综合项目说明

综合项目选题

选题一：葡萄酒质量预测

葡萄酒的质量通常是由专业品酒师通过观察葡萄酒的颜色、质感，嗅闻葡萄酒的香气，并结合品尝体验后按评分指标进行打分来评定的。受品酒师能力水平和个人喜好的影响，评定结果会存在一定的主观性。葡萄酒的理化指标与葡萄酒的质量是密切相关的，如果能够找到两者之间的合理关系，那么即使没有品酒师，也能完成对葡萄酒的质量评价并使评价结果更加客观可信。

课题数据集来源于葡萄牙米尼奥大学的一项研究，可从UCI数据集官网下载（https://archive.ics.uci.edu/ml/datasets/Wine+Quality）。数据集中包括两个文件winequality-red.csv和winequality-white.csv，分别为葡萄牙Vinho Verde葡萄酒的1599个红葡萄酒测试样品数据和4898个白葡萄酒测试样品数据，其中输入特征为葡萄酒的11个物理化学指标值：包括：

(1) 非挥发性酸度（fixed acidity），单位：g/dm^3；

(2) 挥发性酸度（volatile acidity），单位：g/dm^3；

(3) 柠檬酸（citric acid），单位：g/dm^3；

(4) 残糖（residual sugar），单位：g/dm^3；

(5) 氯化物（chlorides），单位：g/dm^3；

(6) 游离二氧化硫（free sulfur dioxide），单位：mg/dm^3；

(7) 总二氧化硫（total sulfur dioxide），单位：mg/dm^3；

(8) 密度（density），单位：g/cm^3；

(9) PH值（pH）；

(10) 硫酸盐（sulphates），单位：g/dm^3；

(11) 酒精含量（alcohol），单位：$\%$。

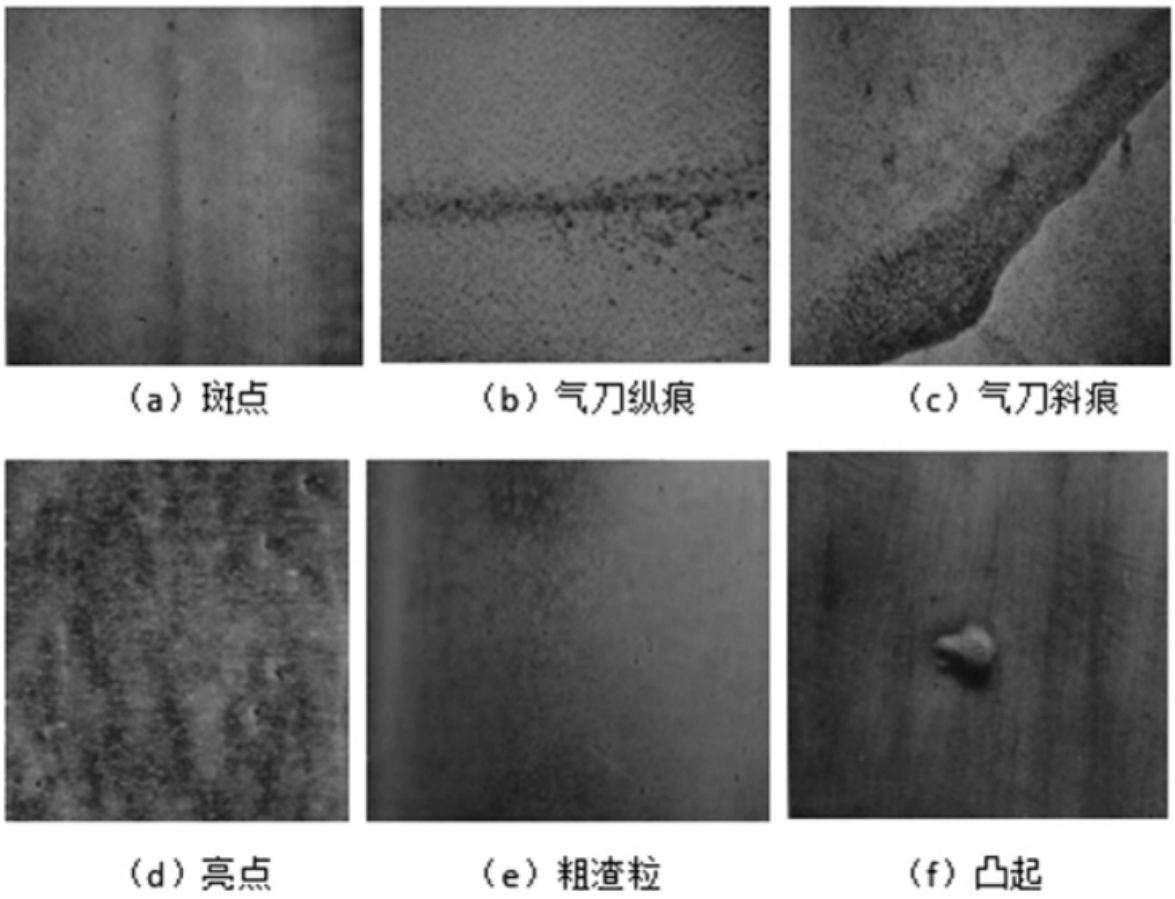
输出特征为葡萄酒的质量评分（score），由三位品酒师通过感官评价给出0~10的评分（0为极坏，10为非常好），选取其中最低的评分。

