**2025春机器学习大作业**

**（计算机选修）**

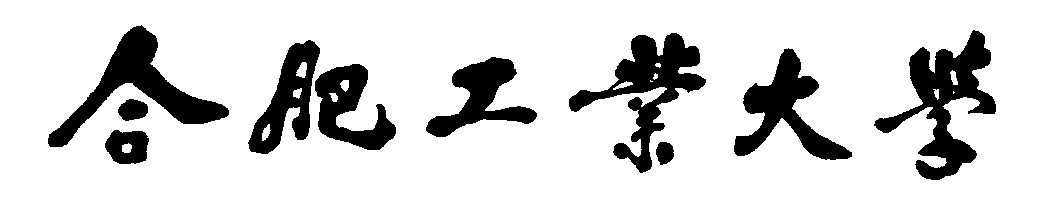
1. **提交要求**

* **提交内容：**
  1. 代码：
     + 可使用开源工具库；
     + **提交ipynb文件**，命名为【25春机器学习】学号+班级+姓名+代码.ipynb；
  2. 报告：
     + 报告不低于8页；
     + 内容需要包括可视化分析结果（即需要包含图表）；
     + **提交word文件**，命名为【25春机器学习】学号+班级+姓名+报告；
  3. PPT：
     + 报告对应的PPT文件，不低于8页；
     + **提交PPT文件**，命名为【25春机器学习】学号+班级+姓名+汇报。
* **提交方式**：以上三项内容**压缩打包**，命名为【25春机器学习】学号+班级+姓名+大作业.zip，提交至**1350650232@qq.com**；
* **邮件主题：**【25春机器学习】学号+班级+姓名+机器学习大作业。

1. **选题**

* 于腾讯文档完成选题<https://docs.qq.com/sheet/DWWhXQW9TREZuT01R>，鼓励1、2、3自拟题目。

1. 自行设计的解决实际问题的题目。
2. 参加国内外机器学习相关项目/比赛。
3. 在kaggle等平台上实际参加的工程项目。
4. 收集一定数量的学生成绩数据（例如10个男生，10个女生，10门课），根据成绩数据，预测该学生是男生还是女生。
5. 针对合肥某个特定小区，在合肥二手房网站上，收集这个小区历史的房屋月均价数据（例如从去年5月到今年3月），预测下个月的房价。
6. 基于scikit-learn软件包，对mnist数据集合进行分类，并可视化。至少要尝试5种不同的分类方法，并比较分类结果。
7. 收集一定量的猫和狗的图片（例如100个猫的图片，100个狗的图片），训练一个分分类器，能够判断输入的图片中是猫还是狗。
8. 从<http://archive.ics.uci.edu/ml/> 上下载任意一个数据集，然后根据其问题类型，用相应的软件包，构建相应的分类器或者回归模型。
9. 收集合肥地区过去一段时间（例如过去一年每个月的平均值）的空气质量（例如pm2.5值），然后构建回归模型，能够预测今年某个月的空气质量值。
10. 收集一定数量的男和女的照片，训练一个分类器，能够直接根据照片分类出男女。
11. 基于字体的性别检测：收集一定数量的男生的手写字（图片）和一定数量的女生手写字（图片），然后训练一个分类器，能够直接根据这个手写字的照片分类出男女。
12. 基于深度学习的面部表情分类：自行查找关于深度学习的资料，基于pytorch等深度学习系统，构建一个面部表情分类，可以检测静态图片中的人脸面部表情，或者面部表情实时监测系统。
13. 新闻文本分类：从新闻网站上，收集各类新闻文本（5-10种），根据课堂上讲解的one-hot特征提取方式，或者自行查找文本特征提取方式，对文本进行提取特征并向量化，利用机器学习模型进行分类，构建一个新闻文本分类系统。
14. 垃圾邮件分类系统：从邮件系统中，收集垃圾邮件和非垃圾邮件每类50-100组，根据课堂上讲解的垃圾邮件检测技术，进行垃圾邮件分类，构建垃圾邮件分类器。
15. 股票价格预测：从股票网站上，收集某股票的历年价格，对历年每月（或每天）价格数据进行建模，预测其下月（或下一天）的股票价格。考虑股票价格和时间的关联。
16. 饭卡消费预测系统：收集自己或同学的饭卡消费数据，从校园网账户中，对每月（每天）的消费数据进行建模，预测下月（或者下一天）的饭卡消费。考虑饭卡消费和月份的关联。
17. 水果检测：设计一个水果检测系统，可以对3-4种水果进行分类，收集3-4种水果的照片，每种水果收集20-50张照片，用机器学习模型进行学习训练，构建桌面应用或者移动端应用，推荐开发移动端应用，实现水果的实时检测与分类。
18. 基于机器学习的图片自动拼接：自行查找相关模型算法，构建一个能够将两个图片进行自动拼接为一个完整图片的系统，类似于全景照片的拼接过程。



**机器学习大作业**

**【题目】**

课程：

姓名：

学号：

班级：

日期：

1. 研究背景

说明本工程所涉及的研究背景，包括主要涉及领域，主要研究方法，主要存在问题，现有解决方案等等，可自行补充其他内容。

1. 模型方法

**详细**说明本工程所使用的模型方法和理论。

1. 系统设计

系统的详细设计，系统流程，系统的每一步的具体流程，例如，如何处理语料，如何训练模型，如何测试模型，最后对模型进行评估。

1. 系统演示与分析

对系统演示结果进行说明，测试一些课程中演示的样例，根据结果说明为什么对或者为什么错。最后对系统提出改进方案。

五、对本门课的感想、意见和建议。

**代码验收成绩**

|  |  |  |
| --- | --- | --- |
| 运用计算机科学与技术及其应用领域工程的思想与方法，从需求分析开始，独立完成机器学习系统的设计与开发，兼顾机器学习系统开发的社会评价。（10分） | 理解机器学习系统运行应遵守的社会规范，能够在机器学习系统开发实践中自觉遵守工程职业道德和规范，履行责任。（10分） | 理解机器学习系统开发相关的工程配置管理方法，掌握领域方向系统项目管理能力。（10分） |
|  |  |  |

**报告成绩**

|  |  |  |
| --- | --- | --- |
| CO1 运用计算机科学与技术及其应用领域工程的思想与方法，从需求分析开始，独立完成机器学习系统的设计与开发，兼顾机器学习系统开发的社会评价。（50分） | CO2 理解机器学习系统运行应遵守的社会规范，能够在领域方向系统开发实践中自觉遵守工程职业道德和规范，履行责任。（10分） | CO4 理解机器学习系统开发相关的工程配置管理方法，掌握领域方向系统项目管理能力。（10分） |
|  |  |  |

**总成绩：**

**指导教师评语（验收/报告）：**

**指导教师签名：**

**日期：**