

副本 lecture01

Goal

上完课之后可以在学术界、工业界、英语和法律教育等构建机器学习应用

Prerequisites

理解基础的计算机技能和概念、比如时间复杂度、栈、队列、二叉树；
有概率论基础，知道随机变量、期望、方差；
有线代基础，知道矩阵，向量，以及他们相乘，特征向量。

homework

会通过若干作业来练习这些ML思想

form study group

建议组建学习小组，可以互相帮助，一起学习

Project

这门课的目标之一：做一个有意义的机器学习项目
浏览之前的课程网站，有许多项目：诊断癌症、地震工程、理解文学等。会让你得到启发

Discussion Section

在每周会有一些讨论环节，讨论内容有：线代、概率论、Python、NumPy、凸优化算法、隐含马尔可夫模型等。

Piazza

学生使用Piazza进行讨论问题，鼓励活跃参与回答问题。私人问题可以发邮件。

这次课程和之前课程的变化：

①从MATLAB改为python；②加入期中考试。
有60小时办公室时间可以问问题，作业提交截止日期在课表上。

CS229A, CS229, CS230的区别

课程	CS229A	CS229	CS230
课程内容	ML	ML	DL
More math	×	√	×
More applied	√	×	√

如果想成为ML领域的专家，建议多学一些像概率统计、凸优化、强化学习等。

学习ML的伟大意义

会让世界变得更美好，比如：改善医疗保健系统、让民主运行的更好等。

ML的一些定义

Arthur Samuel

So field of study gives computers the ability to learn without being explicitly programmed.

Tom Mitchell

a program is said to learn from experience E with respect to task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

Different tools in ML

1.Supervised Learning

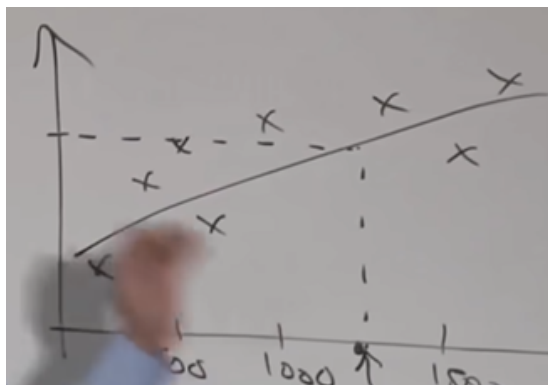
监督学习（最广泛应用）

定义：

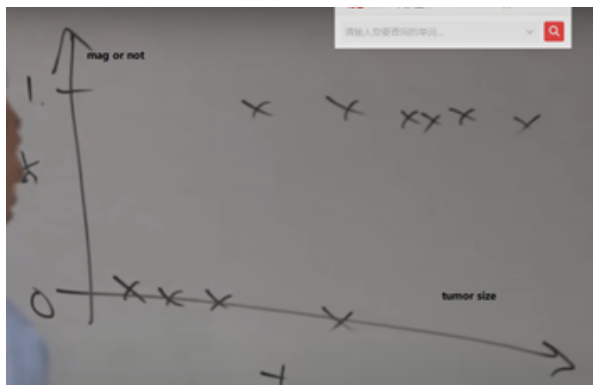
给出若干带有标记Y的X，找出X到Y的映射，然后根据这个映射判断新X。

例子

例1，房价（回归问题）



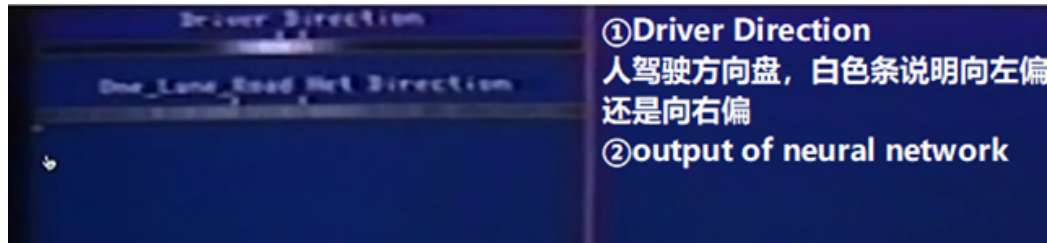
例2，乳腺肿瘤和乳腺癌（分类问题）



支持向量机

用来处理无穷个输入特征。

Video：汽车自动驾驶



输入X为人类操作方向盘，输出Y为轮子偏向，观察若干之后学会自动驾驶。

2.Machine Learning Strategy

3.Deep Learning

ML的一个分支

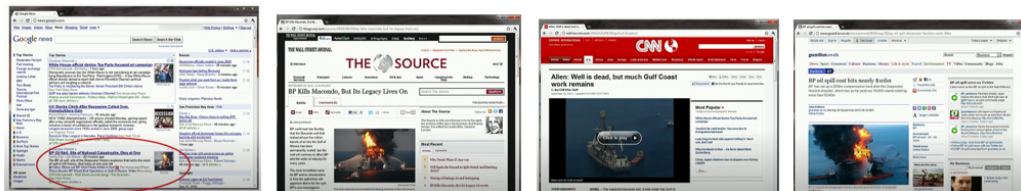
4.Unsupervised Learning

定义

没有标记的数据集，仅有输入x和输出y，从中找到有价值的数据

例子

例1，谷歌新闻对新闻进行聚类分类



例2，基因问题

例3，鸡尾酒会问题

一个房间里，有一个麦克风收集五个人一块讲话，要求分离出每个人说话的声音

例4，类比数据

用 男人->女人 类比： 国王->往后

用 东京->日本 类比： 华盛顿->美国

进而可以从大量无标记数据中，找到规律

5.Reinforcement Learning

例子

例1，直升机倒着飞

例2，训练狗狗，训练直升机

例3, 机器狗

例4, 玩游戏 (AlphaGo)