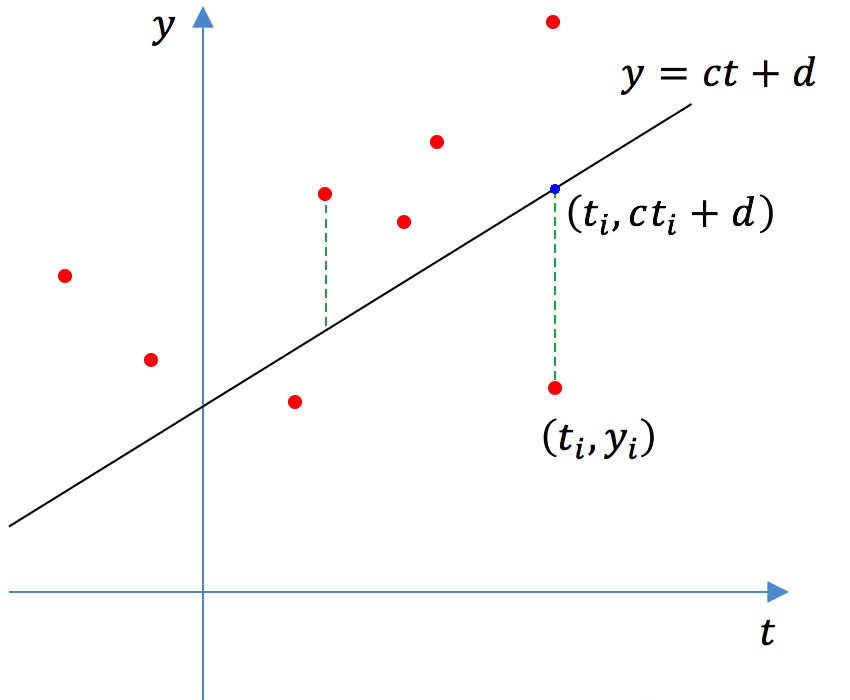
# 最小二乘问题（上）

- 白马负金羁 - 博客频道 - CSDN.NET

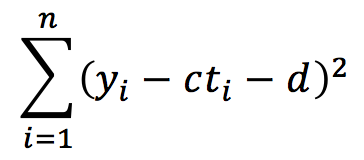
http://blog.csdn.net/baimafujinji/article/details/6476368

**一、问题的引入**

我们同样从回归问题入手引入本文的话题。假设在某些时间点*t*i上，观察到一些数值*y*i，于是有一组二维数据点(*t*i,*y*i), i = 1, 2, ..., n。现在的任务是find the "best" line to represent those data。如下图所示，假设我们的备选直线之方程为 y=ct+d。在*t*i时刻，实际观测值为*y*i，而根据回归直线我们的估计值应该是*ct*i+*d*。



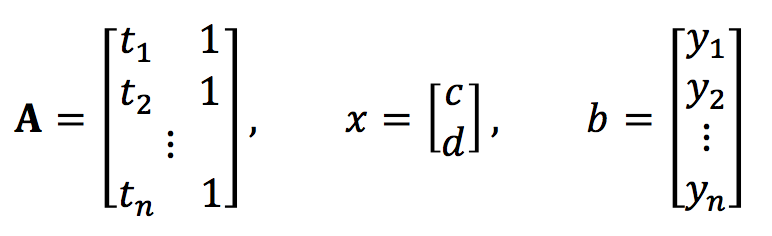
因此，实际观测值与估计值之间的误差就可以表示为|*y*i - *ct*i - *d*|。既然我们的任务是找到the "best" line，那么就可以用“总误差”最小来作为衡量"the best"的标准。而且等价地，我们用平方来取代绝对值计算，注意(*t*i,*y*i)是已知的，于是现在的目标变为求使得下式（总误差）最小的参数c和d：



