1. 基本概念：训练集，测试集，特征值，监督学习，非监督学习，半监督学习，分类，回归

2. 概念学习：人类学习概念：鸟，车，计算机

   定义：概念学习是指从有关某个布尔函数的输入输出训练样例中推断出该布尔函数

3. 例子：学习"享受运动" 这一概念：

   小明进行水上运动，是否享受运动取决于很多因素

样例	天气	温度	湿度	风力	水温	预报	享受运动
1	晴	暖	普通	强	暖	一样	是
2	晴	暖	大	强	暖	一样	是
3	雨	冷	大	强	暖	变化	否
4	晴	暖	大	强	冷	变化	是

   天气：晴，阴，雨
   温度：暖，冷
   湿度：普通，大
   风力：强，弱
   水温：暖，冷
   预报：一样，变化

   享受运动：是，否

   概念定义在实例(instance)集合之上，这个集合表示为X。（X：所有可能的日子，每个日子的值由 天气，温度，湿度，风力，水温，预　　报6个属性表示。
   待学习的概念或目标函数成为目标概念（target concept), 记做c。
   c(x) = 1, 当享受运动时， c(x) = 0 当不享受运动时，c(x)也可叫做y
   x: 每一个实例
   X: 样例, 所有实例的集合
   学习目标：f: X -> Y

4. 训练集(training set/data)/训练样例（training examples): 用来进行训练，也就是产生模型或者算法的数据集
   测试集(testing set/data)/测试样例 (testing examples)：用来专门进行测试已经学习好的模型或者算法的数据集
   特征向量(features/feature vector)：属性的集合，通常用一个向量来表示，附属于一个实例
   标记(label): c(x), 实例类别的标记
   正例(positive example)
   反例(negative example)

5. 例子：研究美国硅谷房价
   影响房价的两个重要因素：面积(平方米），学区（评分1-10）

样例	面积（平方米）	学区 （1-10）	房价 （1000$)
1	100	8	1000
2	120	9	1300
3	60	6	800
4	80	9	1100
5	95	5	850

6. 分类 (classification): 目标标记为类别型数据(category)
   回归(regression): 目标标记为连续性数值 (continuous numeric value)

7. 例子：研究肿瘤良性，恶性于尺寸，颜色的关系
特征值：肿瘤尺寸，颜色
标记：良性/恶性

有监督学习(supervised learning)： 训练集有类别标记(class label)
无监督学习(unsupervised learning)： 无类别标记(class label)
半监督学习（semi-supervised learning)：有类别标记的训练集 + 无标记的训练集

8. 机器学习步骤框架

8.1 把数据拆分为训练集和测试集
8.2 用训练集和训练集的特征向量来训练算法
8.2 用学习来的算法运用在测试集上来评估算法 （可能要设计到调整参数（parameter tuning), 用验证集（validation set）

100 天： 训练集
10天：测试集 （不知道是否 " 享受运动", 知道6个属性，来预测每一天是否享受运动）