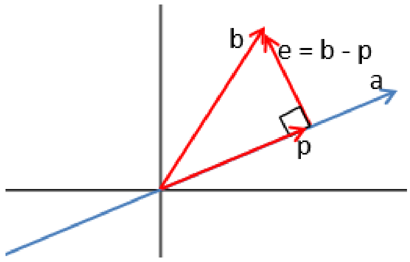
线性代数笔记18——投影矩阵和最小二乘

一维空间的投影矩阵

先来看一维空间内向量的投影：



向量p是b在a上的投影，也称为b在a上的分量，可以用b乘以a方向的单位向量来计算，现在，我们打算尝试用更“贴近”线性代数的方式表达。

因为p趴在a上，所以p实际上是a的一个子空间，可以将它看作a放缩x倍，因此向量p可以用p = xa来表示，只要找出x就可以了。因为a⊥e，所以二者的点积为0：

$$a \cdot e = a^T e = a^T(b-p) = a^T(b-xa) = 0$$

我们希望化简这个式子从而得出x：

$$a^T(b-xa) = a^T b - a^T xa = a^T b - a^T xa = 0$$

x是一个实数，进一步得到x：

$$xa^T a = a^T b$$

$$x = \frac{a^T b}{a^T a}$$

$a^T b$和$a^T a$都是点积运算，最后将得到一个标量数字。这里需要抑制住消去a^T的冲动，向量是不能简单消去的，a和b都是2×1矩阵，矩阵的运算不满足乘法交换律，a^T无法先和$1/a^T$计算。

现在可以写出向量p的表达式，这里的x是个标量：

$$p = xa = ax = a\frac{a^T b}{a^T a}$$

这就是b在a上的投影了，它表明，当b放缩时，p也放缩相同的倍数；a放缩时，p保持不变。

由于向量点积$a^T a$是一个数字，p可以进一步写成：

$$p = a\frac{a^T b}{a^T a} = \frac{a(a^T b)}{a^T a} = \frac{(aa^T)b}{a^T a} = \frac{aa^T}{a^T a}b$$

随笔分类 (211)

★★资源下载★★(1)
Java并发编程(1)
程序员的数学(24)
单变量微积分(31)
多变量微积分(24)
概率(24)
机器学习(27)
软件设计(1)
数据分析(6)
数据结构与算法(27)
随笔(5)
线性代数(34)
项目管理(2)
转载(4)

随笔档案 (205)

2021年2月(1)
2020年3月(2)
2020年2月(6)
2020年1月(4)
2019年12月(7)
2019年11月(15)
2019年9月(3)
2019年8月(6)
2019年7月(1)
2019年6月(8)
2019年5月(3)
2019年4月(5)
2019年3月(7)
2019年2月(3)
2019年1月(7)
更多

阅读排行榜

1. 使用Apriori进行关联分析（一）(29768)
2. 线性代数笔记12——列空间和零空间(28772)
3. FP-growth算法发现频繁项集（一）——构建FP树(24430)
4. 寻找“最好”（2）——欧拉-拉格朗日方程(23099)
5. 多变量微积分笔记3——二元函数的极值(22772)

评论排行榜

1. 隐马尔可夫模型（一）(8)
2. 线性代数笔记12——列空间和零空间(7)
3. 线性代数笔记3——向量2（点积）(7)
4. FP-growth算法发现频繁项集（一）——构建FP树(5)
5. 寻找“最好”（2）——欧拉-拉格朗日方程(4)

推荐排行榜

1. 寻找“最好”（2）——欧拉-拉格朗日方程(7)
2. FP-growth算法发现频繁项集（一）——构建FP树(7)
3. 线性代数笔记3——向量2（点积）(6)
4. FP-growth算法发现频繁项集（二）——发现频繁项集(5)
5. 隐马尔可夫模型（一）(5)

