

知

首发于  
机器学习笔记

写文章

ooo

## 线性回归

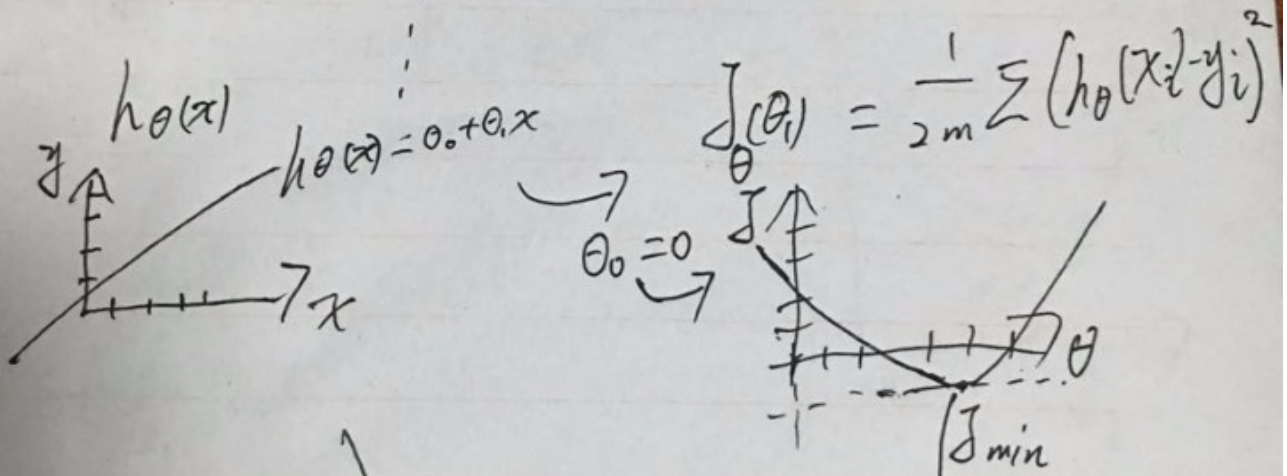
Training Set  
(m)↓  
学习算法

↓

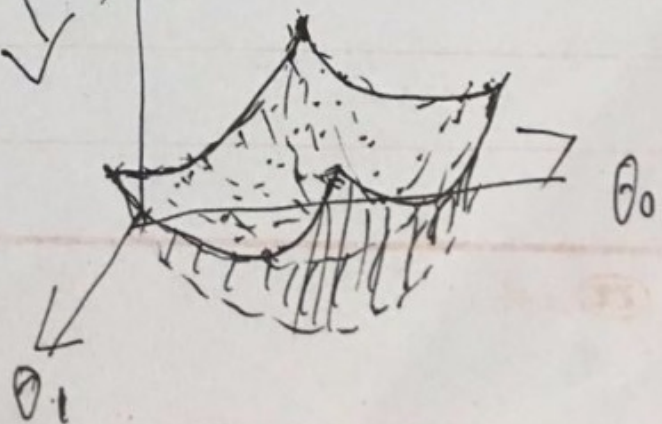
$x$  →  $h$  →  $y$   
 特征值      假设函数      目标值  
 $x_i$                    $h_{\theta}(x) = \theta_0 + \theta_1 x$        $y_i$

$$\text{minimize}_{\theta_0, \theta_1} \frac{1}{2m} \sum_{i=1}^m (h(x_i) - y_i)^2$$

$\min(\theta_0, \theta_1)$   $J(\theta_0, \theta_1)$   
 代价函数 (Cost function)  
 (平方误差函数, 常用于回归问题)



$\theta_0 \neq 0$   
 $J_{\theta}$   
 $\theta_0 = 0$   
 $J_{\min}$   
 $\theta_0 = ? \theta_1 = ?$   
 最佳拟合



知

首发于  
机器学习笔记

写文章

...



子楠 · 2 年前

我们知道，在做数学题的时候，解未知数的方法，是给定自变量和函数，通过函数处理自变量，以获得解。而机器学习就相当于，给定自变量和函数的解，求函数。

类似于：这样：function (x) =y

机器学习就是样本中有大量的x（特征量）和y（目标变量）然后求这个function。

求函数的方法，基于理论上来说，大部分函数都能找到一个近似的泰勒展开式。而机器学习，就是用数据去拟合这个所谓的“近似的泰勒展开式”。

大致可以把机器学习分为Supervised learning（监督学习）和Unsupervised learning（非监督学习）两类。两者区别在于训练样本。

简单理解的话：

监督学习就是你做卷子，做完以后给你正确答案你可以对答案，从而得出获得正确答案的方法。

非监督学习就是只给你一堆题，你做的题多了，自然也能判断出获得正确答案的规律。

监督学习多用于回归分析（求解是连续值，比如某一区间）和分类问题（求解是离散值，比如对错）。非监督学习初步多用于聚类算法（群分析）

线性回归：

假设 $h(x) = y$ ，有大量样本，求 $h(x)$ 是什么。

那么可以假设  $h(x) = \theta_0 * x^0 + \theta_1 * x^1 + \theta_2 * x^2 + \dots + \theta_n * x^n + \dots$  因为  $x^0 = 1$  所以  $\theta_0 * x^0$  可以简化为  $\theta_0$

为了方便，这里先记只有  $\theta_0$  和  $\theta_1$  的情况，原公式就简化为了：  $h(x) = \theta_0 + \theta_1 * x$

假设实际结果为y， $y = f(x)$

那么， $h(x)$  在什么情况下最接近 $f(x)$  呢？

$h(x) - f(x)$  最接近0的情况下啦。

那大量样本（m组样本）呢？

知

首发于  
机器学习笔记

写文章

...

