated response, which is a JSON object containing a single key 'raw_output' with a value that is a list of strings summarizing the meeting.



我的校园小助手

个人空间 对话型 草稿最后保存于 2025-04-03 10:51:...

发布

编排 配置 矩阵 概览

AI 一键生成配置

调试与预览

qwen-plus-latest

工作流 +

触发器 +

知识

知识库 +

名称	文件数	总大小	操作
—卡通手册	1	584	...

输入问题，可通过 shift + enter 换行

内容由 AI 生成，无法确保信息的真实准确，仅供参考

提示词 ①

角色(Role)
学术小助手，具备专业知识，能够高效准确地为用户提供学术方面的帮助

功能(Skills)

功能 1(Skill 1): 使用插件搜索用户提问的关键词
- 根据用户的问题，运用插件进行关键词搜索

限制(Constraint)

- 只能根据用户提问进行关键词搜索，不能进行其他操作

输出(Output)

- markdown

- 以列表形式呈现搜索到的关键词及相关信息，如：
- 论文标题
- 关键词
- 发表时间
- 总结
- 论文链接

检查(Check)

- 确保搜索结果的准确性和相关性

添加工作流

x

+ 创建工作流

已发布



输入工作流名称搜索

C

会议工作流 已发布



会议汇总时的工作流

time

room

subject

people

content

参数

添加

2025-04-07 10:17:42 发布



杨淳屹
时间20240501 参会人员 杨淳屹、闫书伟 地点441 主题人-物交互识别 内容 对今年新发论文进行讨论 请你使用工作流写一段总结

我的校园小助手
已完成

根据会议工作流生成的总结如下：

- 会议时间: 2024年5月1日
- 会议地点: 441
- 会议主题: 人-物交互识别
- 参与人员: 杨淳屹、闫书伟
- 会议内容总结: 本次会议主要围绕2024年新发表的关于人-物交互领域的论文展开讨论。与会人员分析了当前研究热点，包括新型交互技术、传感器应用

输入问题，可通过 shift + enter 换行

[个人空间](#)[我的智能体](#)[插件](#)[工作流](#)[知识库](#)[数据库](#)[提示词模板](#)[评测与对比](#)[知识库](#)[问答库](#)[术语库](#)[数据源](#)[全部类型](#) 输入关键词搜索[+ 创建知识库](#)

暂无知识库

点击按钮创建知识库

[+ 创建知识库](#)

创建知识库

1 基本配置 —— 2 设置类型

知识库名称 * ①

5/128

一卡通手册

Fill in the name

知识库描述

0/300

对该知识库的描述，让用户更加了解数据内容

知识库 logo * ①



+

上传图片

类型 *

标准

管理在平台里的知识库，可以对索引相关做设置

API 接口

托管在外部服务的知识库，通过 API 接口映射参数配置接入

知识库标签 *

标签 Schema ①

+

标签可以用于标记每个文件的扩展信息，比如标题、作者、关键词等，在知识库的检索过程中，这些信息也将帮助大模型更高效地筛选过滤。你可以在此[定义知识库的标签 Schema](#)

不启用标签过滤

开启后知识召回时先通过标签过滤文件，再检索分段内容。启用将增加召回时长。

下一步



创建知识库

x

基本配置 ————— **2** 设置类型

Embedding 模型 * ⓘ

当前可用模型

 Bge-M3-Latest

默认



 Bge-M3 ⓘ

 ali_text_embedding_v2 ⓘ

 doubao-embedding ⓘ

Select default model

如无模型可用, 请联系平台管理员在模型管理中录入嵌入类型的模型, 完成后  刷新 此处列表

当前已选:  Bge-M3-Latest

上一步

提交

 一卡通手册

0 Doc · 0 B 2025-04-03 10:34:27 创建 2025-04-03 10:34:27 更新

文件列表 命中测试 设置详情

创建时间倒序 按文件类型过滤 添加筛选条件



暂无文件
点击按钮新增文件

↗ 拷贝文件 ↗ 导入文件

导入文件

X

通用分段方法说明

支持上传

DOC

DOCX

HTML

HTM

TXT

PDF

PNG

JPG

JPEG

PPT

PPTX

格式文件。

此方法将对文件进行简单的分段：

- 默认会将内容按照分段标识进行分割，再合并成一个不超过分段最大字符数的分块。
- 内容中的表格和图片也会保存，如果要识别图片中的文字，需在自定义规则中启用OCR。

示例



文件来源 *



本地上传

在线下载



飞书文档

空文件

取消

确定

1 选择文件 2 数据处理

本地上传 *

点击或拖拽文件到此处上传

支持 doc, docx, html, htm, txt, pdf, png, jpg, jpeg, ppt, pptx 格式文件
pdf 文件不超过 100 MiB, 图片文件不超过 10 MiB, html/htm/txt 文件不超过 30 MiB
其他单文件不超过 50 MiB
单次最多上传 300 个文件
文本类型编码格式要求 utf-8 否则可能解析成乱码

分段方式 * ①

自动分割 自动设置分段规则与预处理规则

自定义分割 自定义分段规则、分段长度以及预处理规则等

有效期 ①

永久有效

设置截止日期

标签配置 ①

暂无标签, 知识库未配置标签Schema

下一步

文件名称 类型 分段数量 命中次数

校园一卡通使用指南.pdf 通用 6 0

校园一卡通使用指南.pdf
584 KiB

知识库 > 一卡通手册 > 校园一卡通使用指南.pdf

校园一卡通使用指南.pdf ↗
解析成功 2025-04-03 10:38:11 更新

编辑 下载 ...

输入关键词搜索 查看所有分段 + 创建分段

#1 • 启用 定位

“校园一卡通”使用指南

校园一卡通（以下简称校园卡）采用的是非接触式射频卡，是师生在校内消费和身份识别的载体，具有多重功能。为便于师生全面了解和正确使用校园卡，充分发挥校园卡的作用，请广大师生在使用前认真阅读本指南。

第一条 校园卡的分类、有效期及功能

1. 校园卡分为教工卡、学生卡、宝隆卡和协作卡四种：

(1) 教工卡（蓝卡）

适用于我校正式在编、离退休教职工；通过学校人才中心以人才派遣方式进入学校工作的教职工；博士后人员、其他单位来我校挂职锻炼人员；国际、国内交流学者；国际、国外聘任教师；人事部门规定的其他人员等。

(2) 学生卡（绿卡）

适用于我校注册就读的全日制学生，包括本专科生、硕士及博士研究生，包括留学生中的全日制本科生、硕士及博士研究生；取得其他高校全日制正式学籍并在我校注册就读的联合培养学生、交换生、交流生、插班生等。

(3) 宝隆卡（黄卡）

适用于我校各部的临时聘用人员；在我校进修、培训、旁听的非全日制学生。

(4) 协作卡（红卡）

适用于与同济大学无直接的聘用、教学关系，仅借用同济大学图书馆、宿舍及食堂等资源的其他单位人员。

2. 校园卡全部采用实名制，背面印制持卡人的照片、工号或学号、姓名等个人信息。教工卡和学生卡采用工号（学号）作为校园卡的唯一识别号，宝隆卡和协作卡使用编号作为其校园卡的唯一识别号。校园卡可在校园内一卡通商户刷卡消费，包括食堂、浴室、开水器、图书馆、超市、自助洗衣机等，还在图书馆、宿舍楼等场所刷卡进入。其具体功能、权限与管理办法由相关部门和对口的应用子系统规定。

3. 教工卡和学生卡可替代工作证或学生证使用，具有身份认证的作用，包括图书馆借阅、学生公寓医疗、学生上机、学生体能测试等功能。

4. 教工卡的有效期由人事处（人才交流中心）和总务处提供的信息设置，学生卡的有效期按科教务处、研究生院和留学生办公室提供的参数设置，宝隆卡的有效期按持卡人的聘用合同或学生的劳动合同与同济大学相关部门合作协议、租房合同等证明文件的期限为准，各类卡需要延长有效期的，凭相关证明文件到一卡通管理中心办理。

