# 引言

## 1.1 机器学习的作用

第一个视频主要讲机器学习是什么，以及机器学习能干什么

机器学习的作用:

1. 数据挖掘 机器学习被用于数据挖掘的原因之一是网络和自动化技术的增长，，我们有史上最大的数据集，机器学习算法从中分析数据，更好的服务用户。
2. 医疗服务 如分析电子医疗记录
3. 计算生物学 分析人类的基因特性，DNA序列等
4. 计算机视觉 如无人驾驶，图像识别
5. 自然语言处理 机器翻译等
6. 推荐系统

## 1.2 机器学习的定义

机器学习的定义: 一个程序被认为能从经验E中学习，解决任务T，达到性能度量值P，当且仅当，有了经验E后，经过P评判，程序在处理T时的性能有所提升

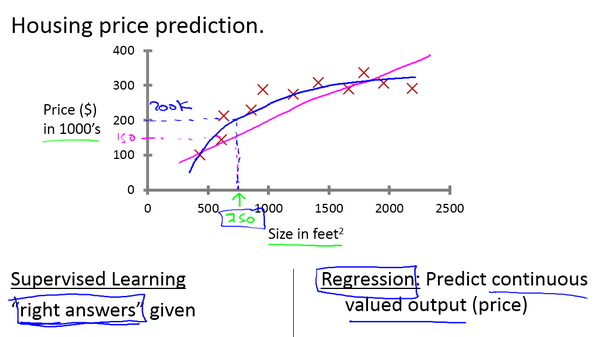
通俗的说，就是一个为了解决某任务的算法只有从某经验中学习才能获得性能的提升，且通过学习必定能获得性能的提升。如垃圾分类系统中，任务T就是识别垃圾邮件，P是是别的准确率，经验E就是相关数据集

## 1.3 监督学习

监督学习何和无监督学习是机器学习中的常见算法。简单的说，监督学习是我们教计算机去完成任务，无监督学习是计算机通过自我学习完成任务。即监督学习中使用的数据集是带有标签的，这个标签就是”标准答案“。而无监督学习使用的数据集是没有标签的。

接下来从一个例子着手解释监督学习

假如说你想预测房价，你花时间收集了大量房价数据，并将它们画出来，如图所示。

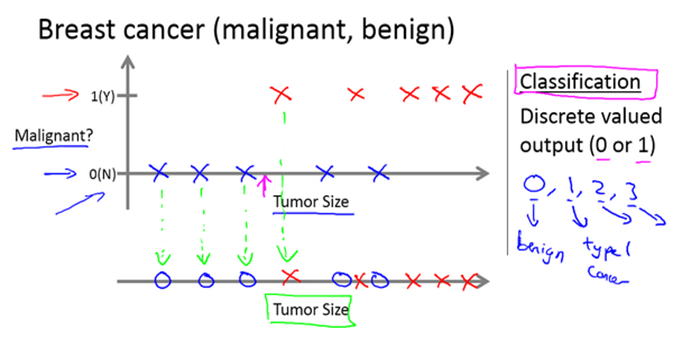


其中x轴表示房子的面子，y轴表示房子的价格。那基于这组数据，假如你有一个朋友，他有一套750平方英尺房子，现在他希望把房子卖掉，他想知道这房子能卖多少钱。

我们应用学习算法，可以从图中拟合出一条直线，以此预测出房子的价格。这就是一个典型的监督学习的问题，针对每一特定的样本输入特征x，都有唯一的y标签与之对应，而我们的学习算法就要找出一条直线，使之预测的y值与样本y值最接近。我们给了一系列房子的数据，我们给定数据集中每个样本的正确价格，即它们实际的售价然后运用学习算法，算出更多的正确答案。

这也就叫做回归问题，我们试着推测出一系列连续的值，即房子的价格。

接下来还有一个监督学习的例子，是分类的问题。如图所示，



我们有一些关于肿瘤的数据集，其中特征是肿瘤的大小，而标签即是肿瘤的恶性与否。1表示恶性，0表示非恶性。我们的任务就在于，给定新样本的大小，能否准确给出是恶性肿瘤的概率。

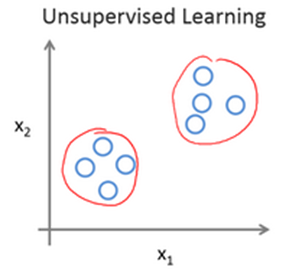
这是一个分类问题，分类指的是，我们的输出是离散值，即只有0和1。

监督学习的基本思想是，对于每个特征都有标准答案，算法通过学习这些“标准答案“，对新样本的答案做出预测。

## 1.4 无监督学习

在监督学习中，每个训练样本都有已知的答案，而在无监督学习中则不同，输入的样本只有特征，而没有确定的标签。针对这个数据集，我们的无监督学习算法能从中找出该数据集具有的特殊结构。