要求基于该数据集构建葡萄酒质量预测回归模型。

选题二：带钢表面缺陷分类识别

带钢是轧钢企业生产的一种窄而长的钢板，广泛用于各种金属产品、机械产品的生产制造。由于加工环境、加工工艺、设备等原因，带钢表面在生产过程中容易出现诸如气孔、孔洞、擦伤等各种缺陷。

课题数据集来源于美国科罗拉多大学的一项研究，可从UCI数据集官网下载（https://archive.ics.uci.edu/ml/datasets/Steel+Plates+Faults）。数据集采集自企业生产实际，包含了1941个样本，从样本缺陷图像中提取了灰度直方图、纹理特征、投影量、图形状这四类共27个特征，分为七种缺陷类型：(1) 斑点（Pastry）；(2) 气刀纵痕（Z_Scratch）；(3) 气刀斜痕（K_Scatch）；(4) 亮点（Stains）；(5) 粗渣粒（Dirtiness）；(6) 凸起（Bumps）；(7) 其它（Other_Faults）。前六种缺陷如下图所示。



要求基于该数据集构建带钢缺陷分类模型。

选题三：污水处理厂运行状态聚类分析

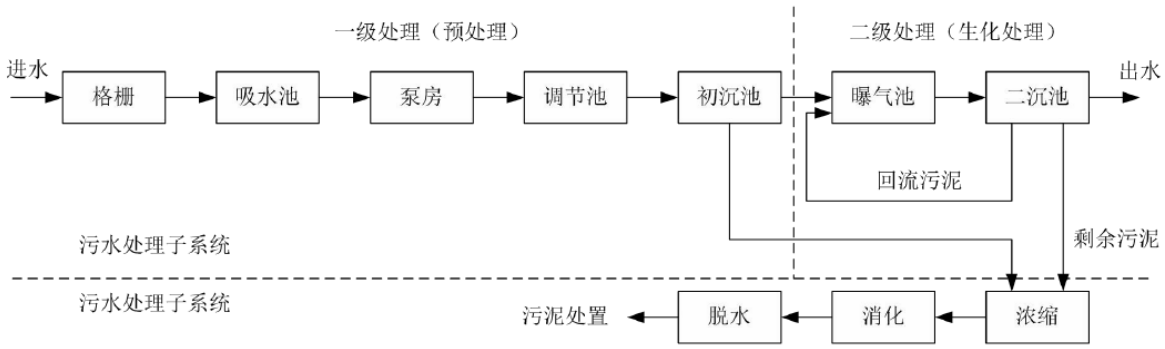
污水处理是将城市生活、工业生产等活动中产生的污水进行加工处理，使其达到国家规定的出水指标的过程，污水处理工业在水资源可持续发展中占据着重要的一环。在污水处理过程中，由于进水流量、进水组分、污染物种类、天气变化等都是被动接受，多种因素都会对污水处理过程产生影响，因此保持污水处理厂的长期稳定运行十分困难。如果不能及时监测到污水处理过程异常工况，可能导致污水处理能力下降、出水水质不达标，甚至引发污水处理过程崩溃，导致不可逆的环境污染事故发生。

