

第 6 周上机作业

时间: 7-9:30PM 11/10/2022

地点: 505/507

班级: 2021 中澳合, 2021 未来网络实验班

截止日期: 23:59 PM 30/10/2022(暂定)

实验目标:

1. 在实际项目中定义自己的类和创建对象
2. 使用 UCI 数据集完成一个基于机器学习的任务, 熟悉机器学习+Java 的实现过程
3. 在 Eclipse 中使用第三方的类库包

实验描述:

一、 基于 UCI 3W 数据集完成一个机器学习的任务

1 准备工作:

1.1 数据集下载链接:

<http://archive.ics.uci.edu/ml/datasets.php> 或者 QQ 群

1.2 第三方的 Java ML 类库下载链接

<https://sourceforge.net/projects/java-ml/>

(可能需要下载其他的依赖库)

1.3 机器学习分类任务或者聚类任务实现

- (1) 基于 3W 数据集, 使用 JavaML 集成的机器学习算法, 完成对数据集的分类或者聚类任务。

或者:

- (2) 基于 3W 数据集, 使用 Java 手动实现一个机器学习算法, 完成对数据集的分类或者聚类任务。

提醒: 数据集需要一定的专业理解或者数据预处理, 请参考网站或者其他资源进行数据集的预处理操作。

1.4 可视化实现 (可选)

基于 1.3 的任务结果, 使用第三方可视化库实现对机器学习结果的可视化显示。

2. 实现要求:

2.1 代码实现要求:

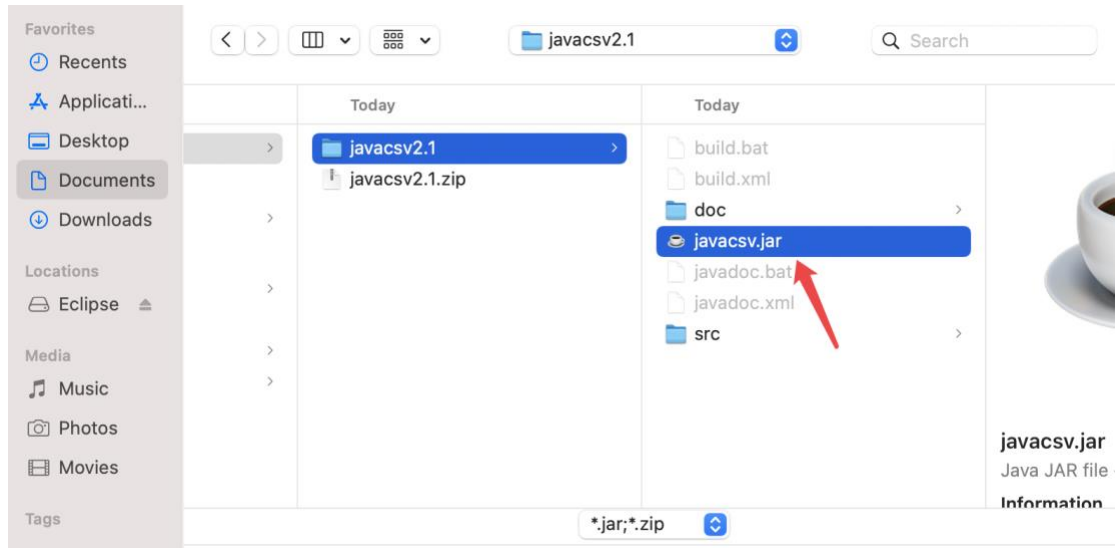
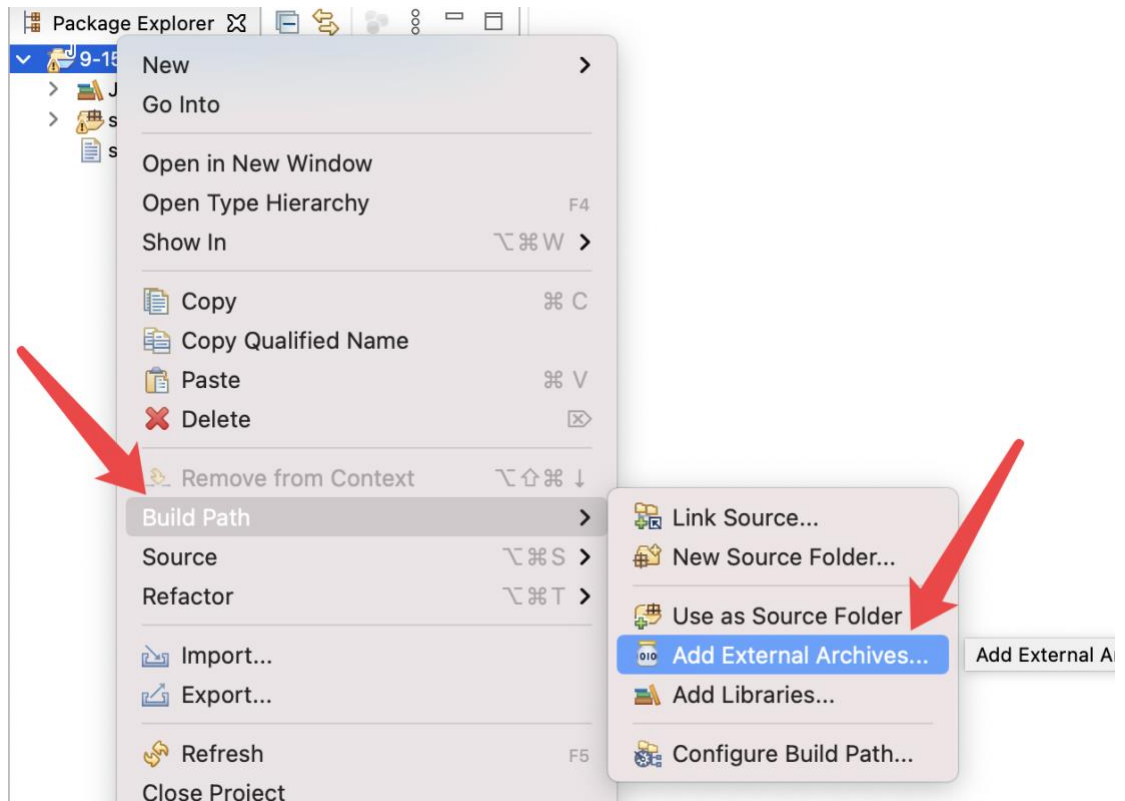
- (1) 可以调包实现机器学习算法或者手动实现机器学习算法。
- (2) 主要功能代码需要使用 Java 语言, 可以进行其他语言的一些代码实现或者调用。
- (3) 不同的功能要封装到不同的方法或者不同的类中。