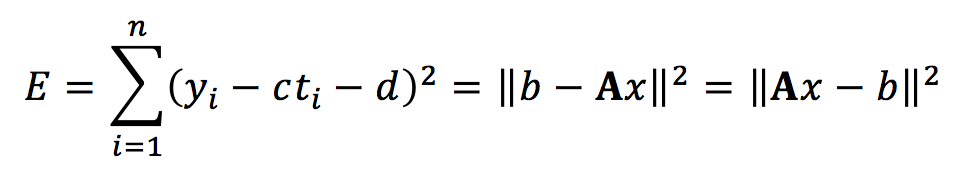
一种可以想到的解决方法就是采用微积分的方法（可以参考《R语言实战：[**机器学习**](http://lib.csdn.net/base/machinelearning)与数据分析》）。今天我们要介绍另外一种基于矩阵的解决方法。矩阵方法与微积分的方法所得之结论都是一样的，但是基于矩阵的方法更加powerful，而且更容易推广。

**二、矩阵形式的引入**

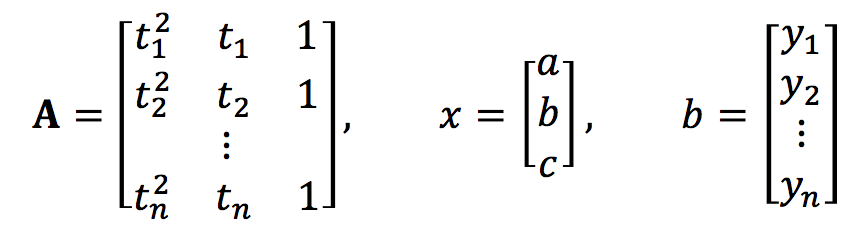
下面来看看如何把上述优化问题转变为矩阵问题。令



此时问题就变成如下这样：

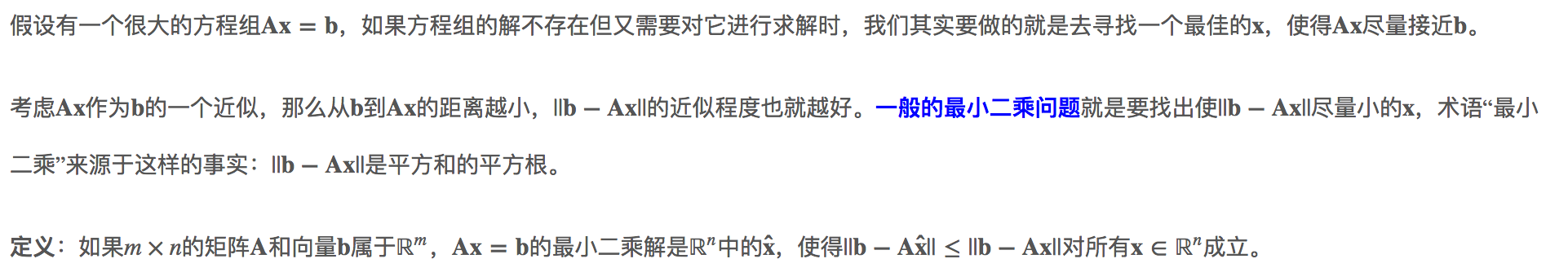


可见这个形式也简单明了。不仅如此，矩阵也非常方便问题的推广，比如说现在我们不是要找直线，而是一个用多项式所表示的曲线（这时问题就变成了多元回归问题），例如二次曲线 y=a*t*2+b*t*+c。这时其实我们只要像下面这样简单地改变一下矩阵的形式即可