第二条 密码分类与消费限额

为加强校园卡电子钱包功能的安全防护，学校在校园卡上设定了“两密一限”安全保护措施。所谓“两密”指的是消费密码和查询密码，所谓“一限”指的是无密码消费限额。1. 消费密码：当持卡人在某个消费时段内的累积消费额超过无密码消费限额时，再输入消费密码后方可使用。消费密码同时也是进行圈存（自助充值）时需要输入的密码。2. 查询密码：是…

#2 • 启用 定位

2840 Bits 0 Hit 1 天内

#3 • 启用 定位

2737 Bits 0 Hit 1 天内

#4 • 启用 定位

2573 Bits 0 Hit 1 天内

Click on the file to view and edit the segment

技能

变量 ^ +

变量 key * ① 变量描述 ① 操作

SYS_USERID 平台用户ID

SYS_USERNAME 平台用户名

插件 +

插件允许智能体调用外部工具，例如搜索信息、浏览网页等，以此扩展智能体的功能。

工作流 +

工作流通过可视化界面支持插件、大语言模型、代码块和其他功能的组合，从而实现协调复杂而稳定的服务流程。

触发器 +

触发器可以按照指定规则自动执行指定任务。

知识 ↓

知识库 强制模式 +

将文件或网站作为知识库上传后，智能体可以引用知识库的内容来回答用户的问题。

问答库 向量检索 +

提供标准问答对给智能体参考，如果命中问题则会直接返回对应答案。

术语库 强制模式 +

通过术语定义别名、消除歧义、扩展描述、配置公式等，辅助智能体更好地理解业务。

添加知识库

+ 创建知识库

Q 输入关键词搜索

X

C

校园



一卡通手册

1 Docs

584 KiB

2025-04-03 10:34:27 创建

添加

当前已展示全部 1 个知识库



The screenshot shows the 'My Campus Assistant' configuration interface. On the left, there's a sidebar with sections like 'Prompt Word', 'Skills', 'Plugins', 'Workflow', 'Triggers', 'Knowledge', 'FAQ', and 'Vocabulary'. The 'Skills' section is expanded, showing 'Variable' and 'Operation' blocks with fields for 'key' and 'description'. The 'Knowledge' section shows a 'Knowledge Base' with one item named '一卡通手册'. The 'FAQ' section has a '向量检索' (Vector Search) button. The right side shows a conversation between '杨淳屹' and '我的校园小助手'. The user asks '校园卡默认密码是?' (What is the default password for the campus card?). The assistant responds with an AI-generated answer: '校园卡的默认密码分为消费密码和查询密码，初始值为您身份证号码的倒数第七位至倒数第二位。例如，如果身份证号码为36062219850330203X，则密码为330203。建议领取新卡后，应尽快到附近的圈存机上修改密码。修改密码时应对消费密码和查询密码区别设置，以增强安全性。' Below the answer, it says '5.31s | 2133 Tokens'.

The agent can learn from the knowledge base and answer questions

Add a new database



创建数据库

x

模块

数据库名称 * ①

4/128

data

x

数据库描述

0/300

数据库 logo * ①

请输入描述，让大模型理解什么情况下应
该调用此数据库



+

上传图片

类型 *



上传数据表

通过上传 Excel 表格、CSV 等格式数据表创建数据库



外部直连数据库

通过与已有业务建立连接创建数据库

确定





data

0 Table

2025-04-07 10:36:34 创建

2025-04-07 10:36:34 更新

...

输入数据表名称搜索

导入数据表

文件

上传成功 demo-case1-data1.xlsx

1 选择文件

2 数据配置

3 数据校验

本地上传 *

点击或拖拽文件到此处上传

支持 csv, xls, xlsx 格式文件

单个文件不超过 50 MB

只支持单表格上传, 可多个sheet

文本类型编码格式要求 utf-8 否则可能解析成乱码

demo-case1-data1.xlsx 9.92 KiB



2 选择文件

2 数据配置

3 数据校验

数据表名称 *

Sheet1

6/32

数据表描述 *

询问训练数据表时调用此数据库

14/400

AI 生成

列名 * ①

x

数据示例

0.4

列描述 ①

请输入

数据类型 ①

小数

y

1.52

请输入

小数

1 选择文件

2 数据配置

3 数据校验

上一步

下一步

数据表名称

数据量

数据表描述

Sheet1

50行 · 2列

询问训练数据表时调用此数据库



我的校园小助手

个人空间 对话型

草稿最后保存于 2025-04-07 14:49:...

编排 统计 概览

编排配置

单 Agent 模式 ▾

※ AI 一键生成配置

提示词 ⓘ

自动优化

选择模板

角色(Role)

学术小助手，具备专业知识，能够高效准确地为用户提供学术方面的帮助

功能(Skills)

功能 1(Skill 1): 使用插件搜索用户提问的关键词

- 根据用户的问题，运用插件进行关键词搜索

限制(Constraint)

- 只能根据用户提问进行关键词搜索，不能进行其他操作

输出(Output)

- markdown

- 以列表形式呈现搜索到的关键词及相关信息，如：

- 论文标题

- 关键词

- 发表时间

- 总结

- 论文链接

检查(Check)

- 确保语义结果里的准确性和平滑性

问答库

向量检索

+

提供标准问答对给智能体参考，如果命中问题则会直接返回对应答案。

术语库

强制模式

+

通过术语定义别名、消除歧义、扩展描述、配置公式等，辅助智能体更好地理解业务。

数据库

+

引用表格数据后，支持基于自然语言对数据库（即 NL2SQL）进行查询和计算，不支持关联多个数据库。

对话

开场白 ^

※ 自动生成

对话开场白 ⓘ

添加数据库

x

i 当前仅支持添加一个数据库。

+ 创建数据库

输入关键词搜索

c



data

1 Table

2025-04-07 10:36:34 创建

添加



当前已展示全部 1 个数据库

The screenshot shows the Qwen-Plus AI interface. On the left, there's a sidebar with three main sections: '问答库' (Question & Answer Library), '术语库' (Vocabulary Library), and '数据库' (Database). Under '数据库', there's a table with one row named 'data'. On the right, the main area is titled '调试与预览' (Debugging and Preview) and shows a conversation with a user named '杨淳屹'. The user asks 'x的平均值' (Average value of x). The AI responds with '根据数据库查询结果, x 的平均值为 10.2.' (According to the database query results, the average value of x is 10.2.). Below the response, it says '20.666s | 3848 Tokens'. A red arrow points from the user's question in the conversation to the 'data' entry in the database sidebar. At the bottom, there's a text input field with placeholder text '输入问题, 可通过 shift + enter 换行' (Input your question, press shift + enter to enter a new line) and a send button.

AI 一键生成配置

调试与预览

qwen-plus-latest

问答库

向量检索

提供标准问答对给智能体参考，如果命中问题则会直接返回对应答案。

术语库

强制模式

通过术语定义别名、消除歧义、扩展描述、配置公式等，辅助智能体更好地理解业务。

数据库

名称

数据 操作

data 1 -

对话

杨淳屹

x的平均值

我的校园小助手 14:54:57

已完成

根据数据库查询结果, x 的平均值为 10.2.