2.2 第三方类库包引用:

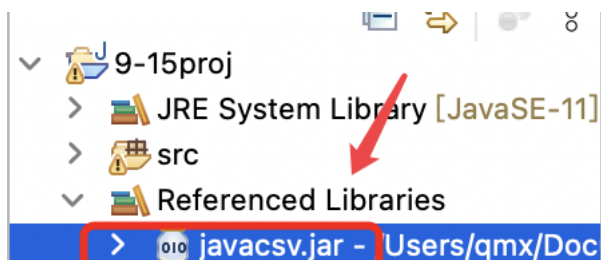
- (1) 以外部导入的类库包形式使用:

步骤 1: 解压 jar 文件的类库包到本地磁盘

步骤 2:



Step4: Project structure is as follows:



(2) Maven 依赖方式使用:

<https://www.cnblogs.com/bnuvincent/p/6940795.html>

备注: 该部分内容在第 10 章介绍。

2.3 作业提交要求:

- (1) 完成一篇类似科研论文的简短图文描述, 具体结构参考附录论文结构 (可以使用 Latex 或者 doc 排版, 最终提交为: 学号姓名.pdf)
- (2) 提交项目文件夹, 命名为: 学号姓名.zip 。需要包含: 项目文件夹 (命名随便)、项目 readme (如果有, 可以附录上 gitee 或者个人博客的地址)

备注: 如果有第三方的代码实现, 请一起提交。

附录:

1. 知乎 ICML-机器学习顶级会议, CCF A 类

<https://www.zhihu.com/topic/25559103/top-answers>

2. Peter Battaglia ICML 22 作品

<https://zhuanlan.zhihu.com/p/534623127>