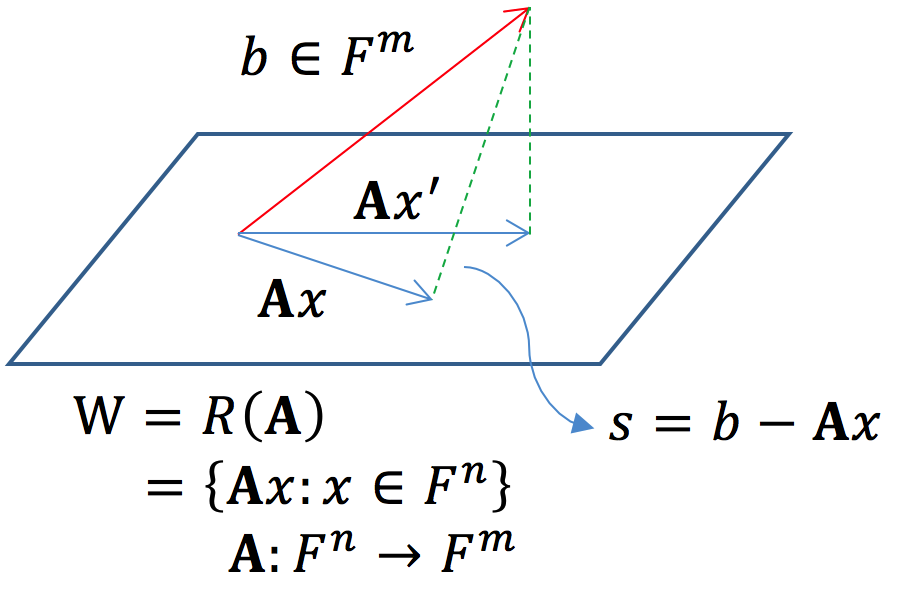
事实上，你已经可以看出对于多元回归问题，我们最终要面对的问题都是要求一个||**A**x-b||2的最佳逼近，采用线性代数的方法，我们就将一个相当复杂的问题统一到一个相对简单的形式上来。

**三、最小二乘问题**



或者用我们更加熟悉的表述方式，还可以把最小二乘问题重述如下：Given A∈Mm×n(F), and b ∈ Fm， find an x' ∈Rn，使得 ||**A**x'-b|| is a minimum。（特别地，对于回归问题而言，就是Given a set (*x*i,*y*i), i = 1, 2, ..., n，of measurements, find a "best" polynomial of degree n-1 to represent such set of measurements）

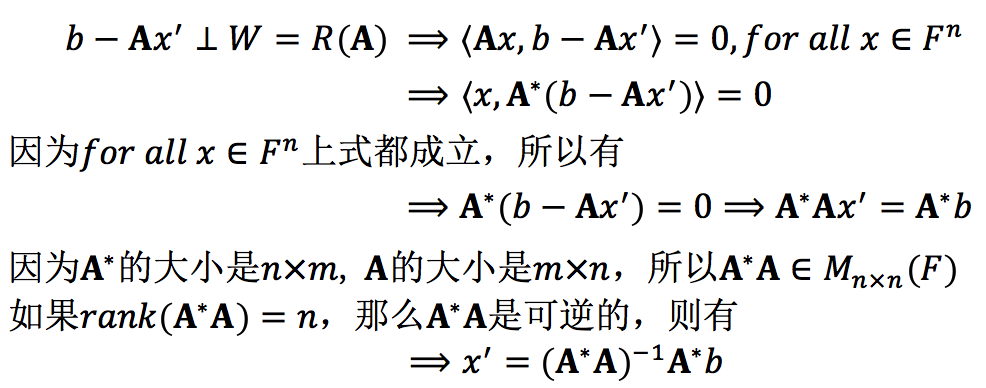
问题说到这其实答案已经很显然了，只要利用我们前面介绍过的“最佳逼近原理”（http://blog.csdn[**.NET**](http://lib.csdn.net/base/dotnet)/baimafujinji/article/details/6478123）即可！如下图所示，子空间W是A的值域，即W中的每一个元素都是Ax。b是Fm空间中的一个元素，如果我们要在W中找一个使得||b-Ax||最小的x'，显然x'就是b在W中的投影。