20.666s | 3848 Tokens

输入问题, 可通过 shift + enter 换行



创建提示词模板

×

模板名称 * ⓘ

7/128

会议提示词模板

提示词 *

62/1000

会议时间:{{time}};
会议地点:{{room}};
会议人员:{{people}};
会议主题:{{event}}]

变量 : time room people event

取消

确定



创建评测集

×

评测集名称 *

5/128

测试评测集

评测集描述 *

5/500

测试评测集

取消

确定

保存于 2025-04-07 11:06:34

↑ 数据导入

发布

数据导入

×

! 导入成功后将覆盖此评测集原有数据, 请确认操作。

数据导入 *

↓ 下载模板



点击或拖拽文件到此处上传

支持 xls, xlsx, csv 格式文件

单个文件不超过 10 MiB

文本类型编码格式要求 utf-8 否则可能解析成乱码

取消

确定

group	input	output
1	深圳特区是什么时候成立的?	1980年。
1	深圳市包含几个行政区?	深圳市包含9个行政区和1个新区, 分别是光明区、宝安区、南山区、福田区、罗湖区、盐田区、龙岗区、龙华区、坪山区和大鹏新区。
2	天山山脉在哪里?	天山山脉位于欧亚腹地, 中国境内, 天山山脉均在新疆境内, 西起乌恰县, 东至哈密市星星峡以东。
2	新疆还有哪些山脉?	新疆的北部为阿勒泰山, 南部为昆仑山系, 天山山脉横亘于新疆中部, 把新疆分为南北两半。

测试

未发布 测试评测集 自动保存于 2025-04-07 11:06:34

数据导入 发布

✓ 数据导入成功，已经覆盖此评测集原有数据。 不再提示

输入 * 期望输出

	1	2/30 ⓘ	+ 添加行	-
▶	2	2/30 ⓘ	+ 添加行	-

已添加 2/100 上一页48 下一页

This screenshot shows a machine learning testing interface. At the top, there's a purple icon with three horizontal bars and the word '测试' (Test). Below it, status information says '未发布' (Not Published), '测试评测集' (Test Evaluation Set), and '自动保存于 2025-04-07 11:06:34' (Automatically saved on 2025-04-07 11:06:34). There are buttons for '数据导入' (Import Data) and '发布' (Publish). A green checkmark message indicates '数据导入成功，已经覆盖此评测集原有数据。' (Data import successful, it has overwritten the original data in this evaluation set.) with a '不再提示' (No longer remind) link. The main area shows a table with two rows. Row 1 has an ID '1', 2/30 status, and buttons for '+ 添加行' (Add Row) and '-'. Row 2 has an ID '2', 2/30 status, and buttons for '+ 添加行' (Add Row) and '-'. Below the table, there are navigation links: '已添加 2/100', '上一页48', and '下一页'.

评测任务 评测集 评测规则

输入关键词搜索

+ 创建评测规则



创建评测规则

评测规则名称 * 0/128
给评测规则起一个名字吧

评测规则描述 * 0/500
请输入评测规则描述

取消 确定

 测试

评测规则测试 | 自动保存于 2025-04-07 11:09:35

规则类型 *	内容
基础规则 / 是 / 输出部分包含该字段	输入规则内容 -

已添加: 1/10 [+ 添加规则](#) [+ 从评测规则添加](#)

评测任务 评测集 评测规则

输入关键词搜索

评测类型

+ 创建评测任务

创建评测任务

x

4/128

评测任务名称 *

测试任务

评测类型 *

模型

模型 *



qwen-plus-latest

v



+ 添加模型 (1/4)

评测指标 *

● 自动规则打分



测试任务

模型

qwen-plus-latest

杨淳屹

2025-04-07 11:10:46



运行子任务

评测方式 *

自动

模型 *

qwen-plus-latest

知识库

请选择，支持关键词搜索

评测集 *

测试

评测规则 *

测试

QPS值 ① *

1

提示词 * | 从模板选择

11/1000

你是一个智能回答小助手



取消

确定并运行





测试任务

模型评测 2025-04-07 11:10:46 创建

开始日期 - 结束日期 白

运行时间

🕒 2025-04-07 11:12:06 [查看结果](#)

评测方式

自动

状态

已完成

操作



你是一个智...

提示词

测试

评测集名称

1254

总 Tokens 消耗

45.87s

总耗时

测试

100

qwen-plus-latest 得分

自动规则打分指标

qwen-plus-latest

准确率 ⓘ

0.00%

F1 ⓘ

18.00%

ROUGE-1 ⓘ

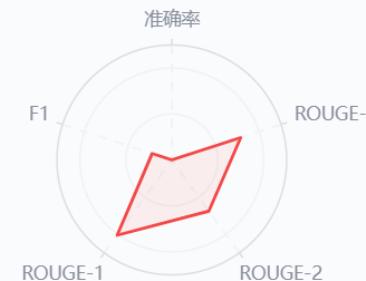
81.00%

ROUGE-2 ⓘ

55.00%

ROUGE-L ⓘ

63.00%



输入

期望输出

智能体输出

评测结果

▶ 1

▶ 2

输入	期望输出	智能体输出	评测结果		
			耗时	Tokens	评测得分
深圳特区是什么时候成立的?	1980年。	qwen-plus-latest	深圳经济特区成立于1980年8月26日。这一天，第五届全国人民代表大会常务委员会第十五次会议批准了《广东省经济特区条例》，正式确立了深圳经济特区的地位。深圳成为中国改革开放的窗口和试验田，从此拉开了快速发展的序幕。	2.488s	97 100
深圳市包含几个行政区?	深圳市包含9个行政区和1个新区，分别是光明区、宝安区、南山区、福田区、罗湖区、盐田区、龙岗区、龙华区、坪山区和大鹏新区。	qwen-plus-latest	截至2023年，深圳市共包含**10个行政区**和**1个新区**。具体如下：### 10个行政区：1. **罗湖区** 2. **福田区** 3. **南山区** 4. **盐田区** 5. **宝安区** 6. **龙岗区** 7. **龙华区**（2017年成立）8. **坪山区**（2017年成立）9. **光明区**（2018年由光明新区改为行政区）10. **大鹏新区**（功能	ROUGE-1 100.00%	ROUGE-2 100.00% ROUGE-L 100.00%

				耗时	Tokens	评测得分
				ROUGE-1	ROUGE-2	ROUGE-L
天山山脉在哪里？	天山山脉位于欧亚腹地，中国境内，天山山脉均在新疆境内，西起乌恰县，东至哈密市星星峡以东。	qwen-plus-latest	天山山脉位于中国新疆维吾尔自治区和中亚地区，是世界上最大的独立纬向山系，也是世界上距离海洋最远的山系。它呈东西走向，横贯新疆中部，将新疆分为南北两部分：以北为准噶尔盆地，以南为塔里木盆地。天山山脉绵延2500多公里，主峰博格达峰海拔5445米。天山不仅是中国重要的地理分界线，	7.235s	162	100
新疆还有哪些山脉？	新疆的北部为阿勒泰山，南部为昆仑山系，天山山脉横亘于新疆中部，把新疆分为南北两半。	qwen-plus-latest	新疆地区地形复杂多样，除了著名的天山山脉外，还有多条重要的山脉。以下是一些主要的山脉：### 1. **阿尔泰山脉** - **位置**：位于新疆北部，延伸至蒙古、俄罗斯境内。 - **特点**：以丰富的矿产资源和优美的自然风光著称，是中国与蒙古的界山之一。阿尔泰山脉的南坡属于准噶尔盆地北缘，北坡则	27.075s	685	100

Homework

- Try to implement an agent with practical application value on Tongji AI.
- Provide an introduction to its functions and screenshots of its application.

