

博客园 首页 博问 闪存 新随笔 订阅 XML 管理

posts - 10, comments - 18, trackbacks - 0

机器学习-----线性回归浅谈 (Linear Regression)

Linear Regreesion

在现实生活中普遍存在着变量之间的关系，有确定的和非确定的。确定关系指的是变量之间可以使用函数关系式表示，还有一种是属于非确定的（相关），比如人的身高和体重，一样的身高体重是不一样的。

线性回归：

1：函数模型（Model）：

$$h_w(x^i) = w_0 + w_1x_1 + wx_2 + \dots + w_nx_n$$

$$h_w(x^j) = w^T x_i = W^T X$$

$$X = \begin{bmatrix} 1 \\ x_1 \\ \dots \\ x_n \end{bmatrix} \quad W = \begin{bmatrix} w_0 \\ w_1 \\ \dots \\ w_n \end{bmatrix}$$

假设有训练数据

$$D = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$$

那么为了方便我们写成矩阵的形式

$$X = \begin{bmatrix} 1, x_1^1, x_1^2, \dots, x_1^n \\ 1, x_2^1, x_2^2, \dots, x_2^n \\ \dots \\ 1, x_n^1, x_n^2, \dots, x_n^n \end{bmatrix} \quad XW = h_w(x^i)$$

2：损失函数（cost）：

现在我们需要根据给定的X求解W的值，这里采用最小二乘法。

a.最小二乘法：

何为最小二乘法，其实很简单。我们有很多的给定点，这时候我们需要找出一条线去拟合它，那么我先假设这个线的方程，然后把数据点代入假设的方程得到观测

昵称：[GJS Blog](#)

园龄：[3年1个月](#)

粉丝：[24](#)

关注：[20](#)

[+加关注](#)

2017年6月						
≤	日	一	二	三	四	五
≥	六					
	28	29	30	31	1	2
	4	5	6	7	8	9
	11	12	13	14	15	16
	18	19	20	21	22	23
	25	26	27	28	29	30
	2	3	4	5	6	7
						8

搜索

常用链接

[我的随笔](#)

[我的评论](#)

[我的参与](#)

[最新评论](#)

[我的标签](#)

随笔分类

[Biology\(1\)](#)

[DM\(1\)](#)

[Ensemble Learning\(3\)](#)

[Java\(5\)](#)

[ML\(11\)](#)

[Weka \(4\)](#)

[算法（人工智能）\(4\)](#)

[网站\(1\)](#)

随笔档案

[2016年7月 \(3\)](#)

[2016年6月 \(1\)](#)

[2016年3月 \(3\)](#)

值，求使得实际值与观测值相减的平方和最小的参数。对变量求偏导联立便可求。

因此损失代价函数为：

$$J(W) = \frac{1}{2M} \sum_{i=0}^M (h_w(x^i) - y^i)^2 = \frac{1}{2M} (XW - y)^T (XW - y)$$

3：算法 (algorithm)：

现在我们的目的就是求解出一个使得代价函数最小的W：

a. 矩阵满秩可求解时 (求导等于0)：

$$\begin{aligned} L(W) &= \frac{1}{2} (XW - y)^T (XW - y) \\ &= \frac{1}{2} [W^T X^T XW - W^T X^T y - y^T XW + y^T y] \\ &= \frac{1}{2} [W^T X^T XW - 2W^T X^T y + y^T y] \end{aligned}$$

$$\frac{\partial L(W)}{\partial W} = 0$$

$$\frac{\partial L(W)}{\partial W} = \frac{1}{2} [W^T X^T X - 2X^T y] = 0$$

$$X^T XW = X^T y$$

$$W = (X^T X)^{-1} X^T y$$

b. 矩阵不满秩时 (梯度下降)：

梯度下降算法是一种求局部最优解的方法，对于F(x)，在a点的梯度是F(x)增长最快的方向，那么它的相反方向则是该点下降最快的方向，具体参考wikipedia。

原理：将函数比作一座山，我们站在某个山坡上，往四周看，从哪个方向向下走一小步，能够下降的最快；

注意：当变量之间大小相差很大时，应该先将他们做处理，使得他们的值在同一个范围，这样比较准确。

1) 首先对θ赋值，这个值可以是随机的，也可以让θ是一个全零的向量。

2) 改变θ的值，使得J(θ)按梯度下降的方向进行减少。

描述一下梯度减少的过程，对于我们的函数J(θ)求偏导J：

Repeat until convergence：{

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

[2015年5月 \(3\)](#)

[2015年2月 \(1\)](#)

[2015年1月 \(1\)](#)

[2014年11月 \(1\)](#)

[2014年10月 \(2\)](#)

[2014年9月 \(4\)](#)

[2014年8月 \(8\)](#)

[2014年7月 \(18\)](#)

[2014年5月 \(1\)](#)

最新评论

[1. Re:libsvm使用方法总结](#)

@asxinyu有空交流一下啊，以前我也常搞.net...

