**第一天：**

统计学习（ statistical learning）是关于计算机基于数据构建概率统计模型并运用模型对数据进行预测与分析的一门学科。

统计学习的特点：

1. 计算机、网络为平台
2. 数据驱动的科学
3. 统计学习方法构建模型
4. 多学科交叉

统计学习的对象：以数据为核心，发现数据中的知识

统计学习目的：如何选择模型和学习模型

统计学习方法：方法多样，大体可以分监督学习、无监督学习和强化学习等。

本书主要讨论的是监督学习。

监督学习就是：从训练数据（已经标记好label的数据）中学习数据一般规律，把这种规律化为模型，最终我们学习到的是一个模型。一般的流程为给标注的数据（包含数据和标签）分为训练集和测试集，用训练集去训练一个模型，用测试集去测试这个模型训练的情况。

监督学习作为统计学习的一个重要的分支，在实际的业务场景中，很多模型也是属于监督学习方法训练的模型。

监督学习模型：线性模型，近邻模型，朴素贝叶斯模型、决策树模型，随机森林，支持向量机等模型。

—————————————————————————————————————

统计学习（ statistical learning）是关于计算机基于数据构建概率统计模型并运用模型对数据进行预测与分析的一门学科。

监督学习利用训练数据集学习一个模型，再用模型对测试样本集进行预测