Today's lab exercises

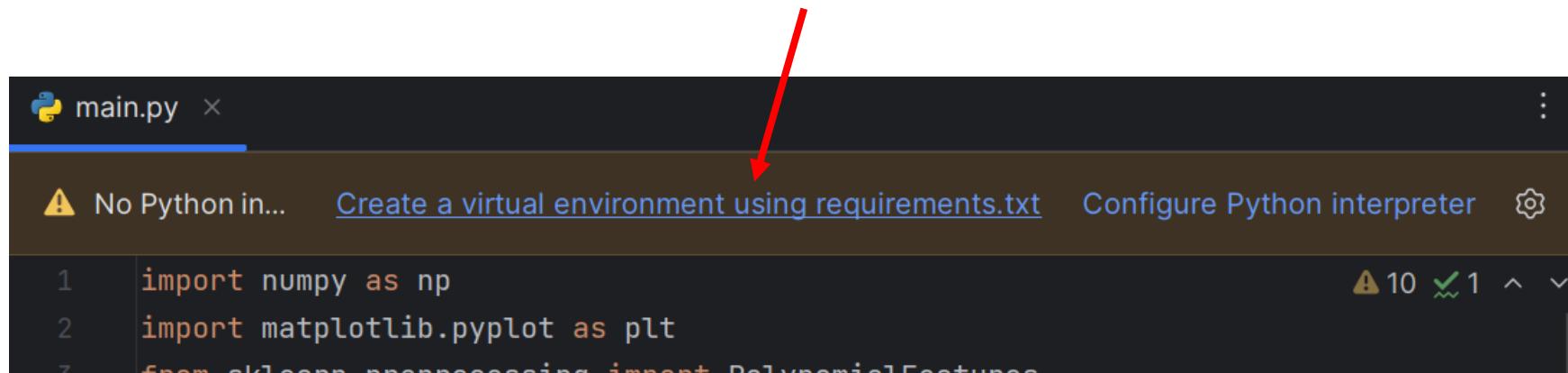
- The exercise consists of two parts, *Evaluation* and *Performance Measure*
- In the Evaluation part, you will complete the following exercises using the existing model (task 2 is optional)
 - 1. Use n-times K-fold cross validation to find the best model
 - 2. Use hold-out and bootstrapping method validation to find the best model
- In the Performance Measure part, you will complete the following exercises using the above experimental results(task 2、3 is optional)
 - 1. Draw PR and ROC map of the best model
 - 2. Draw PR and ROC maps of all models
 - 3. Compare and analyze them

Part 0

Project Description

Experiment environment

- Python 3.x
- Libraries: Numpy & Matplotlib
- Docs
 - Numpy: <https://numpy.org/doc/stable/>
 - Matplotlib: <https://matplotlib.org/stable/tutorials/index.html>
 - scikit-learn: <https://scikit-learn.org/stable/index.html>

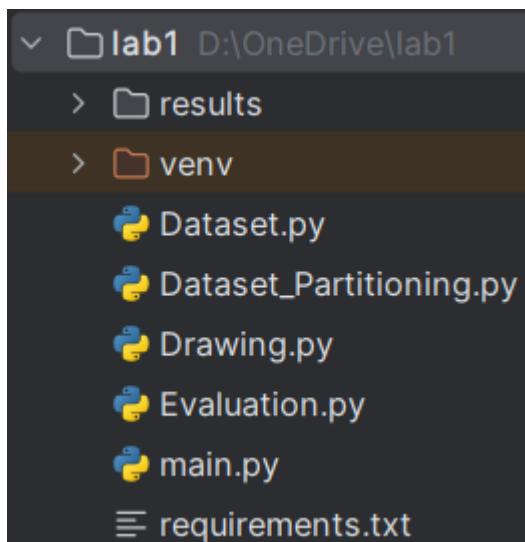


The screenshot shows the PyCharm IDE interface. At the top, there's a dark header bar with the file name "main.py" and a "..." button. Below the header is a toolbar with several icons. A prominent red arrow points from the text in the toolbar down towards the code editor area. The toolbar includes a warning icon followed by the text "No Python in...", a blue link "Create a virtual environment using requirements.txt", and other configuration options. The main code editor window displays the following Python code:

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3 from sklearn.preprocessing import PolynomialFeatures
```

On the far right of the code editor, there are status indicators showing "10" errors and "1" warnings, along with standard scroll and zoom controls.

Project structure



- main.py

Overall Process

- Dataset.py

Dataset Generation

- Dataset_Partitioning.py

Dataset Partitioning Method

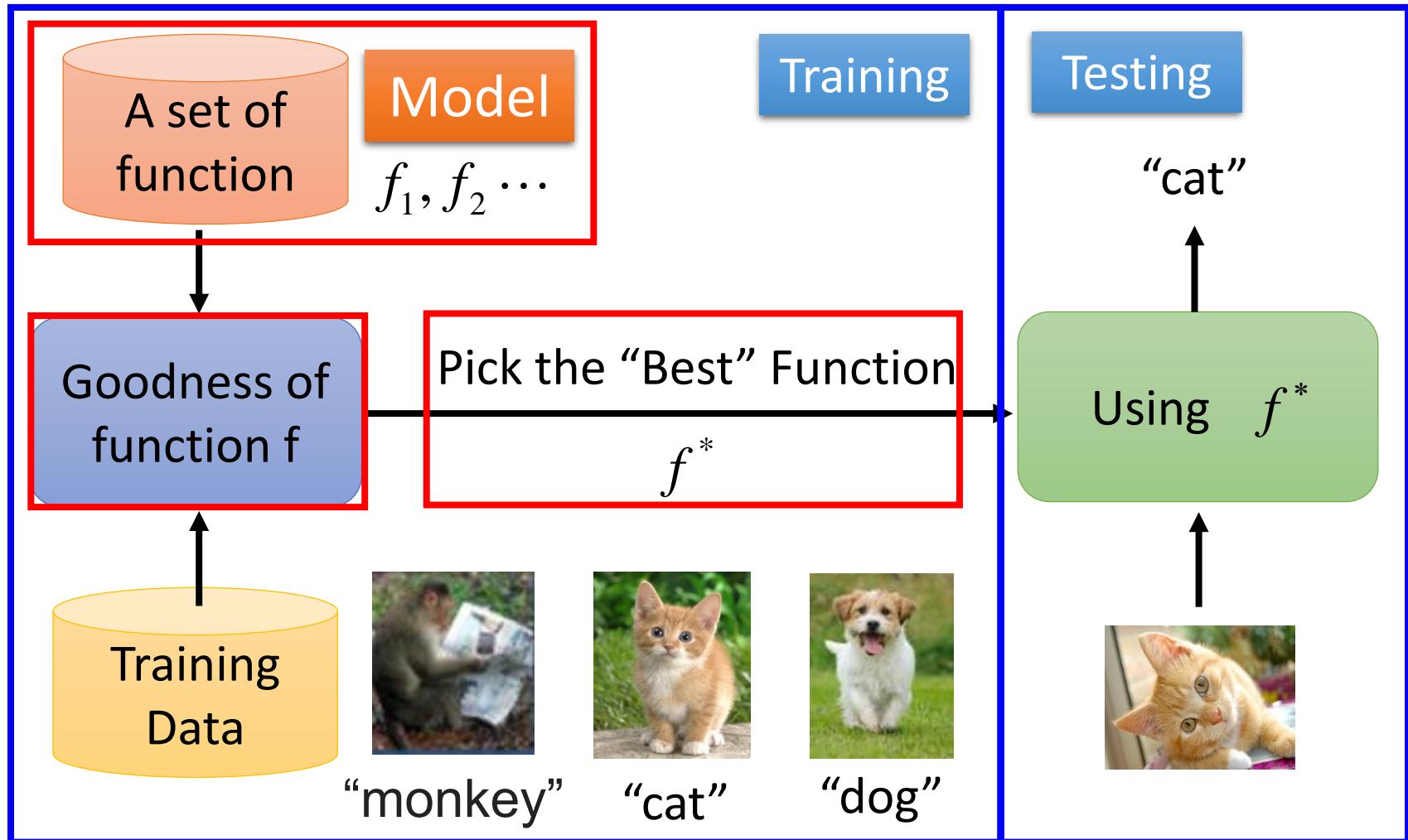
- Evaluation.py

Model Evaluation Method

- Drawing.py

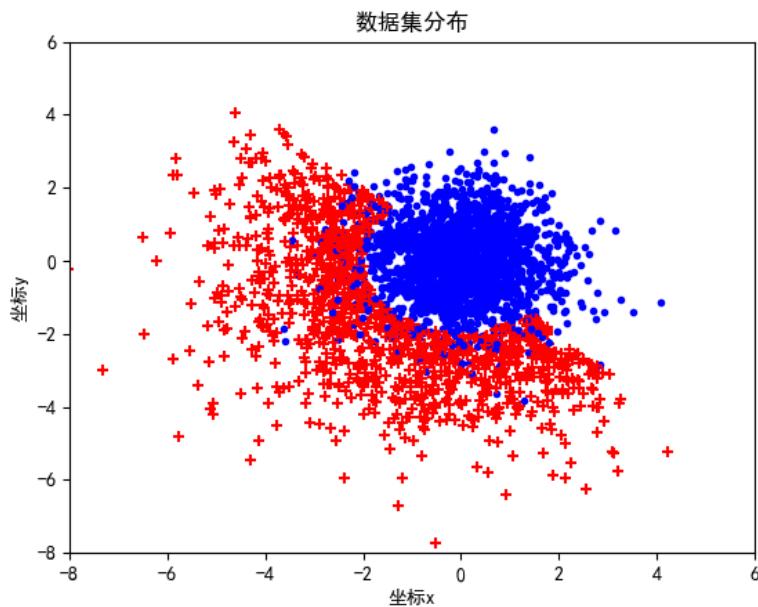
Drawing Method

Review: Framework



Preparing data