--GJS Blog

[2. Re:libsvm使用方法总结](#)

我用的是.net版本，自己修改了一下，方便在项目中自己调用，还不错

--asxinyu

[3. Re:机器学习-介绍](#)

@常建57其实我也刚刚接触机器学习，不过我感觉李航的统计学习方法对于我们其实不错。...

--GJS Blog

[4. Re:机器学习-介绍](#)

楼主和给位大神 有没有 文本分类，机器学习方面的好书推荐一下，中英文版本的都能推荐一下最好了。。。Thanks

--常建57

[5. Re:机器学习-介绍](#)

高大上的问题,Mark

--深蓝医生

阅读排行榜

[1. libsvm使用方法总结\(8270\)](#)

[2. 机器学习——线性回归浅谈 \(Linear Regression \) \(6837\)](#)

[3. Ensemble learning \(集成学习 \) \(4845\)](#)

[4. Java 从单核到多核的多线程 \(并发 \) \(3057\)](#)

[5. 分布式文件系统--GFS\(1291\)](#)

评论排行榜

[1. 机器学习-介绍\(6\)](#)

[2. 决策树介绍\(4\)](#)

[3. Weka 入门2\(2\)](#)

[4. java 小结2 多态问题和容器介绍\(2\)](#)

下面是更新的过程，也就是 θ_i 会向着梯度最小的方向进行减少。 θ_i 表示更新之前的值，-后面的部分表示按梯度方向减少的量， α 表示步长，也就是每次按照梯度减少的方向变化多少。

$$\theta_i = \theta_i - \alpha \frac{\partial}{\partial \theta} J(\theta) = \theta_i - \alpha (h_{\theta}(x) - y) x^{(i)}$$

}

假设有数据集D时：

$$X = \begin{bmatrix} 1, x_1^1, x_1^2, \dots, x_1^n \\ 1, x_2^1, x_2^2, \dots, x_2^n \\ \dots \\ 1, x_n^1, x_n^2, \dots, x_n^n \end{bmatrix} \quad W = \begin{bmatrix} w_0 \\ w_1 \\ \dots \\ w_n \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}$$

$$h = XW$$

对损失函数求偏导如下：

$$\frac{\partial}{\partial W_j} J(W) = \frac{1}{M} \sum_{i=1}^m (h_w^{x(i)} - y^{(i)}) x_j^i$$

$$W_j : W_j - \alpha \frac{\partial}{\partial W_j} J(W)$$

使用矩阵表示（方便计算）

$$E = X - y$$

$$E = \begin{bmatrix} h_1 - y_1 \\ h_2 - y_2 \\ \dots \\ h_n - y_n \end{bmatrix}$$

$$W = W - \frac{1}{M} \alpha X^T E$$

从概率层面解释-回归模型的目标函数：

基本上每个模型都会有一个对应的目标函数，可以通过不同的最优化求解方法（梯度下降，牛顿法等等）对这些对应的目标函数进行求解。线性回归模型，我们知道实际上是通过多个自变量对自变量进行曲线拟合。我们希望找到一条可以较好拟合的曲线，

那我们如何判断一条曲线的拟合程度的好坏。上面讲到，我们采用的是最小二乘法（预测值和真实值得误差的平方和），那为什么要用这个作为目标函数呢？

[5. java /小结3 hashCode和 equals l/o问题\(2\)](#)

推荐排行榜

- [1. 机器学习--介绍\(2\)](#)
- [2. Java 从单核到多核的多线程（并发）\(2\)](#)
- [3. 分布式文件系统-HDFS\(1\)](#)
- [4. boosting和bagging\(1\)](#)
- [5. 关于Eclipse连接MySQL的Jdbc\(1\)](#)

