

# 《模式识别与机器学习》课程设计任务书

参考教材：

- 1、《模式识别》（第3版），张学工，清华大学出版社，2010年。
- 2、西奥多里德斯 著（李晶皎等译），模式识别（第三版），电子工业出版社，2006年。
- 3、机器学习，周志华，清华大学出版社，2016年

## 一、课程设计的目的

《模式识别与机器学习》课程是智能科学与技术等专业教学计划中以应用为基础的一门专业课，是研究如何用机器去模拟人的视觉、听觉、触觉等感觉器官以识别外界环境的理论与方法，《模式识别与机器学习》课程设计的目的是使学生掌握统计模式识别的基本分类方法的算法设计及其验证方法，通过设计性实验的训练，以提高学生设计算法及数值实验的能力，进一步提高分析问题、解决问题的能力。通过本课程设计，学习利用监督或非监督学习方法对生活中的实际问题进行识别分类，掌握模式识别系统的基本设计思路与步骤。

## 二、课程设计的基本内容

1) 观察生活与环境，自选一个题目，至少采用一种算法进行分类或回归。（贝叶斯决策、逻辑回归、Fisher线性判别、感知准则方法、SVM或 K<sub>n</sub> 近邻法、K-L变换法等）。

2) 数据源可以自选，也可以参考UCI数据集：

<http://archive.ics.uci.edu/ml/>。

但要求所选的数据集至少有一个要满足：至少包括300个实例，特征数至少8个以上。

如：回归数据集：load-diabetes（糖尿病），442个患者的10个生理特征；boston房价等。

分类数据集：breast\_cancer，手写数字识别字符集等。

3) 对于分类算法用准确率，ROC曲线等进行评价，对于回归采用均方误差（Mean Squared Error）、均方根误差（Root Mean Squared Error）等进行评价。

4) 编程语言任意，算法必须自己实现，不能使用任何的机器学习工具包，评价函数可以调包。

### 三、课程设计的基本要求

1、每人独立完成课程设计任务

2、课程设计报告内容要求：

(1) 问题描述，选用某种方法的理由，模式采集，特征提取与选择，分类器设计，学习过程，测试结果，结果分析（含不足与展望），设计总结。程序代码作为附录与报告一起提交。

共性的要求如下：

- a. 必须要有评价指标的实现
- b. 必须要有参数的选择和实验分析

选做的题目要求如下（任选一个）：

- a. 实现对标准分类算法的改进，改进的算法与标准算法在 2 个（或以上）的数据集上进行实验比较。如：设计一个算法，能通过动态调整学习率显著提升收敛速度，编程实现该算法，并选择两个 UCI 数据集与标准的 BP 算法进行实验比较，如：BP 算法的实现与改进。
- b. 实现一个模式识别系统，该系统包括有特征提取（选择）和分类算法，比较有特征提取（选择）和无特征提取的分类结果，如：基于 XXX、XXX 的模式识别系统的设计与实现。
- c. 实现 2 种以上分类算法，将算法在 2 个（或以上）数据集上比较有特征提取（选择）和无特征提取的分类结果，如：朴素贝叶斯分类器与 KNN 分类器的比较。
- d. 实现基本分类器算法和分类器的集成算法，并将其进行对比。
- e. 上述四题为基本要求，可以任意组合。

(2) 内容应层次分明，文字简练，说明准确，数据可靠，结构合理，表格图规范，符合国家标准。

3、课程设计报告格式要求：

a. 摘要

b. 目录

c. 正文

**概述：**研究内容和数据集的描述

**算法描述：**说明所选算法的特点，为什么该算法适用于所设计的系统；说明所采用的算法的主要思想和实现的基本步骤。

**实验结果与分析：**训练过程中的参数是如何设置的、可能需要训练样本数量，与其他算法方案进行比较的结果，分析为什么会有这样的结果

d. 结论

通过本系统的设计可得到什么样的结论，如某算法适合于某系统的设计等等。

e. 参考文献：

至少 5 篇以上，格式要规范。

## 四、课程设计的安排

- 1、时间：第五学期 17-18 周进行课程设计。
- 2、1 人一题。

## 五、考核方式与评分办法

- 1、学生在设计过程中的态度（10%）
- 2、完成工作量，提问、分析所涉及的问题的深度和广度（20%）
- 3、设计报告的格式规范性、层次合理性，语言通顺及图表的正确性（40%）
- 4、综合性面试考核：主要考核设计算法的有效性和解决实际识别问题的能力（30%）
- 5、综合上述四方面内容给出课程设计成绩（优、良、中、及格、不及格），不及格者重做。