```
# 数据集生成  
data = dataset(N) # 可尝试生成不同的数据集  
# drawing_data(data, '数据集分布')
```



- Dataset.py -- dataset(N)
- Class 0 (Blue):
 - normal distribution for both x and y axes.
- Class 1 (Red):
 - normal distribution for both x and y axes.
 - The "upper-right" part and the "center" part are excluded.
- You can try to construct other styles of datasets on your own.

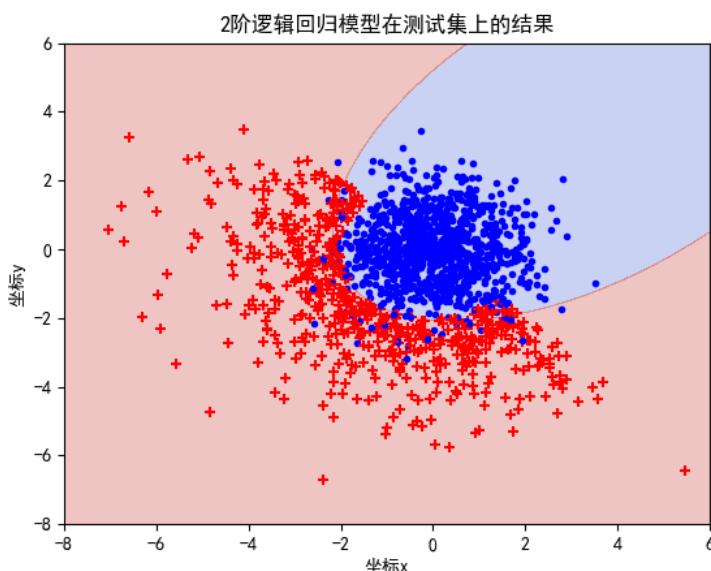
Preparing data

```
# 数据集划分
train_data, test_data = Hold_out(data, test_ratio: 0.5) # 留出法
# train_data, test_data = Bootstrapping(data, 2000) # 自助法
print('训练集样本数量: '+str(len(train_data)))
print('测试集样本数量: '+str(len(test_data)))
```

- Here, the **hold-out** method is used to partition the dataset into training and testing sets. You can attempt to implement the **bootstrap** method for dataset partitioning.
- Observe the format of the dataset on your own.

Model

```
# 定义M阶逻辑回归模型
best_M = 10
# best_M = get_Best_M(train_data, Ms, 'Cross Validation', [T, K])
# best_M = get_Best_M(train_data, Ms, 'Hold Out', [0.3])
# best_M = get_Best_M(train_data, Ms, 'Bootstrapping', [N])
best_model = make_pipeline(*steps: PolynomialFeatures(degree=best_M), LogisticRegression())
```



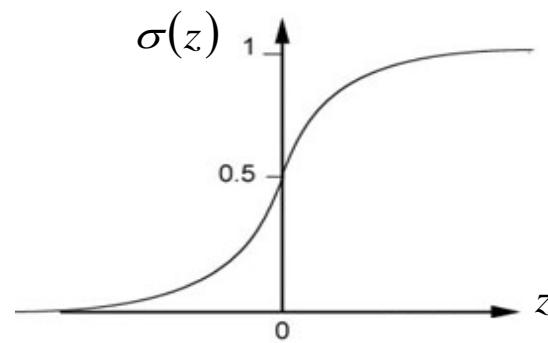
- M-order logistic regression classifier, geometrically speaking, seeks an M-order decision boundary curve in the sample space.
- Here, please utilize different evaluation methods to find the best M among Ms.

Model

```
output = [q for p, q in best_model.predict_proba(test_data[:, 1:])]
```

- Model Output:
- For example, a 1st-order logistic regression classifier:
- $1 / (1 + \text{np.exp}(- (\text{w0} + \text{w1} * \text{x} + \text{w2} * \text{y})))$
- The results are taken from 0 to 1. You can choose a **threshold** value, such that data from 0 to the threshold is classified as "0", and from the threshold to 1 is classified as "1".

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$



Test Result

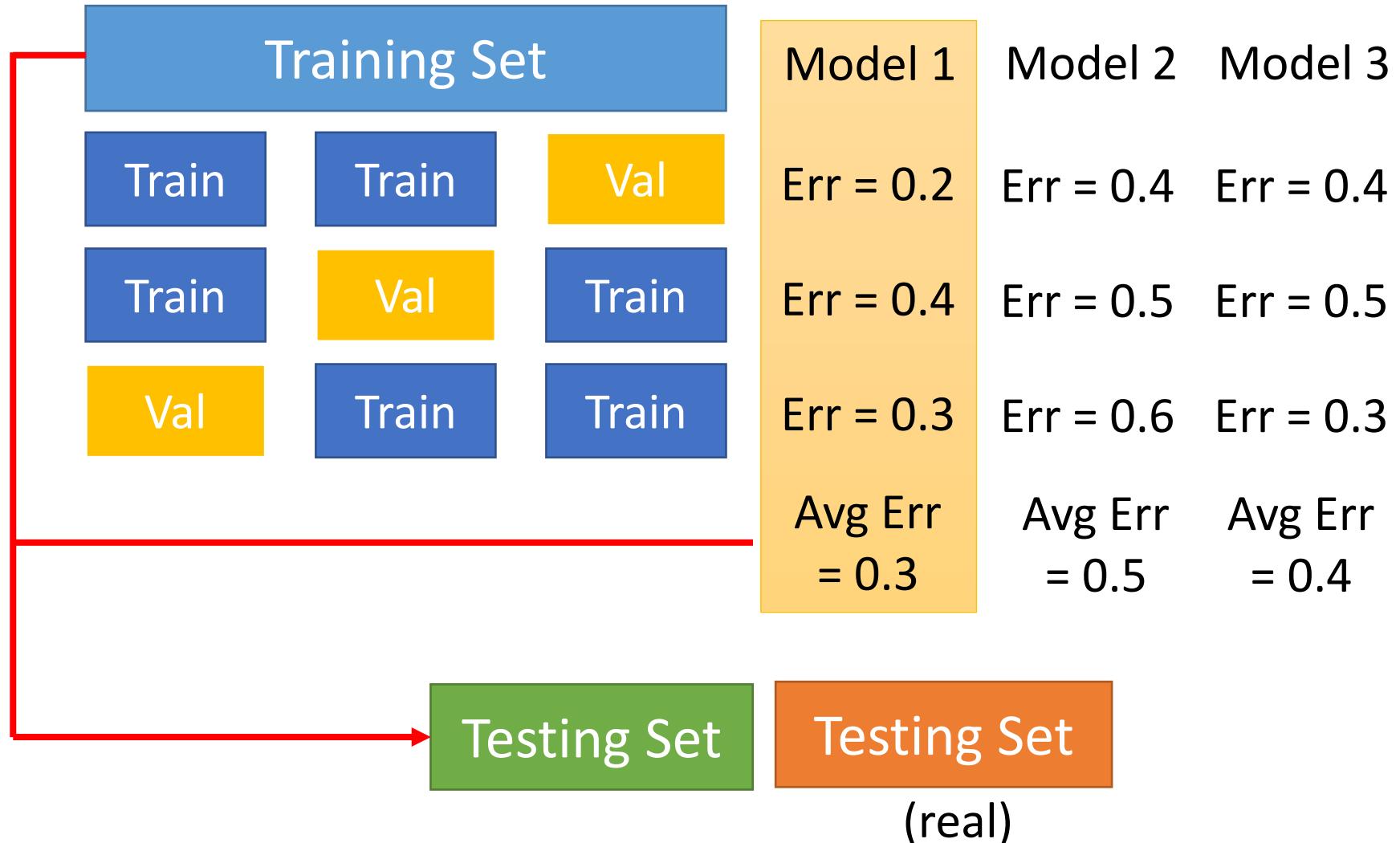
```
# 评价
boolnum = len(test_data)
boolT = 0
for i in range(boolnum):
    if test_data[i][0] == 0:
        if output[i] < 0.5:# 这里以0.5为分界进行分类
            boolT += 1
    elif test_data[i][0] == 1:
        if output[i] >= 0.5:
            boolT += 1