可以从中心极限定理、高斯分布来分析：

1.中心极限定理：

设有 n 个随机变量， $X_1, X_2, X_3, \dots, X_n$ ，他们之间相互独立，并且有相同的数学期望和均值。 $E(X)=\mu; D(x)=\sigma^2$ 。令 Y_n 为这 n 个随机变量之和。

$$Y_n = X_1 + X_2 + \dots + X_n$$

$$Z_n = \frac{Y_n - E(Y_n)}{\sqrt{D(Y_n)}} = \frac{Y_n - n\mu}{\sqrt{n}\sigma} \rightarrow N(0, 1)$$

Z_n 为 X 这几个变量的规范和。

2.高斯分布

假如给定一个输入样本 x ，我们得到预测值和真实值间的存在的误差 e ，那么他们的关系如下：

$$y^{(i)} = w^T x^{(i)} + \varepsilon^{(i)}$$

而这里，我们就可以假设 e 服从标准的高斯分布。

为什么呢？回归模型的最终目标是建立自变量 x 和 y 之间的关系，我们希望通过 x 可以较为准确的表示结果 y 。而在实际应用场景中，很难甚至不可能把导致 y 结果的所有变量（特征）都找到，放到回归模型里面。

我们只存放那些认为比较重要的特征。根据中心极限定理，把那些对结果影响比较小的（假设独立分布）之和认为是符合正态分布是合理的。

$$\varepsilon^{(i)} = y^{(i)} - w^T x^{(i)} \rightarrow N(0, \sigma^2)$$

那么 x 和 y 的条件概率：

$$Y_n = X_1 + X_2 + \dots + X_n$$

$$Z_n = \frac{Y_n - E(Y_n)}{\sqrt{D(Y_n)}} = \frac{Y_n - n\mu}{\sqrt{n}\sigma} \rightarrow N(0, 1)$$

那么知道一条样本的概率，我们就可以通过极大估计求似然函数，优化的目标函数如下：

$$\max_w L(w) = \prod_{i=1}^m \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y^{(i)} - wx)^2}{2\sigma^2}\right)$$

通过取对数我们可以发现极大似然估计的目标函数和最小平方误差是一样。

在概率模型中，目标函数的极大和极小与极大似然估计是等价的。

假设随机变量为 Y ，和普通变量 x 存在相关关系，由于 Y 是随机变量，对于 x 的各个确定值， Y 有它的分布（高斯）。

假设为：

$$Y \sim N(W^T X, \sigma^2)$$

$$Y = W^T X + \varepsilon, \quad \varepsilon \sim (0, \sigma^2)$$

使用极大似然估计可求解。

我们知道对于下面公式：

$$\sum_{i=1}^m (y_i - c_i)^2$$

y 为随机变量，在 $c=E(y)$ 时达到最小，这表明以 $E(y)$ 作为 y 的近似是最好的。

从菜鸟走向大神，这是道路。

分类: [ML](#)

好文要顶

关注我

收藏该文



[GJS Blog](#)

关注 - 20

粉丝 - 24

[+加关注](#)

1

0

« 上一篇: [分布式文件系统-HDFS](#)

» 下一篇: [Java 从单核到多核的多线程（并发）](#)

posted on 2014-08-21 21:48 [GJS Blog](#) 阅读(6837) 评论(0) [编辑](#) [收藏](#)

[刷新评论](#) [刷新页面](#) [返回顶部](#)

注册用户登录后才能发表评论，请 [登录](#) 或 [注册](#)，[访问网站首页](#)。

[【推荐】50万行VC++源码：大型组态工控、电力仿真CAD与GIS源码库](#)

[【阿里云】云计算科技红利邀您免费体验！云服务器、云数据库等35+产品，6个月免费使用！](#)

[【免费】从零开始学编程，开发者专属实验平台免费实践！](#)



最新IT新闻：

- [任正非最新讲话：强调精神文明不是光喊口号不发钱](#)
 - [NASA的实验性太阳能电池阵列已出现在太空中](#)
 - [被问阿里对手是谁 马云说我想不出](#)
 - [谷歌开源T2T模型库，深度学习系统进入模块化时代！](#)
 - [FB图片分享应用日活跃用户达2.5亿 碾压Snapchat](#)
- » [更多新闻...](#)



最新知识库文章：

- [小printf的故事：什么是真正的程序员？](#)
 - [程序员的工作、学习与绩效](#)
 - [软件开发为什么很难](#)
 - [唱吧DevOps的落地，微服务CI/CD的范本技术解读](#)
 - [程序员，如何从平庸走向理想？](#)
- » [更多知识库文章...](#)

Copyright ©2017 GJS Blog Powered By[博客园](#) 模板提供：[沪江博客](#)