

# 机器学习大作业

这门课程的主要目标是希望大家将机器学习算法应用于实际任务中，并具备相应的算法开发基础。

## 项目要求

1. 选择一项任务，使用课程上讲过的机器学习方法进行，可以使用信号处理或特征提取方法进行辅助。
2. 给出相应任务结果，训练集和测试集精度，以及特征降维后的可视化。

## 项目主题

首先，需要选择一个项目主题，可以从下面两种项目中选择一个：

1. 应用项目：选择感兴趣的应用场景，探索如何最好地应用机器学习算法解决它。
2. 算法项目：针对目前机器学习算法存在的一些问题，对现有算法进行改进或开发新的算法来解决这些问题。

当然，也有一些项目需要结合应用和算法两方面的任务。

一些可能的项目主题：

1. 无人机目标检测

具体介绍&dataset: <http://aiskyeye.com/>

2. 情感分析：可以将用户情感分为正面、负面或中性。包括视觉情感识别和语音情感识别。

视觉情感识别：

dataset: <http://ai.stanford.edu/~amaas/data/sentiment/>

语音情感识别：使用音频数据对说话者的情绪进行分类，例如快乐、悲伤、

惊讶、愤怒等。

dataset: <https://drive.google.com/file/d/1wWsrN2Ep7x6lWqOXfr4rpKGYrJhWc8z7/view>

3. 工业数据分析：包括故障诊断、寿命预测等

dataset: <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/>

4. 电影推荐系统：根据目标客户的兴趣推荐产品。

dataset: <https://drive.google.com/file/d/1Dn1BZD3YxgBQJSljbfNnmCFIDW2idQGD/view>

5. 虚假新闻识别系统：虚假信息传播迅速，在当前时代是一个大问题，学习如何将假新闻和真实新闻区分，可以使用监督学习实现。

dataset: <https://drive.google.com/file/d/1er9NJTLUA3qnRuyhfzuN0XUsolC4a-q/view>

当然也可以选择其他任务，例如音乐/音频分析、自然语言处理、医学影像分析等，希望同学们选择一些感兴趣的方向和任务。

## 最终提交：项目主题、大作业实验报告和 PPT 学术报告

1. 项目主题。具体包括以下信息：

动机：想解决什么问题？

方法：打算使用哪些机器学习技术？

预期实验：打算进行什么实验？如何评估自己的机器学习算法？

提交 dataset 链接，自己了解到的相关研究示例，以及自己实现的算法代码。

2. 实验报告：包括方法、实验流程、结果和分析等，同学们也可以尝试将自己完成的项目作业形成一篇学术论文。

3. 最后每位同学需要整理自己所做的工作，并在课堂上做一个 5 分钟学术报告，方便同学们了解彼此的工作，并进行后续的合作。

## 作业提交：

1. 提交时间：2023 年 5 月 17 日；

2. 提交方式：将实验报告、讲解 PPT、算法代码等材料压缩打包后以姓名+学号为文件名，发送给王梦宇老师。