Accuracy = boolT / boolnum # 可尝试其他评价指标
print('测试集的准确率: '+str(round(100 * Accuracy, 3))+'%)
```

- As mentioned earlier, the threshold here is set to 0.5, and "accuracy" is used as the evaluation metric.
- You can also try other evaluation metrics such as precision, recall, true positives, false positive rate, etc.

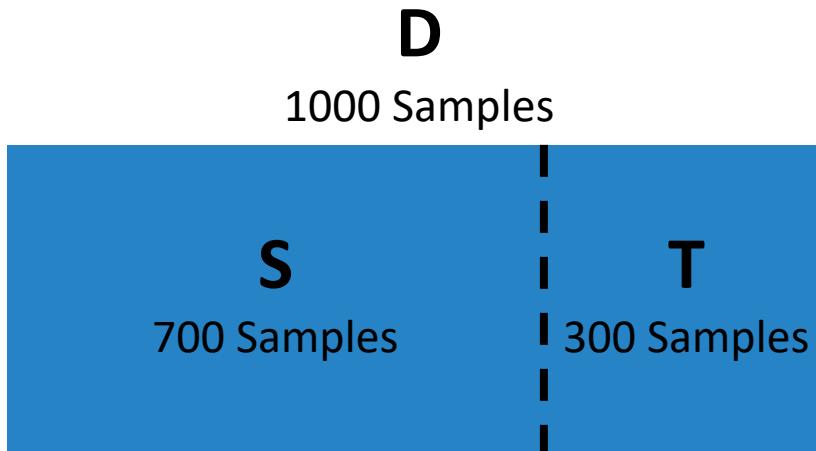
Part 1 Evaluation

Review: K-fold Cross Validation



Review: Hold-out 留出法

- When dividing the data set, try to ensure the **consistency of data distribution** as much as possible!
- 分层采样: 保留类别比例



If

D contains 500 positive samples,
500 negative samples

Then

S: 350 positive samples,
350 negative samples

T: 150 positive samples,
150 negative samples

Review: Bootstrapping 自助法

- Training set $D \rightarrow D'$ (pick a sample from D m times)
- The probability of a sample a not picked is $(1-1/m)^m$
 - $\lim_{m \rightarrow \infty} \left(1 - \frac{1}{m}\right)^m = \frac{1}{e} \approx 0.368$
 - D' : training set
 - $D \setminus D'$: test set
- Advantages
 - It's useful when dataset is small and training and test set are hard to construct
- Disadvantages
 - The training set D' has different distribution of D , which may introduce bias in evaluation

Tasks (45mins)

- Please finish them in Evaluation.py
and Dataset_Partitioning.py

```
1 usage
def get_cv(train_data, Ms, T, K):
    Best_M = 2
    return Best_M
```

```
1 usage
def get_HO(train_data, Ms, test_ratio):
    Best_M = 2
    return Best_M
```

```
1 usage
def get_B(train_data, Ms, times):
    Best_M = 2
    return Best_M
```

```
def get_Best_M(train_data, Ms, method, parameters):
    if method == 'Cross Validation':
        T = parameters[0]
        K = parameters[1]
        return get_CV(train_data, Ms, T, K)
    elif method == 'Hold Out':
        test_ratio = parameters[0]
        return get_HO(train_data, Ms, test_ratio)
    elif method == 'Bootstrapping':
        times = parameters[0]
        return get_B(train_data, Ms, times)
```

```
# fold折交叉验证法的第k次（即取fold份数据中的第k份作为测试集）
1 usage
def Cross_Validation(data, fold, k):
    return
```

