

一、机器学习

也就是说机器学习不需要制定具体的模型，而是让计算机根据庞大的数据量自己训练模型，与之相对的，例如 CFD 软件，是建立在物理模型之上的，例如输运方程等。

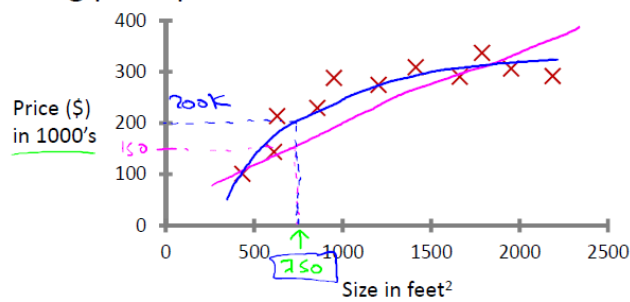
二、监督学习 (Supervised learning)

数据集中的每个样本有相应的“正确答案”，根据这些样本做出预测，分有两类：回归问题和分类问题

回归问题

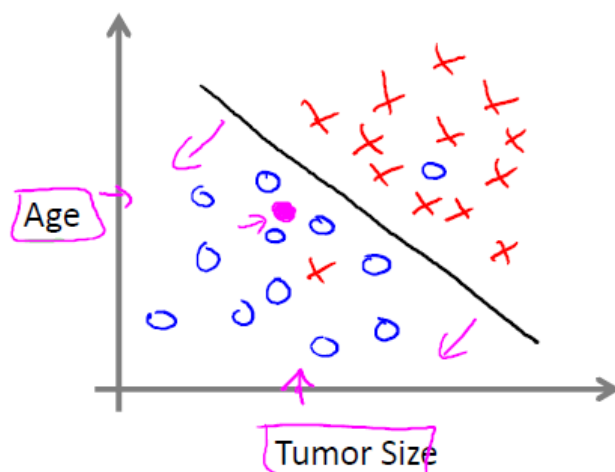
给定一个新的模式，根据训练集推断它所对应的输出值（实数）是多少，是一种定量输出，也叫连续变量预测。例如预测房价，根据样本集拟合出一条连续曲线（预测的结果为连续的值比如数字、温度）

Housing price prediction.



(2) 分类问题

给定一个新的模式，根据训练集推断它所对应的类别（如： $+1$, -1 ），是一种定性输出，也叫离散变量预测。例如：根据肿瘤特征判断良性还是恶性，得到的是结果是“良性”或者“恶性”，是离散的



二、非监督学习 (Unsupervised learning)

非监督学习的数据集跟监督学习不同，没有任何标签，即没有相应的“正确答案”。从数据集中可以通过非监督学习得到数据的某种结构，可能是把数据分成两个不同的聚集簇，称为聚类算法。

回归与分类的区别

回归问题和分类问题的本质一样，都是针对一个输入做出一个输出预测，其区别在于输出变量的类型。分类问题

1、分类问题：给定一个新的模式，根据训练集推断它所对应的类别（如： $+1$ ， -1 ），是一种定性输出，也叫离散变量预测。

2、回归问题是指，给定一个新的模式，根据训练集推断它所对应的输出值（实数）是多少，是一种定量输出，也叫连续变量预测。

如果预测的结果为连续的值，是回归问题；如为离散的值，是分类问题
题目中的 **number** 也是连续的值,所以也是回归，温度是连续的值,所以是回归

区分监督与非监督学习,关键是看有没有标准

选项 1: 给定 1000 篇文章，自动把这些文章分为相似的还是相关的.没有给出分类标准属于非监督学习

选项 2: 给定儿童的身高与年年龄的历史数据,去预测一个儿童在某岁时的身高,是监督学习

选项 3: 给定 50 个男作者与 50 个女作者的文章,再给一个文章去预测是男作者还是女作者,是监督学习

选项 4: 给定一堆垃圾邮件,然后去发现垃圾邮件的子类型,没有分类标准所以是非监督学习

以上 2 的标准就是这个儿童的身高, 3 的标准就是作者的性别,所以是监督学习

1 与 4 找不到标准,所以是非监督学习