我们给  $h(x) - f(x)$  做一个平方，保证它为正，为正了以后，就是求最小了对吧。方便很多：

$$(h(x) - f(x))^2$$

然后把这些样本相加： $\sum_i^m (h(x_i) - f(x_i))^2$  这个相加后的值最小的时候，就是我们的假定函数  $h(x)$  最接近实际函数  $f(x)$  的时候。

由于方程中没有  $f(x)$ ，而  $f(x) = y$ ，所以方程可以写为： $\sum_i^m (h(x_i) - y_i)^2$

假设我们有  $m$  个样本，那么我们的目标就是：计算这些样本下，找到至少一个  $\theta_0$  和  $\theta_1$  的值，使得在这个值下，样本所有结果得到的偏差最小。现在我们有两个方程：

$$\sum_i^m (h(x_i) - y_i)^2$$

$$h(x) = \theta_0 + \theta_1 * x$$

我们的目标就是，找到至少一个  $\theta_0$  和  $\theta_1$  的值，使得  $\sum_i^m (h(x_i) - y_i)^2$  最小，这个值下得到的函数  $h(x) = \theta_0 + \theta_1 * x$  就是最接近真实函数的假定函数。

实际操作很简单.....

$$h(x) = \theta_0 + \theta_1 * x, \text{ 所以 } \min(\theta_0, \theta_1) = \sum_i^m (h(x_i) - y_i)^2 = \sum_i^m (\theta_0 + \theta_1 * x_i - y_i)^2$$

假设这个  $\min(\theta_0, \theta_1)$  的平方差函数为  $J_\theta(\theta_0, \theta_1)$  那么，

$$J_\theta(\theta_0, \theta_1) = 1/2m * \sum_i^m (\theta_0 + \theta_1 * x_i - y_i)^2$$

那么当这个函数  $J_\theta(\theta_0, \theta_1)$  最小的时候，就是我们得到  $\theta_0, \theta_1$  的时候。

具体做法叫做“梯度下降”

对方程  $J_\theta = 1/2m * \sum_i^m (\theta_0 + \theta_1 * x_i - y_i)^2$  对  $\theta_j$  进行求导，得到：

$$\frac{d}{d\theta_j} J_\theta = \frac{d}{d\theta_j} 1/2m * \sum_i^m (\theta_0 + \theta_1 * x_i - y_i)^2$$

当  $\theta_i = \theta_0$  的时候， $\frac{d}{d\theta} J_\theta = 1/m * \sum_i^m (\theta_0 + \theta_1 * x_i - y_i)$

当  $\theta_j = \theta_1$  的时候,  $\frac{d}{d\theta_1} J_\theta = 1/m * \sum_i^m (\theta_0 + \theta_1 * x_i - y_i) * x_i$

(注意, 求导的意思是, 得到一个微小变量, 所以上面的  $\frac{d}{d\theta_0} J_\theta$  和  $\frac{d}{d\theta_1} J_\theta$  可以看作分别是对于  $\theta_0$  和  $\theta_1$  的变化的一个变化量 (方向) 的预测。你也可以写为、 $\theta_0$  和、 $\theta_1$ , 看个人习惯。)

那么, 因为我们的目标是要得  $\theta_0$  和  $\theta_1$  的值, 这个值使得  $J_\theta(\theta_0, \theta_1)$  最小。那么我们就可以随便选个起始点  $J_\theta(\theta_0, \theta_1)$ , 假设选取的点是  $J_\theta(0, 0)$ , 你可以选  $J_\theta(3, 7)$ ,  $J_\theta(6, 9)$ , 无所谓。反正最终会使得  $\theta_0$  和  $\theta_1$  不断改变的.....

具体做法是, 分别对于  $\theta_0$  和  $\theta_1$ ,  $\pm\lambda$ , 使得  $J_\theta(\theta_0 \pm \lambda, \theta_1 \pm \lambda) < J_\theta(\theta_0, \theta_1)$  不断重复, 直到再也找不到一个  $\lambda$  (或者找到的  $\lambda$  太小可以忽略不计) 使得  $J_\theta(\theta_0 \pm \lambda, \theta_1 \pm \lambda) < J_\theta(\theta_0, \theta_1)$ , 那么此时的  $\theta_0$  和  $\theta_1$  的值就是我们想要的。使得  $J_\theta(\theta_0, \theta_1)$  最小的值。这时候得到的  $h(x) = \theta_0 + \theta_1 * x$ , 就是我们想要的最接近目标函数的假定函数。