特别地，如果rank(A)=n，则有唯一解 x' = (A\*A)-1A\*b。

Remark: Let *x*i=1,2,...m be all distinct. Assume m≥n，则有rank(A)=n。

下面我们来推导这个结论。我们声称 x' is the solution to the problem. 意思就是需要证明 ||Ax'-b||≤||Ax-b|| for all any x ∈ Fn。证明：||Ax-b||2 = ||Ax'-b+Ax-Ax'||2，根据勾股定理可得， ||Ax'-b+Ax-Ax'||2 =||Ax'-b||2+||Ax-Ax'||2≥||Ax'-b||2。



上述过程中的几个步骤可以实现，由下面两个引理来保证：

http://img.blog.csdn.net/20161219210145548

http://img.blog.csdn.net/20161219210046687

综上所述，结论得证。

在《最小二乘问题》系列文章的下篇中，我们将给出一些实际计算的例子（http://blog.csdn[**.Net**](http://lib.csdn.net/base/dotnet)/baimafujinji/article/details/53313784）

(本文完)

本文主要根据台湾交通大学开放课程线性代数（莊重 特聘教授主讲）之授课内容整理，并参考以下书籍：

【1】S.H. Friedberg, A.J. Insel, L.E Spence, 4th edition, Linear Algebra, Prentice-Hall, 2003

【2】David C. Lay. 刘深泉，等译. 线性代数及其应用（原书第3版），机械工业出版社，2005

# 2@最小二乘问题（下）

- 白马负金羁 - 博客频道 - CSDN.NET

http://blog.csdn.net/baimafujinji/article/details/53313784?locationNum=2&fps=1

**引言**

牛顿说：“把简单的问题看得复杂，可以发现新领域；把复杂的问题看得简单，可以发现新规律。”而从历史的角度来看，一个学科的发展也亦是如此。随着学科的发展，最开始的一个主干方向会不断衍生出各自相对独立的分支，这也就是所谓“把简单的问题看得复杂”的过程。然而，一旦学科发展到一定程度之后，某些分支学科又开始被抽象综合起来，这也就是所谓“把复杂的问题看得简单”的过程。

例如，在很长一段时间里，物理学家们都把电和磁看成是两种独立的物理现象在研究，当学科研究积累到一定程度时，麦克斯韦就创立了电磁学从而完成了物理学中的一次大综合。而在数学发展的历史中，几何与代数也曾经在很长的一段时间里是彼此独立的。直到笛卡尔引入了直角坐标系的概念之后，人们才开始建立了一种代数与几何之间的联系，也就是所谓的解析几何。

泛函分析也是对以往许多数学问题或者领域进行高度抽象和综合的结果，其主要研究对象之一是抽象空间。其实在学习线性代数的过程中，读者已经建立了一种从矩阵到线性方程组之间的一种联系。而在泛函分析中，实数系、矩阵、多项式以及函数族这些看似关联不大的概念都可以抽成空间。

最小二乘法是求解线性回归问题时会介绍到的一种重要方法。本文会跳出线性回归的框架，讨论更泛化的最小二乘问题。而我们审视这个问题的高度，你可以认为是从线性代数或者矩阵分析的层级出发的。但如果你感觉我们所讨论的术语跟过去线性代数里学过的好像不太一样，那是因为我们正试图从泛函的角度来解读它。

**最小二乘问题**

假设有一个很大的方程组Ax=b，如果方程组的解不存在但又需要对它进行求解时，我们其实要做的就是去寻找一个最佳的x，使得Ax尽量接近b。

考虑Ax作为b的一个近似，那么从b到Ax的距离越小，||b−Ax||的近似程度也就越好。**一般的最小二乘问题**就是要找出使||b−Ax||尽量小的x，术语“最小二乘”来源于这样的事实：||b−Ax||是平方和的平方根。

