



# 大作业要求



- 本次课程大作业为**单人**课题，不可组队。每个人从以下两题**任选一题**，独立完成项目设计、代码开发、实验验证、报告撰写。
- 作业目的：
  - 旨在要求学生通过掌握的机器学习基本知识以及课程上所运用的数学优化方法，**从0实现经典分类算法。不可使用现有的机器学习库函数。**
- 提交要求：
  - PDF格式实验报告及其latex源码以及代码。
  - 代码部分需要提交**全部**的训练与测试代码以及可执行，供助教运行检查。
- 评分标准：
  - 1) 设计的合理与新颖性;
  - 2) 代码的正确与整洁;
  - 3) 实验的充分以及性能;
  - 4) 报告的方法陈述质量与实验结果论证质量。
- 开发语言不限。
- 图片的特征提取不能使用神经网络！大作业全过程不能使用神经网络！
- **抄袭必究!!! -互相抄袭，与从互联网抄袭同样惩罚。**
- 截止日期：**第16周结束的周日晚上24点。**不会延期！



# SVM的实现以及核函数探究



## 题目一：SVM的实现以及核函数探究

### 层次一

从0实现SVM的训练与测试代码，并在MNIST，CIFAR数据集进行实验。

结果应包括分类精度、效率【训练、测试】、多类别策略、参数消融实验等。

报告应通过数据表、图表、讨论等形式反映算法的性能。

重点是训练优化算法的实现与调优。

### 层次二

多核SVM比起单核SVM具有高精度的优势。

请构建一个多核学习的SVM，核的种类不限，层次一中的测试项重复一遍，并且添加多核的种类、多核的组合公式等多核相关参数的消融实验等。

重点是多核组合方式与组合系数优化的算法构建与调优。

注：由于CIFAR10数据集的训练开销部分同学可能无法承担。在本次大作业中，请同学取原每类训练集前500样本为训练集，原每类测试集前100样本为测试集。共计5000张训练图片与1000张测试图片。



# 随机森林的实现以及节点分类函数探究



## 题目二：随机森林的实现以及节点分类函数探究

### 层次一

从0实现随机森林分类器的训练与测试代码，并在MNIST，CIFAR数据集进行实验。结果应包括分类精度、效率【训练、测试】、多类别策略、参数消融实验等。报告应通过数据表、图表、讨论等形式反映算法的性能。重点是节点分类的高效实现。

### 层次二

探索将决策树的分类函数从简单的单维度二分法，扩展为复杂的多维度分类函数，分类种类不限，层次一中的测试项重复一遍，并且添加节点分类函数种类等相关参数的消融实验等。重点是节点分类函数复杂度与整体算法复杂度的折中设计。

注：由于CIFAR10数据集的训练开销部分同学可能无法承担。在本次大作业中，请同学取原每类训练集前500样本为训练集，原每类测试集前100样本为测试集。共计5000张训练图片与1000张测试图片。