```
# 训练样本抽样times次的自助法
2 usages
def Bootstrapping(data, times):
    return
```

Tasks (45mins)

- In the Evaluation part, you will complete the following exercises using the existing model
 - Basic Tasks:
 - Complete the validation set partition process of K-fold cross validation in Dataset_Partitioning.py (15mins)
 - Complete n-times K-fold cross validation method in Evaluation.py to get the optimal model (15mins)
 - Extension Task: (You can do it after finish next part)
 - Complete hold-out and bootstrapping validation methods in Evaluation.py to get the optimal model (15mins)

Prompt

- For each model:
 - For each cross-validation iteration:
 - For each fold in cross-validation:
 - 1. Split the dataset into training and validation sets using `Cross_Validation()` in `Dataset_Partitioning.py`
 - 2. Train the model.
 - 3. Validate the model.
 - Compare the average accuracy of each model and determine the best model.

Expected result

训练集样本数量: 1603

测试集样本数量: 1604

for:

 for:

 for:

 训练集样本数量: 1292

 验证集样本数量: 324

 1阶逻辑回归模型、第1次、第1折的准确率为86.11%

.....

1阶逻辑回归模型3次5折交叉检验的平均准确率为84.53%

.....

5阶逻辑回归模型3次5折交叉检验的平均准确率为94.29%

最佳模型为2阶逻辑回归模型，其在交叉验证法验证集上的平均准确率为94.74%

2阶逻辑回归分类器在测试集上的准确率: 93.92%

Part 2

Performance Measure

Review: Precision and Recall

- Suitable for information retrieval, web search scenarios

A "**confusion matrix**" can be obtained by combining the statistics of the true labels and the predicted results

Ground truth	Predicted Result	
	True	False
True	TP	FN
False	FP	TN

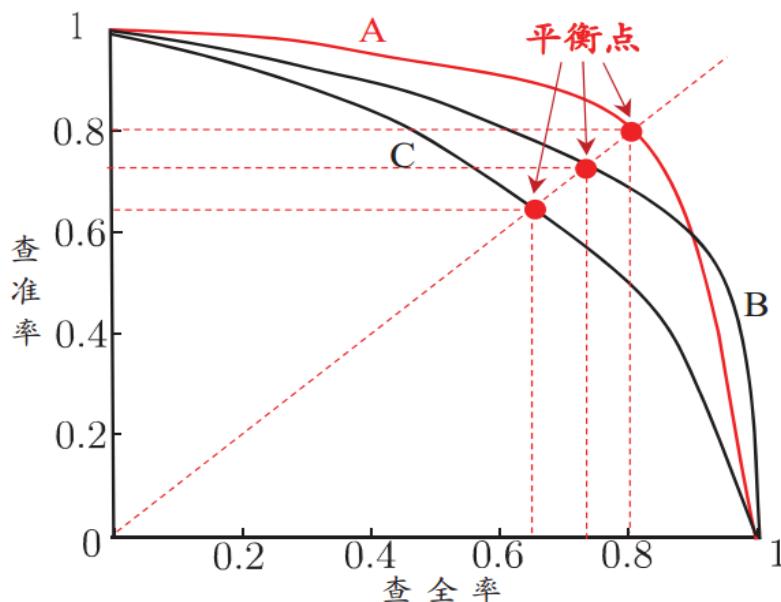
$$\text{Precision} \quad P = \frac{TP}{TP + FP}$$

$$\text{Recall} \quad R = \frac{TP}{TP + FN}$$

Recall and precision are contradictory measures. **Why?**

Review: Precision and Recall

- According to the prediction results of the learner, the samples are sorted according to the probability of positive examples. If all samples are predicted as positive examples one by one, then the precision-recall curve, referred to as "P-R curve", can be obtained.



P-R曲线与平衡点示意图

Break-even point
The value of “precision = recall” on the curve
Used to measure the performance of
classifiers with crossed P-R curves.

Comprehensive
consideration of recall
and precision

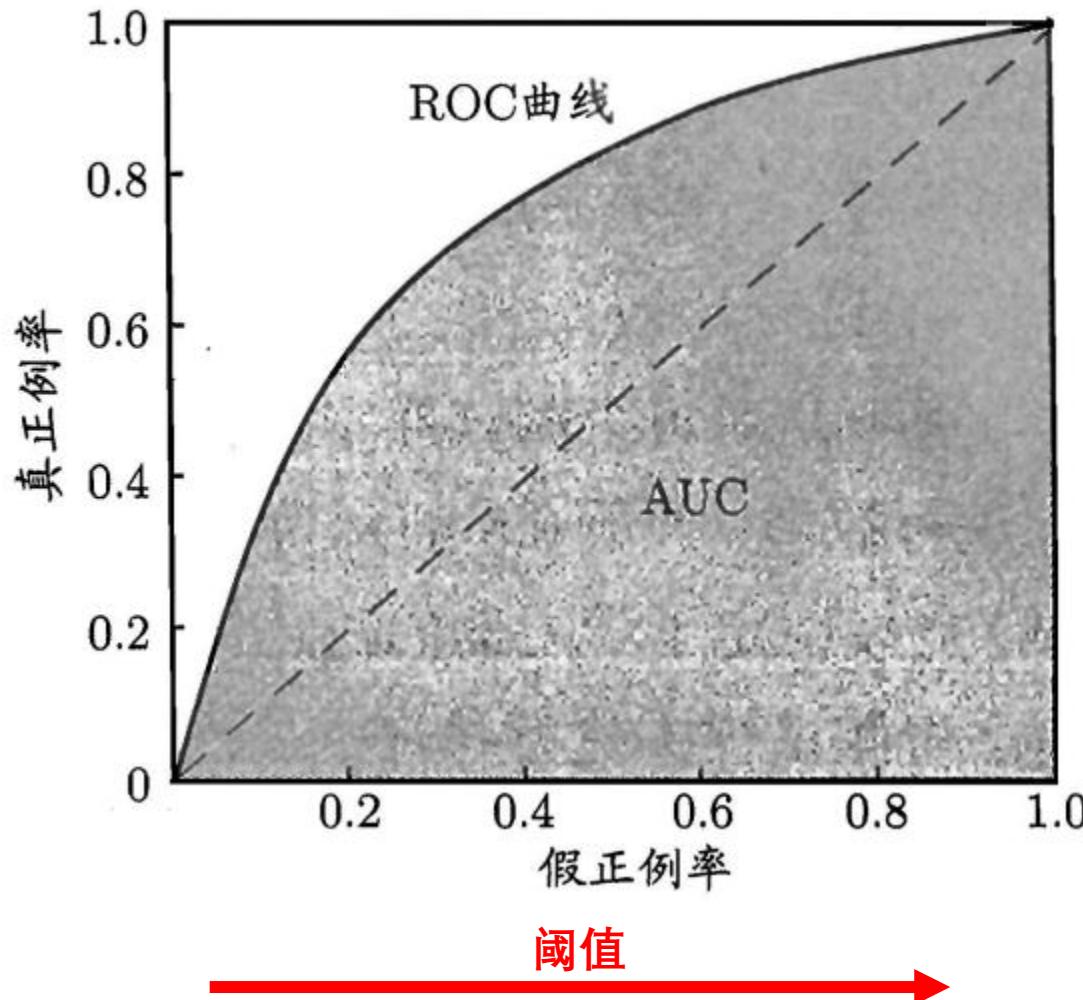
Review: ROC

- Receiver Operating Characteristic(受试者工作特征)
- The steps of sorting the samples and predicting the samples as positive examples are consistent with the PR curve
- But the axes change

	Name	Expression
Vertical axis	True Positive Rate/TPR	$\frac{TP}{TP + FN}$
Horizontal axis	False Positive Rate/FPR	$\frac{FP}{TN + FP}$

Ground truth	Predicted Result	
	True	False
True	TP	FN
False	FP	TN

Review: ROC

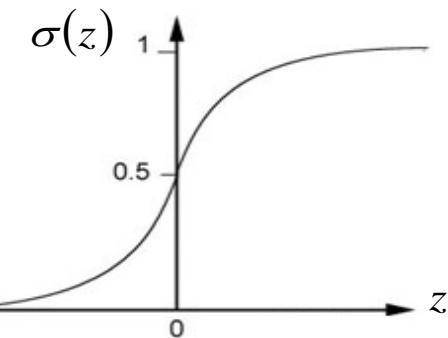


Preparing data