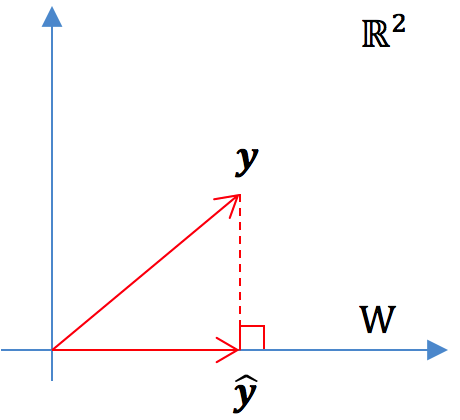
**定义**：如果m×n的矩阵A和向量b属于Rm，Ax=b的最小二乘解是Rn中的x^，使得||b−Ax^||≤||b−Ax||对所有x∈Rn成立。

请大家回忆一下列空间的概念。假设大小为m×n的矩阵A=[a1,⋯,an]，此处ai表示矩阵中的一列，其中1≤i≤n，那么矩阵A的列空间（记为ColA）就是由A中各列的所有线性组合构成的集合，即ColA=Span{a1,⋯,an}。

很显然，ColA=Span{a1,⋯,an}是一个子空间，进而我们可以知道m×n的矩阵A的列空间是Rm的一个子空间。此外，我们还注意到ColA中的一个典型向量（或者称为元素）可以写成Ax的形式，其中x是某个向量，记号Ax表示A的列向量的一个线性组合，也就是说ColA={b :b=Ax, x∈Rn}，于是该式告诉我们：1）记号Ax代表ColA空间中的向量；2）ColA是线性变换 x↦Ax 的值域。

假设W是 Rn 空间中的一个子空间，y是Rn中的任意向量，y^是y在W上的正交投影，那么y^是W中最接近y的点，也就是说||y−y^||<||y−v||对所有属于W又异于y^的v成立。这个结论也称为最佳逼近原理。其中向量y^称为W中元素对y的最佳逼近。对于给定元素y，可以被某个给定子空间中的元素代替或“逼近”，用||y−v||表示的从y到v的距离，可以认为是用v代替y的“误差”，而最佳逼近原理说明误差在v=y^处取得最小值。

最佳逼近原理看似抽象，其实它的几何意义也是相当直观的。如下图所示，我们所取的向量空间为R2，其中的一个子空间W就是横轴，y是R2中的任意向量，y^是y在W上的正交投影，那么y^显然是W中最接近y的点。更要的是，这个结论不仅仅在欧几里得几何空间中成立，因为泛函中所讨论的“向量”可能是我们通常所说的向量，但也可能是函数、矩阵，甚至多项式等等，但最佳逼近原理皆成立。



通过前面的分析，我们知道，最小二乘问题的最重要特征是无论怎么选取x，向量Ax必然位于列空间ColA中。而最小二乘问题的本质任务就是要寻找一个x，使得Ax是ColA中最接近b的向量。

**一般最小二乘问题的解**

对于上面给定的A和b，运用最佳逼近定理于ColA空间，取 

b^=projColAb

由于b属于A的列空间，方程Ax=b^是相容的且存在一个属于Rn 的x^使得Ax^=b^。由于b^是ColA中最接近b的点，一个向量x^是Ax=b的一个最小二乘解的成分必要条件是x^满足Ax^=b^。这个属于Rn的x^是一系列由A构造b^的权值。（显然如果方程Ax^=b^中存在自由变量，则这个方程会有多个解。）

如果x^满足Ax^=b^，根据正交分解定理，投影b^具有这样的一个性质：(b−b^)与ColA正交，即(b−Ax^)正交于A的每一列，如果aj是A的任意列，那么aj⋅(b−Ax^)=0且aTj⋅(b−Ax^)=0，由于每一个aj是AT的行，于是可得 

AT(b−Ax^)=0

因此有 

ATb−ATAx^=0

或者 

ATAx^=ATb

这其实表明Ax=b的每个最小二乘解满足方程 

ATAx=ATb

上述矩阵方程表示的线性方程组常称为Ax=b的法方程，其解通常用x^表示。最后我们还可以得出如下定理（证明从略）。   
**定理**：方程Ax=b的最小二乘解集和法方程ATAx=ATb的非空解集一致。

