华科小涛

追求卓越，成功就会不经意间的找上你！

联系 管理

公告

昵称：华科小涛
园龄：3年2个月
粉丝：24
关注：15
+加关注

< 2017年2月 >

日	一	二	三	四	五	六
29	30	31	1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	1	2	3	4
5	6	7	8	9	10	11

我的标签

C++学习之路基础篇(23)
互联网知识积累(10)
数据结构与算法(10)
机器学习与数据挖掘(9)
Java 学习(8)
Linux基础(8)
Python(6)
链接、装载与库(5)
网络协议(5)
设计模式(2)
更多

积分与排名

积分 - 20787
排名 - 12100

随笔-88 文章-0 评论-49

Linear Regression(线性回归)（一）—LMS algorithm

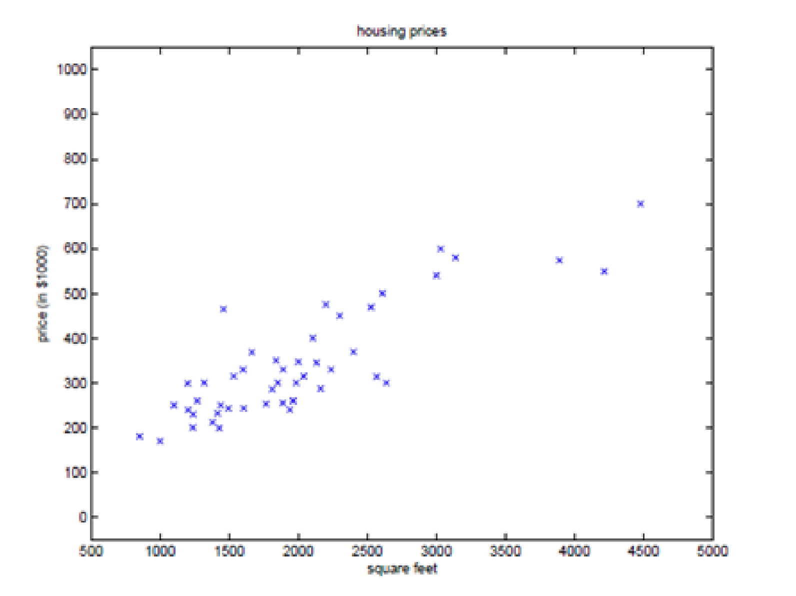
（整理自AndrewNG的课件，转载请注明。整理者：华科小涛@http://www.cnblogs.com/hust-ghtao/）

1.问题的引出

先从一个简单的例子说起吧，房地产公司有一些关于Portland,Oregon的房子信息，下表是房子的面积和价格的对照表：

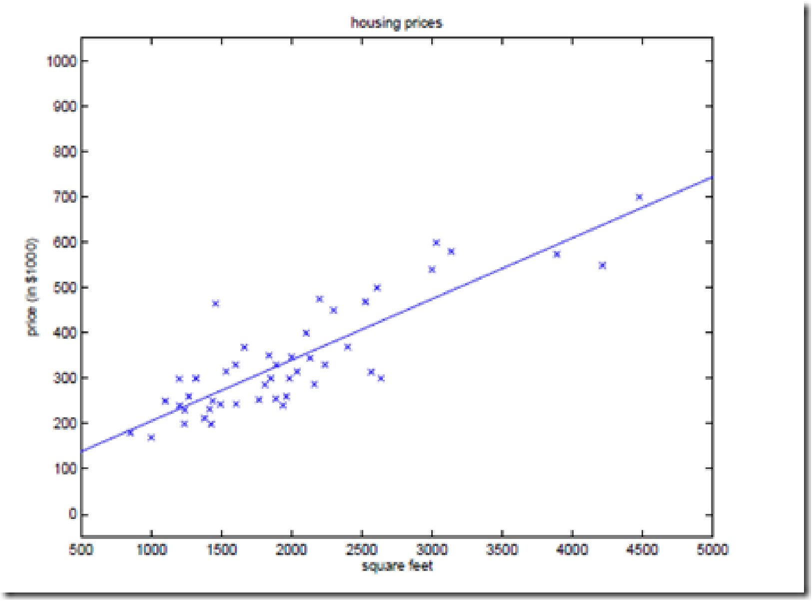
Living area(feet^2)	Price(1000＄s)
2104	400
1600	330
2400	369
1416	232
3000	540
……	……

将点画在二维坐标下表示：



那么问题就来了，面积为2000的房子，价格是多少呢？表中没有直接给出答案，考虑能否根据表中给定的数据，得出价格和面积之间的一个函数，对于重新给定的面积来对其价格进行预测，这就是我们

