**《机器学习（Machine Learning）》教学大纲**

**学分：**5

**学时：**90

**开课单位：**软件学院

**任课教师及职称（学位）：**汪浩教授、龙浩副教授、李第平博士

**教学内容及要求**

**1、课程简介：**世界的本质是数据（参见维克托·迈尔·舍恩伯格《大数据时代》），数据挖掘的任务是从大量数据中发现隐藏的、未知的和潜在的知识或模式的过程。机器学习(Machine Learning, ML)是数据挖掘的基础，为数据挖掘提供了基本工具，是大数据时代重要的课程。机器学习涉及概率论、统计学、最优化理论、算法复杂性理论等多门课程，本课程将讨论机器学习方法背后的基本数学原理，机器学习方法的基本数学模型，以及机器学习的基本方法和算法。

**2、教学目的与要求：**了解和掌握回归分析、分类、聚类、降维等机器学习的基本算法，同时掌握这些方法背后的数学原理，学习这些算法的基本原理，并能利用对应的软件包进行数据分析和挖掘。

**3、主要内容**

1. **引言（10课时）**

**教学目的和基本要求：**了解机器学习的基本内容，机器学习的分类，机器学习的三要素，机器学习的模型评估和选择。

**教学重点、难点：**机器学习三要素，机器学习模型评估与选择

**教学基本内容：**

* 1. 什么是机器学习
  2. 机器学习分类
* 有监督学习
* 无监督学习
* 半监督学习
  1. 机器学习三要素
* 模型
* 策略
* 算法
  1. 机器学习模型评估与选择
* 训练误差与测试误差
* 过拟合与模型选择

1. **回归分析（12课时）**

**教学目的和基本要求：**了解回归分析的作用，掌握回归分析的基本方法。

**教学重点、难点：**最小二乘法、最大似然法。

**教学基本内容：**

* 1. 线性建模：最小二乘法
  2. 回归分析用于预测
  3. 正则化最小二乘法
  4. 线性建模：最大似然法

1. **机器学习的贝叶斯方法（12课时）**

**教学目的和基本要求：**学习掌握共轭模型和非共轭模型等两种贝叶斯推理的基本方法。

**教学重点、难点：**图模型、共轭模型、最大后验估计MAP，拉普拉斯近似等。

**教学基本内容：**

3.1 图模型

3.2 贝叶斯估计的共轭模型

* 硬币游戏
* 后验概率
* 三个场景
* 边缘似然估计

3.3 贝叶斯计的非共轭模型

* 非共轭模型
* 二值响应
* 最大后验估计MAP
* 拉普拉斯近似
* 抽样技术

1. **分类（12课时）**

**教学目的和基本要求：**了解什么是分类，学习机器分类概率分类器和非概率分类器。

**教学重点、难点：**贝叶斯分类器，SVM及核方法

**教学基本内容：**

4.1 概率分类器

* **贝叶斯分类器**
* 逻辑回归

4.2 非概率分类器

* K紧邻法
* SVM及核方法

1. **聚类分析（12课时）**

**教学目的和基本要求：了解什么是聚类，及其与分类的区别，学习聚类的两种基本方法**

**教学重点、难点：**混合模型及EM算法。

**教学基本内容：**

5.1 K均值聚类

5.2 混合模型及EM算法

1. **主成分分析与隐变量（12课时）**

**教学目的和基本要求：**了解主成分分析的作用和目标，学习主成分分析、隐变量模型的基本内容。

**教学重点、难点：**主成分分析、隐变量模型、变分贝叶斯。

**教学基本内容：**

6.1 主成分分析

6.2 隐变量模型

6.3 变分贝叶斯

6.4 主成分分析的概率模型

**教材及参考书目**

1. SimonRogers, Mark Girolami. A First Course in Machine Leaening. CRC Press 2012. (中译本：郭茂祖，王春宇，刘杨，刘晓燕.机器学习基础教程机械工业出版社，2013)

1. 李航. 统计学习方法. 清华大学出版社，2012.