课程大纲：

第一章 真实世界中的机器学习

1 机器学习项目的流程

1.1 机器学习是什么，机器学习能干什么，不能干什么？

1.2 问题转化：真实世界中的问题怎样转化为机器学习问题

1.3 数据采集：哪些数据可能对手上的问题有帮助

1.4 建模：构建特征与目标间的函数关系

1.5 评估：模型指标与业务指标，评估与再优化

2 目标与评估指标

2.1 回归问题常用的评估指标

2.2 分类问题常用的评估指标

2.3 真实世界中的评估指标

3 机器学习的基本方法：损失函数最小化

3.1 损失函数与评估指标的区别

3.2 回归问题的损失函数

3.3 分类问题的损失函数

4 损失函数最小化算法：梯度下降

4.1 梯度下降算法

4.2 梯度下降算法的变种

5 过拟合与数据集划分

5.1 损失函数和评估指标的gap

5.2 过拟合现象：为什么要划分训练集、验证集和测试集

5.3 交叉验证方法

5.4 正则项与损失函数

第二周： 二分类

案例驱动： “魔镜杯”风控算法大赛

1. 如何EDA

2. 如何处理缺失值以及异常值

3. 分类特征如何处理

4. 时间特征如何处理

5. 数据不平衡如何处理

第三章 常见模型和模型调参

1 线性模型

1.1 常见线性模型

1.2 特征选择

1.3 正则项调参

2 集成学习模型

2.1 常见集成学习模型

2.2 集成学习模型的特征选择

2.3 集成学习模型的调参

3 深度学习模型

3.1 深度学习模型与传统模型的差异

3.2 深度学习模型的特征输入

3.3 深度学习模型的调参

第四周：

对上个案例总结：

1. 特征工程的常见方法

2. 训练集测试集划分方法

3. 数据不平衡方法总结

作业：马上金融违约用户风险预测案例建模

第五周 时间序列专题

1 时间序列预测概述

1.1 时间序列问题与传统问题的差异

1.2 时间序列问题的数据集划分与交叉验证方法

2 时间序列分析的基本框架

2.1 季节性规律

2.2 趋势规律

2.3 平稳序列与非平稳序列

3 从IJCAI客流量预测入门时间序列预测

3.1 预测任务

3.2 数据集划分

3.3 规则法解决时间序列问题

4 更复杂的时间序列模型

4.1 arima模型

4.2 rnn模型

第六周：

讨论作业：马上金融违约用户风险预测

比赛常用代码介绍，快速的写成baseline

作业：比赛 IEEE-CIS Fraud Detection

第七周 广告推荐算法

第八周：

比赛题目讨论

第十周：

比赛讨论，其他知识点讲解（如Prophet进行实践序列建模）