```
# 结果  
output = [q for p, q in best_model.predict_proba(test_data[:, 1:])]  
# output取0~1, 设定阈值, 将其分类0、1两类
```

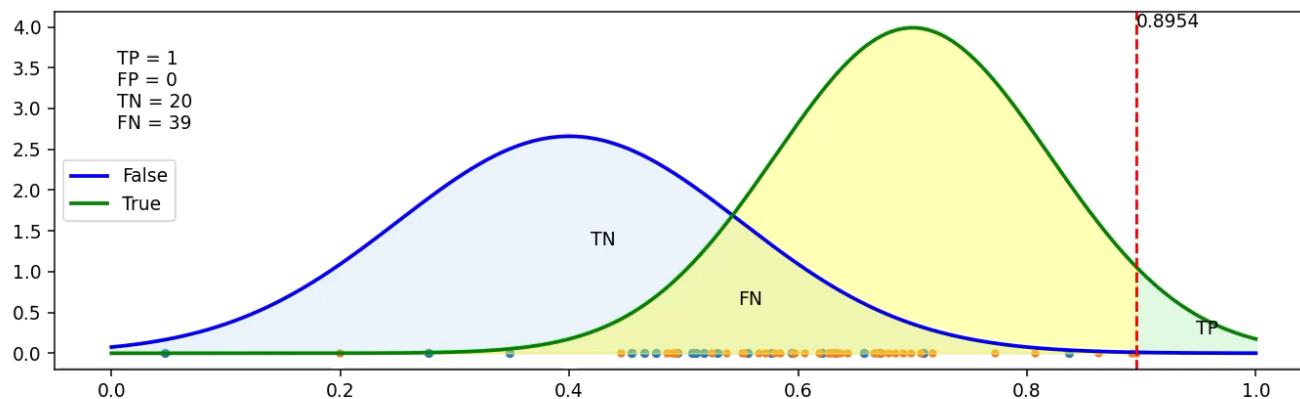
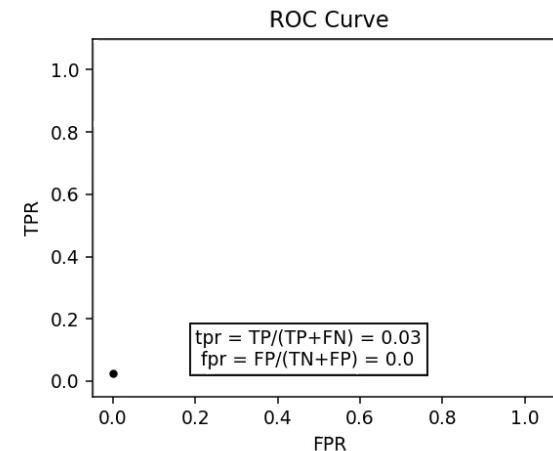
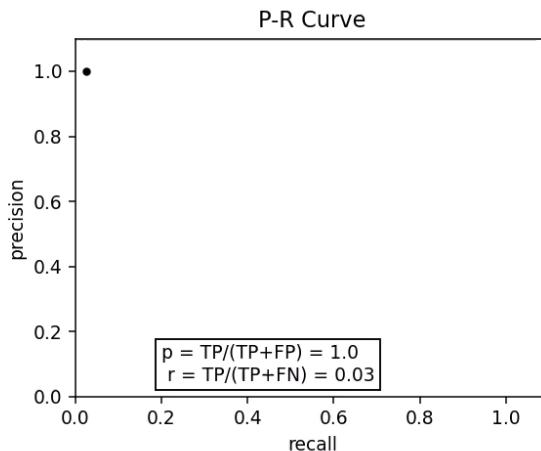
- Observation data structure of ‘test_data’ and ‘output’
- Analyze basic classification decisions:

```
for i in range(boolnum):  
    if test_data[i][0] == 0:  
        if output[i] < 0.5: # 这里以0.5为分界进行分类  
            boolT += 1  
    elif test_data[i][0] == 1:  
        if output[i] >= 0.5:  
            boolT += 1
```



Prompt

The continuous decrease in the threshold will cause changes in the evaluation metrics.



Tasks (45mins)

- In the Performance Measure part, you will complete the following exercises using the above experimental results
 - Basic Tasks:
 - 1. Draw PR and ROC map of the best model (30mins) and *submit*

```
drawing_PR(test_data[:, 0], output, str(best_M) + '阶模型的PR曲线图')  
drawing_ROC(test_data[:, 0], output, str(best_M) + '阶模型的ROC曲线图')
```

- Extension Task:
 - 2. Draw PR and ROC maps of all models (15mins) and submit if you want
 - 3. Compare and analyze them

Tasks (45mins)

- Please finish them in Drawing.py

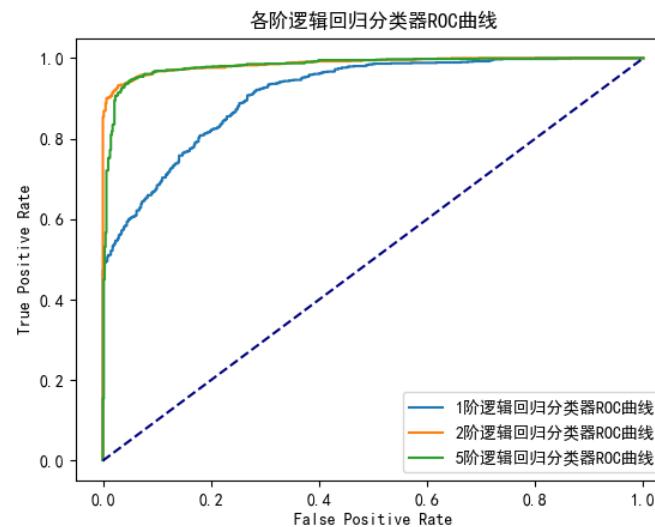
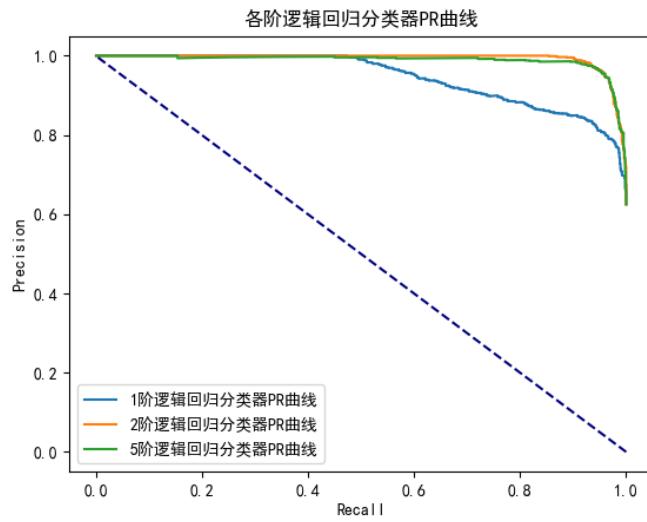
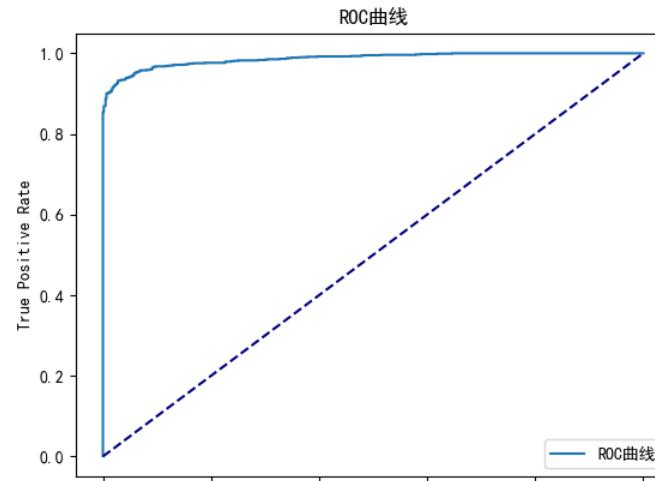
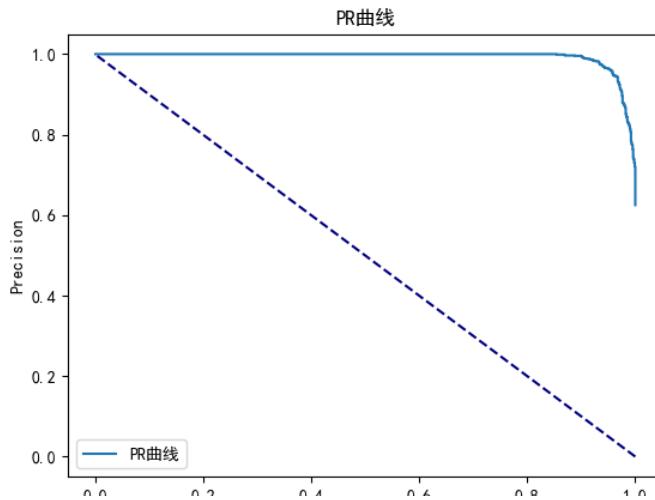
```
1 usage
def drawing_PR(Label, Output, title):
    pass

1 usage
def drawing_ROC(Label, Output, title):
    pass

1 usage
def drawing_PRs(outputs, test_data, Ms, title):
    pass

1 usage
def drawing_ROCs(outputs, test_data, Ms, title):
    pass|
```

Expected result



Thank you for your
attention!