要解决的问题。有人说这不是数据拟合吗？没错这就是数据拟合！若为线性拟合，则表达式为：$y=\theta_0+\theta_1 x$，拟合结果如些图：



如果给matlab这组数据，再用一个线性拟合的指令，很快便能得到θ_0和θ_1得值。问题是在计算机内部是如何实现这种线性的数据拟合的，没错这就是线性回归的方法。

为了便于更一般描述线性回归的问题，也为了更加严谨，先对本课程中出现的一些基本符号与概念作出定义：

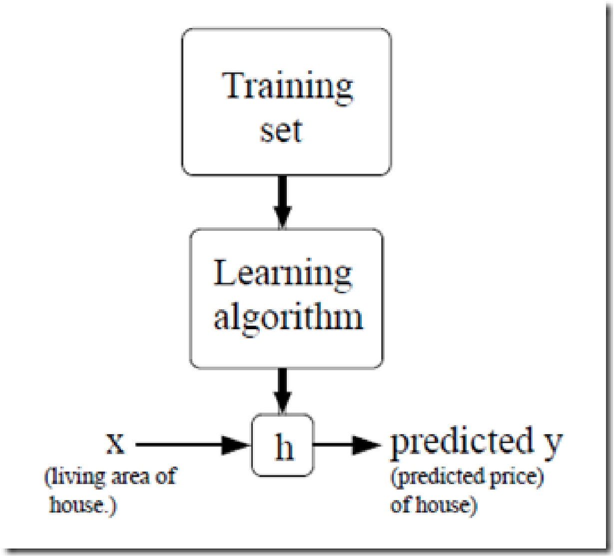
$x^{(i)}$：输入变量（例如本例中的Living area），也称为输入特征；

$y^{(i)}$：输出变量（例如本例中的Price），也成为目标变量；

一对数据$(x^{(i)},y^{(i)})$，称为一个训练样本；我们用来学习的训练样本的集合$\{(x^{(i)},y^{(i)});i=1,...,m\}$称为训练集。这里的上标$(i)$指的是样本在训练集中的序号。

X称为输入变量空间，Y称为输出空间变量。

把我们的问题描述的更加正式：给定训练集，去学习一个函数h：$X\rightarrow Y$，这样就可以根据给定的x来预测y了。我们也把这个函数称为假设。过程用图表示如下：



当我们要预测的目标变量是连续时，例如本例中的房屋价格，我们把这种学习问题称为回归问题（regression problem）；当目标变量只能取一些离散的值时，我们称这种问题为分类问题（classification problem）。

更一般地，为了使我们的问题更加一般化，假设输入特征可以多于一个，像在本例中除了Living area，还有#bedrooms、whether each house has a fireplace、#bathrooms，and so on。如果对于一个线性回归问题，有n个输入输入变量，则假设$h_\theta(x)$可以写成：$h_\theta(x)=\theta_0+\theta_1x_1+\dots+\theta_nx_n$，$\theta_i$就是我们需要确定的参数。为了使表述更加简洁，我们令$x_0=1$，所以假设可以写成：$$h(x)=\sum_{i=0}^{n}\theta_ix_i=\theta^Tx$$，等号右侧θ和都是向量，是输入向量的个数（不含x_0）。那么给定一个训练集，我们应该如何选择或者学习？有种想法很显然，就是使$h(x)$更加接近，至少对于训练样本来说是这样的。为了使这个条件公式化，我们需要提出一个衡量标准的函数，我们定义了一个成本函数（cost function）：

$$J(\theta)=\frac{1}{2}\sum_{i=1}^{m}\left(h_\theta(x^{(i)})-y^{(i)}\right)^2$$

。

至此，线性回归问题就是学习，使$J(\theta)$最小。

2 .LMS算法

既然已经定义了线性回归问题，那么用什么算法求解呢？在数值分析中，遇到过类似的问题，先给一组初始估计值，通过不断地改变，使不断变小，直到收敛。这里我们应用梯度下降法，它的更新规则是：$$\theta_j:=\theta_j-\alpha\frac{\partial}{\partial\theta_j}J(\theta)$$（对$j=0\dots n$同时更新）。其中α称为学习率，通过调整此参数，可以改变迭代速度。在等式右侧，需要计算的偏导数：

$$\frac{\partial}{\partial\theta_j}J(\theta)=\frac{\partial}{\partial\theta_j}\frac{1}{2}(h_\theta(x)-y)^2$$