下面我们通过一个例子来说明最小二乘问题的解法。例：求不相容方程Ax=b的最小二乘解，其中 

A=⎡⎣⎢401021⎤⎦⎥,b=⎡⎣⎢2011⎤⎦⎥

利用前面给出的公式计算 

ATA=[400211]⎡⎣⎢401021⎤⎦⎥=[17115]

ATb=[400211]⎡⎣⎢2011⎤⎦⎥=[1911]

那么法方程ATAx=ATb就变成了 

[17115][x1x2]=[1911]

行变换可用于解此方程组，但由于ATA是2×2的可逆矩阵，于是很容易得到 

(ATA)−1=184[5−1−117]

那么可以解ATAx=ATb如下： 

x^=(ATA)−1ATb=184[5−1−117][1911]=[12]

在很多计算中，ATA是可逆的，但也并非都是如此，下面例子中的矩阵常常出现在统计学中的方差分析问题里。

例：求Ax=b的最小二乘解，其中 

A=⎡⎣⎢⎢⎢⎢⎢⎢⎢⎢111111110000001100000011⎤⎦⎥⎥⎥⎥⎥⎥⎥⎥,b=⎡⎣⎢⎢⎢⎢⎢⎢⎢⎢−3−10251⎤⎦⎥⎥⎥⎥⎥⎥⎥⎥

类似地，我们可以算得 

ATA=⎡⎣⎢⎢⎢110011001010101010011001⎤⎦⎥⎥⎥⎡⎣⎢⎢⎢⎢⎢⎢⎢⎢111111110000001100000011⎤⎦⎥⎥⎥⎥⎥⎥⎥⎥=⎡⎣⎢⎢⎢6222220020202002⎤⎦⎥⎥⎥

ATb=⎡⎣⎢⎢⎢110011001010101010011001⎤⎦⎥⎥⎥⎡⎣⎢⎢⎢⎢⎢⎢⎢⎢−3−10251⎤⎦⎥⎥⎥⎥⎥⎥⎥⎥=⎡⎣⎢⎢⎢4−426⎤⎦⎥⎥⎥

矩阵方程ATAx=ATb的增广矩阵为： 

⎡⎣⎢⎢⎢62222200202020024−426⎤⎦⎥⎥⎥∼⎡⎣⎢⎢⎢1000010000101−1−103−5−20⎤⎦⎥⎥⎥

通解是：x1=3−x4，x2=−5+x4，x3=−2+x4，x4是自由变量。   
所以，Ax=b的最小二乘通解具有如下形式： 

x^=⎡⎣⎢⎢⎢3−5−20⎤⎦⎥⎥⎥+x4⎡⎣⎢⎢⎢−1111⎤⎦⎥⎥⎥

在什么条件下，方程Ax=b的最小二乘解是唯一的？下面的定理给出了判断准则（当然，正交投影总是唯一的），我们不具体讨论该定理的证明。   
**定理**：矩阵ATA是可逆的，其充分必要条件是 A 的列是线性无关的，在这种情况下，方程Ax=b有唯一最小二乘解x^且它有下面的表达式 

x^=(ATA)−1ATb

**最小二乘解的另一个计算**

下面的例子表明，当A的列向量不正交时，该如何找到Ax=b的最小二乘解。这类矩阵在线性回归中常被用到。

例：找出Ax=b的最小二乘解，其中 

A=⎡⎣⎢⎢⎢1111−6−217⎤⎦⎥⎥⎥,b=⎡⎣⎢⎢⎢−1216⎤⎦⎥⎥⎥

解：由于A的列a1和a2相互正交，b在ColA的正交投影如下： 

b^=b⋅a1a1⋅a1⋅a1+b⋅a2a2⋅a2⋅a2=84⋅a1+4590⋅a2=⎡⎣⎢⎢⎢2222⎤⎦⎥⎥⎥+⎡⎣⎢⎢⎢⎢−3−11/27/2⎤⎦⎥⎥⎥⎥=⎡⎣⎢⎢⎢⎢−115/211/2⎤⎦⎥⎥⎥⎥