具体做法如下:

分别对于  $\theta_0$  和  $\theta_1$ , 求出其  $\frac{d}{d\theta_0} J_\theta$  和  $\frac{d}{d\theta_1} J_\theta$  的值, 分别乘以一个参数  $\alpha$  控制下降速度。让计算机重复如下运算:

$$\begin{aligned}\theta_0 &= \theta_0 - \alpha * \frac{d}{d\theta_0} J_\theta \leadsto \theta_0 = \theta_0 - \alpha/m * \sum_i^m (\theta_0 + \theta_1 * x_i - y_i) \\ \theta_1 &= \theta_1 - \alpha * \frac{d}{d\theta_1} J_\theta \leadsto \theta_1 = \theta_1 - \alpha/m * \sum_i^m (\theta_0 + \theta_1 * x_i - y_i) * x_i\end{aligned}$$

控制重复的方法很多。比如可以每次计算  $J_\theta(\theta_0, \theta_1)$  和上一次的  $J_\theta(\theta_0, \theta_1)$  的差值 ( $J_\theta(\theta_0 \pm \lambda, \theta_1 \pm \lambda) - J_\theta(\theta_0, \theta_1)$ ), 如果两次得到的差值  $> 0$ , 也就是变化方向反了, 那就控制函数改方向 (改变  $\alpha$  的值, 加个+号或者-号)。如果每次计算都得到的差值都很奇怪, 比如一下大一下小, 一下正一下负, 可能是控制幅度太大, 缩小  $\alpha$  的值即可。不断重复直到差值小于一个你觉得合适的极小值。方法很多, 不一一枚举。

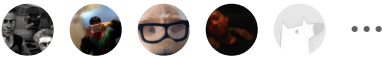
注意: 梯度下降, 每一次都需要用到所有样本 (或者一堆样本, 其它用来做测试, 但是对于你用到的样本来说, 每一次训练依然是用到所有样本。) 来训练数据。这叫做“批量梯度下降”。编程的时候不需要用到循环语句, 用矩阵相乘, 或者向量相乘即可。方便很多。

第一周笔记到此结束。安德鲁ng的课程在第一周最后几节课及作业是给大家复习线性代数的。这里就不继续做笔记了。当然我的纸质笔记本上是记了这部分的.....如果有很多人留言说有需要那我就贴上来就是了.....

机器学习    线性回归

☆ 收藏    ↗ 分享    ⚠ 举报

 66



32 条评论



写下你的评论...



馨园

楼主，我想问下  
这个公式中m是必须的么，如果去掉 $\frac{1}{2}m$ ，整个公式就很好解释其意思，加上就不太明白这个函数想表达的意思。是不是去掉最好，去掉也不影响其求导结果。

1 年前

5 赞

以上为精选评论



李嘉

mark  
2 年前



剧小白

好帖，请问楼楼后来认证了吗  
1 年前



李龙

感谢楼主的笔记，有个小疑惑  
1 年前

知

首发于  
机器学习笔记

写文章

...



$x_0 == 0$ , 你确定?

1 年前



大龙瑟

mark

1 年前



子楠 (作者) 回复 剧小白

查看对话

认证了, 拿到证书了。纯当支持吧。我coursera都花了5000多块钱了, 感觉证书没啥用。

1 年前



子楠 (作者) 回复 李龙

查看对话

有疑问直接提

1 年前



方长

手工点个赞, 看来得先好好复习一下高数

1 年前



王学琴

子楠的笔记好详细, 多谢分享。学习ing...求线性代数笔记。

1 年前

下一页

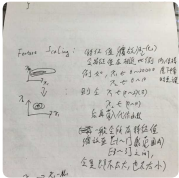
文章被以下专栏收录



机器学习笔记

就是笔记而已。

进入专栏



第二周笔记：多元线性回归

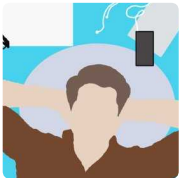
第一周的笔记里，我们假设函数是这样的：也就是，自变量只有一个， $x$ 。无论是 $x$ 的1次方还是 $x$ 的... 查看全文 >

子楠 · 2 年前 · 发表于 机器学习笔记

专栏简介：

该专栏只是个人学机器学习（学习的是coursera上面安德鲁ng的课程，笔记的进度也是按照上面每周的方式记的。）的笔记而已。发出来：1是为了自己复习巩固。2是为了方便自己查阅。3是为了给某... 查看全文 >

子楠 · 2 年前 · 发表于 机器学习笔记



《机器学习》笔记入门篇 | 详谈线性回归与逻辑回归算法

「摘要」不管是哪个行业的人，或多或少都能感受到人工智能那越来越强劲的风力，而机器学习作... 查看全文 >

京东白条 · 2 个月前

nsor Flc  
机器学习

一文打尽：线性回归和逻辑斯蒂线性回归

最近忙着弄毕业论文实验和一个竞赛，好久没发文章了，想着该发一篇了吧。。。正好最近也在看... 查看全文 >

王乐 · 2 个月前