最新评论

1. Re:线性代数笔记3——向量2（点积）
如果点积小于0，即夹角小于90°，这个写错了吧。应该是夹角大于90°
--猫猫猫猫猫大人

在一维空间中，分子是一个2×2矩阵，这说明向量b的在a上的投影p是一个矩阵作用在b上得到的，这个矩阵就叫做投影矩阵（Projection Matrix），用大写的P表达：

$$p = Pb = \frac{aa^T}{a^Ta}b, \qquad P = \frac{aa^T}{a^Ta}$$

推广到n维空间，a是n维向量，投影矩阵就是n×n的方阵。观察投影矩阵会法发现，它是由一个列向量乘以一个行向量得到的：

$$aa^T = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix} = \begin{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} x_1 & \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} x_2 \end{bmatrix}$$

可以看出aa^T的列向量是线性相关的，所以它的列空间和行空间的维度都是1，这说明它的秩是1，aa^T是一个秩一矩阵，并且是一个对称矩阵。由于a^Ta是一个标量，所以aa^T决定了投影矩阵的性质：

$$P = P^T, \qquad rank(P) = 1$$

投影矩阵还有另外一个性质：

$$P = P^2$$

它的几何意义是，对一个向量投影两次和投影一次相同，b在a上的投影是p，再投影一次仍然是p。

投影的意义

向量投影到子空间有什么意义呢？这要从方程Ax = b说起。

对于Ax = b来说，并不是任何时候都有解，实际上大多数这种类型的方程都无解。A的列空间的含义是方程组有解时b的取值空间，当b不在A的列空间时，方程无解。具体来说，当A是行数大于列数的长方矩阵时，方程组中的方程大于未知数个数，肯定无解。

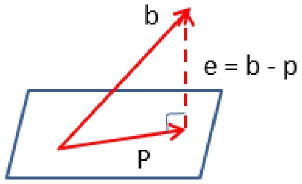
虽然方程无解，但我们还是希望能够运算下去，这就需要我们换个思路——不追求可解，转而寻找最接近可解问题的解。对于无解方程Ax = b，Ax总是在列空间里（因为列空间本来就是由Ax确定的，和b无关），而b就不一定了，所以需要微调b，将b变成列空间中最接近它的一个，Ax = b变成了：

$$A\hat{x} = p$$

p就是A的列空间的投影， x帽表示此时的x已经不是原来Ax = b中的x，$A\hat{x} \neq b$，当然，因为方程无解，所以本来也不可能有Ax = b。此时问题转换为寻找最好的，让它与原方程的差距最小：

$$A\hat{x} = p \approx b$$

假设A的秩是2，此时列空间的维度也是2，列空间将是一个平面，平面上的向量有无数个，其中最接近b的当然是b在平面上的投影，只有b – p才能产生最小的误差向量：



所以说，想要求解不等方程Ax = b，需要将b微调成它在A的列空间上的投影（列空间上的向量很多，b在列空间上的投影是唯一的），这就是投影的意义。

从上图可知，如果b在列空间中，b的投影就是它自己；如果b正交于列空间，它的投影是0。

投影矩阵的多维推广

向量b在子空间上的投影是b在向量上投影的推广。如果A的列空间是一个二维平面，则列空间的基包含两个向量，我们用a_1和a_2表示。在一维空间中，b的投影是p = xa，多维空间也一样，只不过a从向量变成了矩阵，x从标量变成了向量：

$$p = a_1\widehat{x_1} + a_2\widehat{x_2} = A\hat{x}$$

我们知道如何找到列空间的基向量， 如何求解呢？

误差向量e垂直于列空间的平面，这意味它垂直于平面上的任意向量，投影p和基向量a_1，a_2都在平面上，这意味着它们和e的点积为0：

$$a_1 \cdot e = a_1^T(b - p) = 0$$

$$a_2 \cdot e = a_2^T(b - p) = 0$$

将二者归纳为一个矩阵方程：

$$\begin{bmatrix} a_1^T \\ a_2^T \end{bmatrix}(b - p) = \begin{bmatrix} a_1^T \\ a_2^T \end{bmatrix}(b - A\hat{x}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$A^T(b - A\hat{x}) = O$$