既然b^已知，我们可以解Ax^=b^。这个很容易，因为我们已经知道b^用A的列线性（通过线性组合来）表示时的权值。于是从上式可以立刻得到： 

x^=⎡⎣⎢⎢844590⎤⎦⎥⎥=⎡⎣212⎤⎦

某些时候，最小二乘解问题的法方程可能是病态的，也就是ATA中元素在计算中出现较小的误差，可导致解x^出现较大的误差。如果A的列线性无关，最小二乘解常常可通过A的QR分解更可靠地求出。来看最后一个定理。

**定理**：给定一个m×n的矩阵A，且具有线性无关的列，取A=QR是A的QR分解，那么对每一个属于Rm的b，方程Ax=b有唯一的最小二乘解，其解为 

x^=R−1QTb

这个定理的证明非常简单，我们不再赘述。更进一步，基于QR分解的知识，我们知道Q的列形成Col\mathbf{A}的正交基，因此，QQTb是b在ColA上的正交投影b^，那么Ax^=b^说明x^是Ax=b的最小二乘解。

此外，由于上述定理中的R是上三角形矩阵，x^可从方程 Rx=QTb计算得到。求解该方程时，通过回代过程或行变换会比较高效。

例：求出Ax = b的最小二乘解，其中

A=⎡⎣⎢⎢⎢111131135023⎤⎦⎥⎥⎥,b=⎡⎣⎢⎢⎢357−3⎤⎦⎥⎥⎥

解：我们可以计算矩阵A的QR分解为 

A=QR=⎡⎣⎢⎢⎢⎢1/21/21/21/21/2−1/2−1/21/21/2−1/21/2−1/2⎤⎦⎥⎥⎥⎥⎡⎣⎢200420532⎤⎦⎥

那么 

QTb=⎡⎣⎢1/21/21/21/2−1/2−1/21/2−1/21/21/21/2−1/2⎤⎦⎥⎡⎣⎢⎢⎢357−3⎤⎦⎥⎥⎥=⎡⎣⎢6−64⎤⎦⎥

满足Rx=QTb的最小二乘解是x^，也就是 

⎡⎣⎢200420532⎤⎦⎥⎡⎣⎢x1x2x3⎤⎦⎥=⎡⎣⎢6−64⎤⎦⎥

这个方程很容易解出得： 

x^=⎡⎣⎢10−62⎤⎦⎥

**参考文献与推荐阅读材料**

本文中的例子主要来自   
[1] David C. Lay. 刘深泉，等译. 线性代数及其应用（原书第3版），机械工业出版社，2005   
同时推荐如下视频课程（从泛函的角度讲解线性代数）：   
[2] 台湾国立交通大学开放课程：线性代数（一）、（二），莊重教授主讲，<http://ocw.nctu.edu.tw/>