如下图所示，以下是没有标签的数据集，我们的无监督学习算法能自动将该数据集分为两类



该算法被称为聚类算法

聚类算法有如下应用：

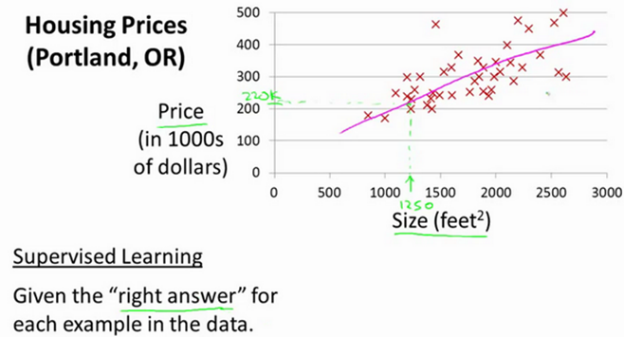
1. 谷歌新闻中的算法能自动为新闻归类
2. 能够根据DNA信息进行身份匹配
3. 组织计算机集群
4. 根据用户信息进行市场切分
5. “鸡尾酒算法”

# 单变量线性回归

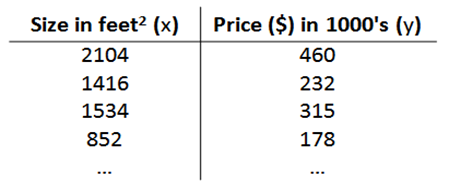
## 2.1 模型表示

我们第一个学习的算法是线性回归算法。

我们继续从房价预测的例子来说。为了预测房价，我们已获得了许多关于房价的数据

如图所示 从图中数据所看，我们似乎可以拟合出一条直线，并以此预测房价。

我们的数据集如下表所示



我们将要用来描述这个回归问题的标记如下:

m 代表训练集中实例的数量

x 代表特征/输入变量

y 代表目标变量/输出变量

(x,y) 代表训练集中的实例

)代表第i 个观察实例

h代表我们的假设函数

h根据我们样本的输入特征x，得出预测值y’

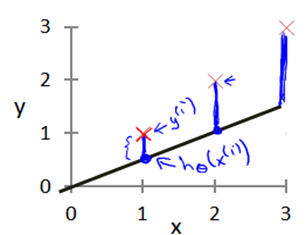
那么问题来了，我们该如何表示我们的假设函数h呢？

一种可能的表示是，，因为只有一个变量，所以这样的问题叫做单变量回归问题

## 2.2 代价函数

我们将定义代价函数的概念，这有助与拟合假设函数

若我们随机为模型选择参数，这样得到的结果往往是不靠谱的。为了能得到最为精准的预测，我们不得不精心选择参数。在选择参数前，我们首先要评估模型的准确度。

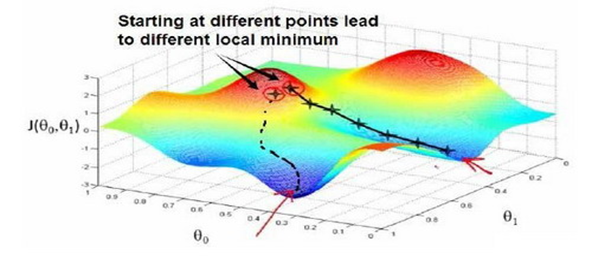
我们可以观察到，我们的假设函数值与实际标签间是存在差距的，为了衡量这种差距的大小，我们引入，即每个预测值与实际值差值的平方均值。我们的目标是选择合理的参数，能够使得最小。

由数学知识可得，要求J的最小值，可直接令J对各个参数求偏导，并让偏导值为0，如此即可得出J的最小值，但是对于成百个参数时，这样解偏导会十分复杂。

## 2.3 梯度下降

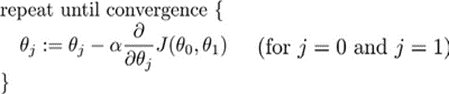
梯度下降是一种能够用来求函数最小值的方法，背后的数学原理我会在另一篇文章中讨论。

简单理解梯度下降的原理就是，开始时我们随机选择一个参数的组合(θ\_0,θ\_1,......,θ\_n )，计算代价函数，然后我们寻找下一个能让代价函数值下降最多的参数组合。我们持续这么做知道找到能令代价函数最小的参数。但因为我们并没有尝试完所有的参数组合，所以不能确定我们得到的局部最小值是否便是全局最小值。

如图所示 

每次都选择下降最快的点，即梯度进行更新。

梯度下降的算法公式为

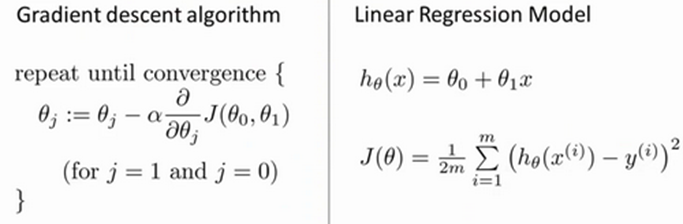


其中α是学习率， 决定下降的不乏，α太大会使函数不收敛，太小会占用太多训练时间。

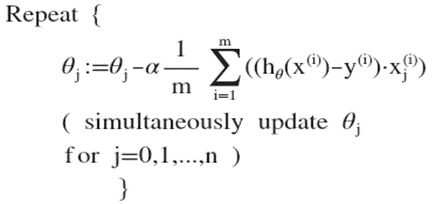
进行梯度下降时，我们需要同时更新各个参数。

## 2.4 通过梯度下降实现线性回归

梯度下降和线性回归算法如图所示



要更新参数θ，我们需要对J求偏导，解得



该算法也称为梯度批量下降，指在梯度下降的每一步中，我们都用到了所有的样本。