对比平面上的投影方程，发现二者很相似：

$$A^T(b - A\hat{x}) = A^T e = O$$

$$a^T(b - ax) = a^T e = 0$$

在一维空间中a是向量，x是标量；在多维空间中，a变成了矩阵A，x变成了向量，它们都长高了。

误差向量e是否属于A的某个基本子空间呢？在上一章中提到过，列空间和零空间正交，e垂直于列空间上的所有向量，所以e属于A的零空间，也就是$b - A\hat{x}$属于a的零空间。

通过矩阵方程可以进一步求解投影向量和投影矩阵：

$$A^T(b - A\hat{x}) = 0$$

$$A^TA\hat{x} = A^Tb$$

$$\hat{x} = (A^TA)^{-1}A^Tb$$

$$p = A\hat{x} = A(A^TA)^{-1}A^Tb$$

$$matrix\ \ P = A\hat{x} = A(A^TA)^{-1}A^T$$

这就是多维空间的投影矩阵，满足和一维空间同样的性质：

$$P = P^T, \qquad P = P^2$$

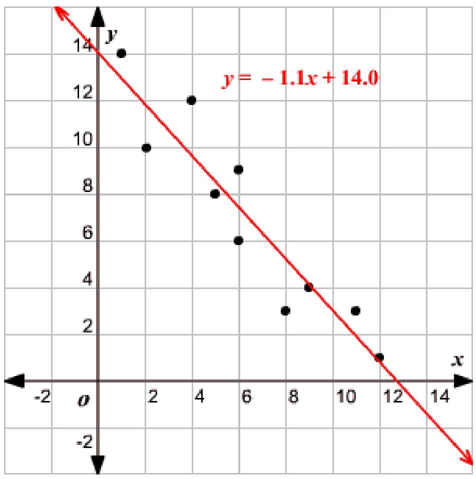
投影矩阵的表达式并不友好，也许可以进一步化简：

$$P = A(A^TA)^{-1}A^T = AA^{-1}(A^T)^{-1}A^T = I$$

看起来就不对，所以一定要抑制住化简的冲动——这里A不是方阵，所以A^{-1}不存在。

最小二乘法

最小二乘是投影矩阵的重要应用，常用于回归分析中。一元线性回归是回归分析中最简单的一种，它有一个自变量和一个因变量，能够根据一系列训练数据$(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)})\ldots(x^{(n)}, y^{(n)})$找出一条最佳拟合直线y = ax + b，用这条直线作为模型，近似地表示x和y的关系，从而对新鲜样本进行预测。下图是一条典型的拟合直线：



《多变量微积分2——最小二乘法》和《寻找“最好”（5）——无解之解》从微积分和约等方程的视角解释过最小二乘，其实用线性代数的观点解释会更加直观。

最小二乘建立了一个约等方程组，目标是使总误差最小化，约等方程组可以用矩阵写成：

$$Ax = b$$

问题是A是一个长方矩阵，想要求解必须将A转换为方阵，方法是等式两侧同时左乘A^T：

$$A^TA\hat{x} = A^Tb$$

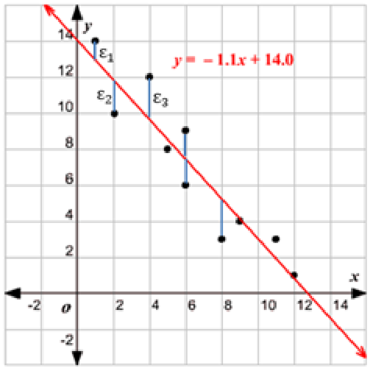
$$\hat{x} = (A^TA)^{-1}A^Tb$$

这就和投影矩阵的相关公式相同，可以直接求解了。

最小二乘的几何意义也可以用投影来解释，平方和损失函数用矩阵表示是：

$$\|A\hat{x} - b\|^2 = \|e\|^2 = e_1^2 + e_2^2 + \cdots + e_n^2$$

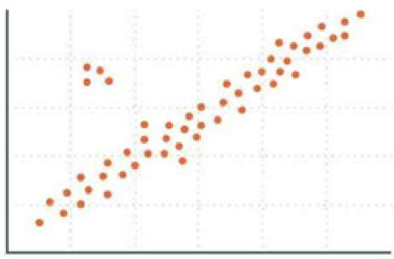
$A\hat{x}$ 正是投影矩阵，整个式子表示所有向量与这些向量在列空间上的投影的误差的平方和，它的几何意义是所有点到拟合直线的距离之和最小：