课题数据集来源于西班牙巴塞罗那大学的一项研究，可从UCI数据集官网下载（https://archive.ics.uci.edu/ml/datasets/Water+Treatment+Plant）。该数据集于1990年开始历时两年时间在西班牙Manresa镇上的一个污水处理厂中以平均每天1个样本的频率采集得到，共527个样本。每个样本由38个特征描述，其中29个特征值是通过污水处理厂的传感器测量获得，9个特征值是通过实验室分析获得，各特征的名称及含义如下表所示。

序号	名称	含义	序号	名称	含义
1	Q-E	进水流量	2	ZN-E	进水中锌的含量
3	PH-E	进水酸碱度	4	DBO-E	进水生化耗氧量
5	DQO-E	进水化学耗氧量	6	SS-E	进水悬浮固体含量
7	SSV-E	进水挥发性悬浮固体含量	8	SED-E	进水沉淀物含量
9	COND-E	进水导电率	10	PH-P	初沉池进水酸碱度
11	DBO-P	初沉池进水生化耗氧量	12	SS-P	初沉池进水悬浮固体含量
13	SSV-P	初沉池进水挥发性悬浮固体含量	14	SED-P	初沉池进水沉淀物含量
15	COND-P	初沉池进水导电率	16	PH-D	二沉池进水酸碱度
17	DBO-D	二沉池进水生化耗氧量	18	DQO-D	二沉池进水化学耗氧量
19	SS-D	二沉池进水悬浮固体含量	20	SSV-D	二沉池进水挥发性悬浮固体含量
21	SED-D	二沉池进水沉淀物含量	22	COND-D	二沉池进水导电率
23	PH-S	瀑气池酸碱度	24	DBO-S	瀑气池生化耗氧量
25	DQO-S	瀑气池化学耗氧量	26	SS-S	瀑气池悬浮固体含量
27	SSV-S	瀑气池挥发性悬浮固体含量	28	SED-S	瀑气池沉淀物含量
29	COND-S	瀑气池导电率	30	RD-DBO-P	初沉池进水平均生化耗氧量
31	RD-SS-P	初沉池进水平均悬浮固体含量	32	RD-SED-P	初沉池平均进水沉淀物含量
33	RD-DBO-S	二沉池进水平均生化耗氧量	34	RD-DQO-S	二沉池进水平均化学耗氧量
35	RD-DBO-G	整个处理过程的平均进水生化耗氧量	36	RD-DQO-G	整个处理过程的平均进水化学耗氧量
37	RD-SS-G	整个处理过程的平均进水悬浮固体含量	38	RD-SED-G	整个处理过程的平均进水沉淀物含量

该污水处理厂采用的活性污泥法运用活性污泥中包含的微生物来对污水进行净化处理，是一种广泛使用的污水处理工艺，其主要工艺流程如下图所示。原始污水经过第一级预处理后，不溶于水的污染物沉淀到初沉池中，污水由初沉池出水到曝气池，进入第二级生化处理环节。曝气池中有着含大量微生物的活性污泥生物絮团，能对污水中的有机污染物进行吸附和氧化分解，转化为无污染的无机物，然后污水进

入二沉池，此时无机物开始沉淀，顶部清水即为净化完毕的污水，可以排放进入江河湖泊等接纳水体，大部分污泥回流到曝气池，多余部分则排出。



要求基于该数据集构建聚类模型，通过聚类结果对污水处理厂的运行状态进行分析。

任务点

任务点1：问题分析及文献调研

要求：每组选择一个选题，进行问题分析和文献调研，撰写实验报告的引言部分，内容包括：实验对象的详细描述；实验任务重要性阐述；针对实验任务，通过查阅文献，综述解决该实验任务的方法有哪些（需要添加参考文献的编号）。

任务点2：模型设计

要求：设计模型及实验流程、方法，撰写实验报告的实验步骤部分。

任务点3：程序实现（4学时）

要求：完成实验验证，对实验结果进行分析，完成实验报告的撰写。

任务点4：课堂演示（提交实验报告和课堂演示文稿）

要求：制作课堂演示文稿，完成课堂演示，回答同学和老师的提问。