Machine learning

# 1 基本概念

## 1.1 what is machine learning?

Arthur Samuel:

Field of study that gives computers the ability to learn without being explicitly programmed.

不用刻意地用程序取指令，便能使机器能自我学习的领域。

Tom Mitchell:

A computer program is said to learn from experience E with respect to some task T and some performance measure P(性能度量，可以理解为做事的效果) ,if its performance on T,as measured by P,improves with experience E.

以自动识别垃圾邮件（spam）程序为例：

要完成自动区分垃圾邮件这个目标（这是任务T），关注你对垃圾邮件与非垃圾邮件的标记（这是经验E），提高此后区分垃圾邮件的成功率（这是性能度量P）。

## 1.2 监督学习与无监督学习（supervised learning & unsupervised learning）

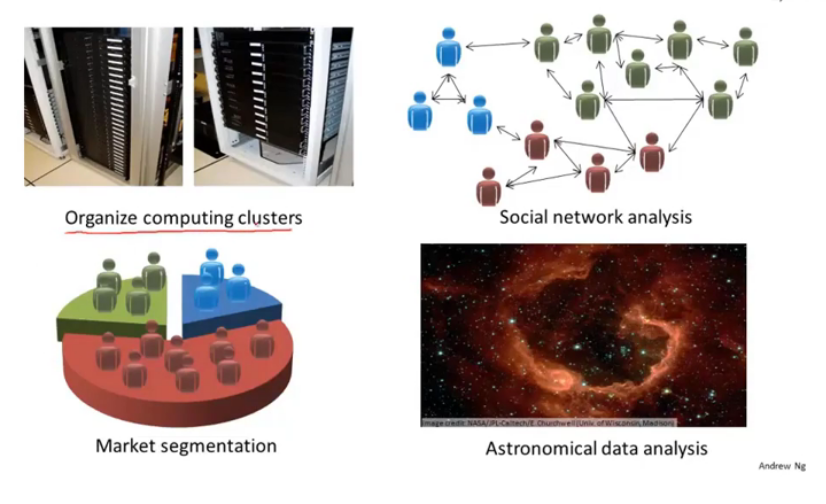
监督学习：人为规定了明确的判断结果，比如：

（1）离散型：给出历史数据预先标明什么情况的肿瘤是恶性的，什么样的是良性的，进而判断新的肿瘤样本是恶性的还是良性的。

（2）连续型：给出一系列历史房屋尺寸与价格的数据（或许可以进行线性拟合），从而预测新的一套房子可以卖出的价格。

无监督学习（clustering algorithm）：不会用所谓的答案来标定数据。

让程序自行分析数据，通过数据的结构，来得到结果。



## 1.3 回归任务与分类任务

回归任务与分类任务都属于监督学习。如果输出的结果是离散的，比如预测明天是晴天，雨天，阴天等，这样的问题就属于分类任务；如果输出的结果是连续的，比如预测明天的温度，可能是25.9 , 26 , 26.1 , 这样的问题就属于回归任务。

## 1.4 最小二乘法

最小二乘法（又称最小平方法）是一种数学[优化](https://baike.baidu.com/item/%E4%BC%98%E5%8C%96)技术。它通过最小化误差的平方和寻找数据的最佳[函数](https://baike.baidu.com/item/%E5%87%BD%E6%95%B0/301912)匹配。利用最小二乘法可以简便地求得未知的数据，并使得这些求得的数据与实际数据之间误差的平方和为最小。最小二乘法还可用于**[曲线拟合](https://baike.baidu.com/item/%E6%9B%B2%E7%BA%BF%E6%8B%9F%E5%90%88/5893992)**。其他一些优化问题也可通过最小化能量或最大化熵用[最小二乘法](https://baike.baidu.com/item/%E6%9C%80%E5%B0%8F%E4%BA%8C%E4%B9%98%E6%B3%95/2522346)来表达。

## 1.5 课程中的常用标记

m—训练样本的数量

x—输入（特征feature）

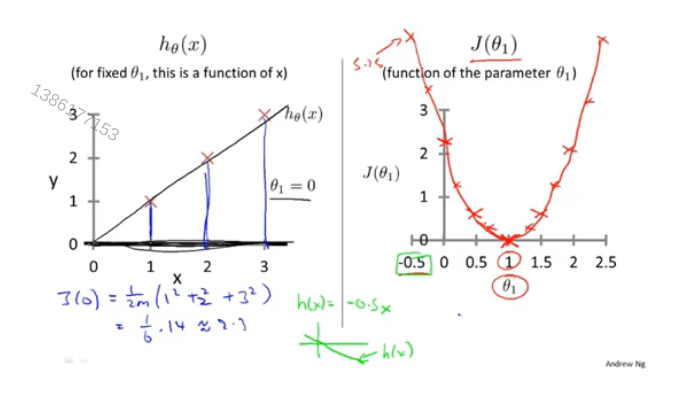
y—输出（目标target）

（x,y）—一个训练样本

（x(i),y(i)）—第i个训练样本

h（hypothesis）—假设函数：从输入得到输出

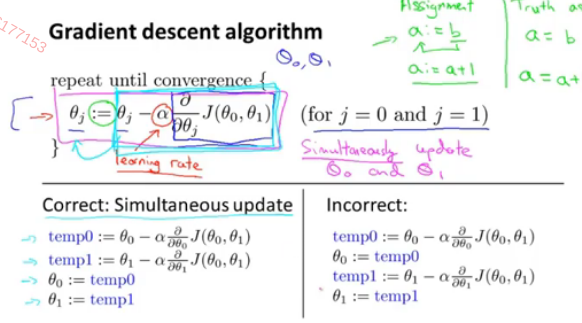
## 1.6 假设函数与代价函数（cost function）



假设函数：是关于输入x的函数，显示的是输入与输出的关系。

代价函数：是关于系数θ的函数， J（θ0，θ1，...）的函数图像表示的是不同的系数与其所造成的误差的关系。这条曲线用来评价系数的取值是否合理。

## 1.7 梯度下降（gradient descent）



符号注释：

：= 赋值，a:=b表示将b的值赋给a

= 判断，a=b表示判断a，b的值是否相等

α：学习效率，α越大，梯度下降的也越块，但太快可能会造成错误的结果

α越小，梯度下降的越慢，太慢可能导致收敛太慢

深蓝框的偏倒：代表的是图像上那一点的斜率。注意，当点越靠近局部最优解（倒数为0）时，斜率会越小，也就是说θj的更新程度会自动越来越小。

注意：

1.这里的j可以取0和1，代表θj和θo。

2.θj和θo的更新应该同时进行，不然一个变量的更新会影响另一个变量的变化。