最后的问题是A^TA是否是可逆的，如果A的各列线性无关，A^TA就是可逆的，这在《线性代数笔记12——列空间和零空间》中有所说明。实际应用中，每个列代表的含义都不同，比如在房价预测中，第一列代表房屋面积，第二列代表房间数，第三列代表楼层数，这就可以确定A的各列一定是线性无关的。

害群之马

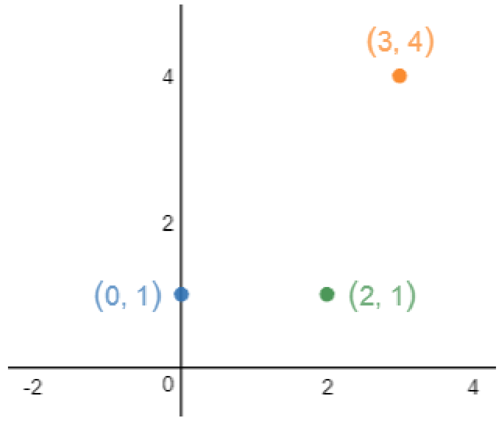
有一个因素会严重影响数据拟合：



一部分数据点极不合群，它们远离了大部队，就像害群之马一样，会拉低整体的平均素质。这样的数据在统计学中称为离群量，也叫逸出量，在数据分析中称为数据噪声。实际上数据噪声很常见，比如预测某个大都市里电视机的价格走势，如果把双十一或店铺周年庆的价格也统计在内，就会严重影响预测结果，所以在拟合数据时，还需要预先去除数据噪声。

示例

假设有 (0, 1), (2, 1), (3, 4)三组数据，现在尝试使用直线对他们进行拟合，数据点的分布如下所示：



直线的方程是y = ax + b，通过图中的三点建立一个方程组：

$$\begin{cases} 0a + b = 1 \\ 2a + b = 1 \\ 3a + b = 4 \end{cases}$$

三个方程两个未知数，这是个无解的方程组，说明三点并不在同一条直线上，我们需要寻找一条能够使三点误差和最小的直线。

将最初的方程组转换为 的形式：

$$\begin{cases} 0a + b = 1 \\ 2a + b = 1 \\ 3a + b = 4 \end{cases} \Rightarrow \underbrace{\begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_{\hat{x}} = \underbrace{\begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}}_{b}$$

可以根据投影矩阵的知识直接求解了：

$$\hat{x} = (A^T A)^{-1} A^T b$$

最终a = 6/7，b = 4/7

现在更改一下需求，将数据拟合为曲线y = ax2 + bx +c。把三点代入假设函数，得到一个方程组：

$$\begin{cases} 0a + 0b + c = 1 \\ 4a + 2b + c = 1 \\ 9a + 3b + c = 4 \end{cases}$$

实际上这个方程组有唯一解：

$$\begin{cases} a = 1 \\ b = -2 \\ c = 1 \end{cases}$$

即便如此，我们仍然想用最小二乘的思想求解并验证一下。

$$\begin{cases} 0a + 0b + c = 1 \\ 4a + 2b + c = 1 \\ 9a + 3b + c = 4 \end{cases} \Rightarrow \underbrace{\begin{bmatrix} 0 & 0 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a \\ b \\ c \end{bmatrix}}_{\hat{x}} = \underbrace{\begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}}_{b}$$

无论是通过消元法还是使用投影矩阵，最终都将得到唯一解。

作者：我是8位的

出处：http://www.cnblogs.com/bigmonkey

随笔

分类: 线性代数

标签: 投影矩阵, 向量投影, 最小二乘

« 上一篇： 线性代数笔记17——正交向量与正交子空间
» 下一篇： 寻找“最好”（8）——牛顿法