$$=2\cdot\frac{1}{2}(h_\theta(x)-y)\cdot\frac{\partial}{\partial\theta_j}(h_\theta(x)-y)$$

$$=(h_\theta(x)-y)\cdot\frac{\partial}{\partial\theta_j}(\sum_{i=0}^{n}\theta_i x_i-y)$$

$$=(h_\theta(x)-y)x_j$$

所以更新规则变为：$\theta_j := \theta_j+\alpha(y^{(i)}-h_\theta(x^{(i)}))x_j^{(i)}$。

这个规则就是LMS（最小均方规则）。可以看出每次更新的值是和$(y^{(i)}-h_\theta(x^{(i)}))$成比例的，当的值较大时，每次改变的值就较大，反之较小。当已经很小时，说明已经达到拟合要求，的值就不变了。当只有一个训练样本时，我们得到LMS算法，将LMS算法应用到线性回归有两种方式：批处理梯度下降法和随机梯度下降法。

2.1 批处理梯度下降法

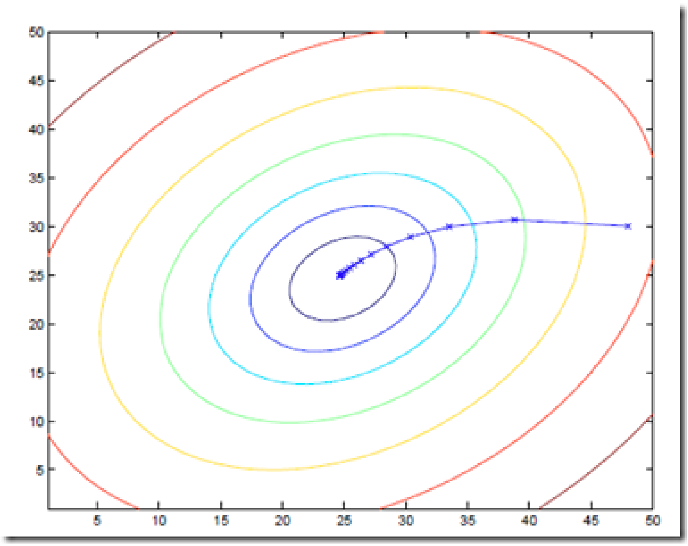
算法：

$Repeat\ until\ convergence$ {

$$\theta_j := \theta_j+\alpha\sum_{i=1}^{m}(y^{(i)}-h_\theta(x^{(i)}))x_j^{(i)}(for\ every\ j.)$$

}

可以看出，的值每更新一次都要遍历样本集中所有的样本，得到新的θ_j，看是否满足阈值要求，若满足，则迭代结束，根据此值就可得到；否则继续迭代。注意到，虽然梯度下降法易受目标函数的局部极小值的影响，但是一般的线性规划问题只有一个极小值，所以梯度下降法一般可以收敛到全局的最小值。例如，J是二次凸函数，则梯度下降法的示意图：



图中，一圈上表示代价函数的函数值相同类似于地理上的等高线哈，从外圈开始逐渐迭代，最终收敛全局最小值。

2.2 随机梯度下降法

算法：

$Loop$ {

$$for\ i=1\ to\ m\ \{$$

$$(for\ every\ j)$$

}

}

在这个算法中，我们每次更新只用到一个训练样本。若根据当前样本进行迭代得到一个，此时会得到一个，有新样本进来之后，在此基础上继续迭代，有得到一组新的和，以此类推。

总结一下这两个梯度算法：批处理梯度下降法，每更新一次，需要用到样本集中的所有样本；随机梯度下降法，每更新一次，只用到训练集中的一个训练样本，所以一般来说，随机梯度下降法能更快的使目标函数达到最小值（新样本的加入，随机梯度下降法有可能会使目标函数突然变大，迭代过程中在变小。所以实在全局最小值附近徘徊，但对于实际应用来说，误差完全能满足要求。）。另外，对于批处理梯度下降法，如果样本集中增加了些许训练样本，就要重新开始迭代。由于以上原因，当训练样本集较大时，一般应用随机梯度下降法。

标签: 机器学习与数据挖掘

« 上一篇：串的模式匹配算法（一）—朴素的模式匹配算法
» 下一篇：Linear Regression(线性回归)（二）—正规方程（normal equations）

posted @ 2014-03-02 01:16 华科小涛 阅读(892) 评论(1) 编辑 收藏

评论列表

#1楼[楼主] 2014-09-17 21:54 华科小涛

这篇阅读量最多了！

支持(0) 反对(0)