# （数学）最小二乘的几何意义及投影矩阵

- AndyJee - 博客园

http://www.cnblogs.com/AndyJee/p/5053354.html

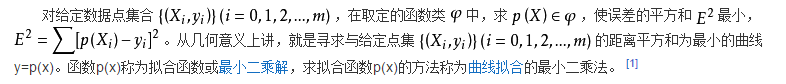
主要内容：

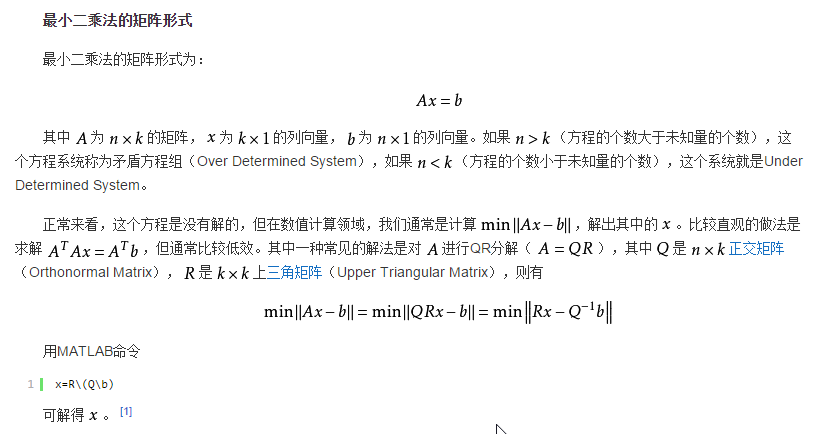
1. 什么是最小二乘
2. 最小二乘的几何意义
3. 正交投影矩阵

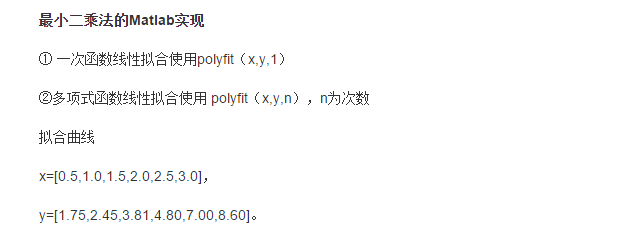
什么是最小二乘？

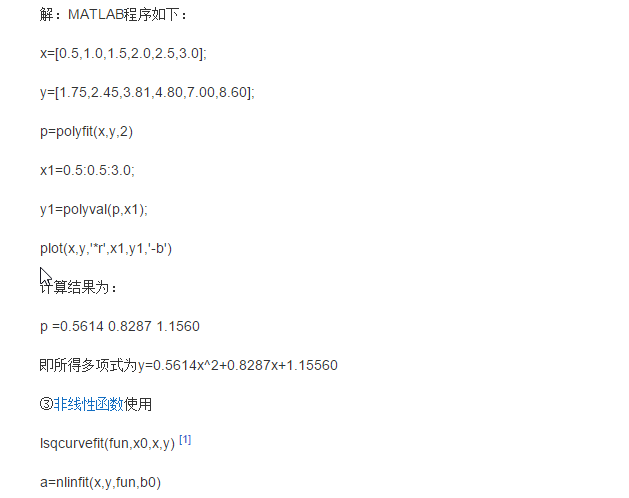
假设我们手上有n组成对的数据，{(xi,yi):i=1…n}，为了探究y变量与x变量的关系，我们希望用一个多项式来匹配它，可是多项式中的系数怎么确定呢？拿来拼凑肯定是不行的，最小二乘法告诉我们，这个多项式的系数应该让每个点的误差的平方之和最小。

**（百度百科）**最小二乘法（又称最小平方法）是一种数学优化技术。它通过最小化误差的平方和寻找数据的最佳函数匹配。利用最小二乘法可以简便地求得未知的数据，并使得这些求得的数据与实际数据之间误差的平方和为最小。最小二乘法还可用于曲线拟合。其他一些优化问题也可通过最小化能量或最大化熵用最小二乘法来表达。





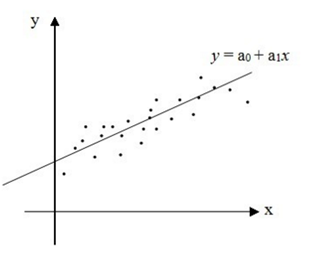




最小二乘的几何意义

**最小二乘的几何意义：最小二乘法中的几何意义是高维空间中的一个向量在低维子空间的投影。**

从上面的定义中，我们很难想象到最小二乘的几何意义，那么我们通过一个简单的例子来推导一下：



我们根据定义中的误差平方之和最小化来拟合直线：

每个点的误差表示：http://images2015.cnblogs.com/blog/585228/201512/585228-20151217104606943-1858233857.png

最小误差的平方和：http://images2015.cnblogs.com/blog/585228/201512/585228-20151217104609052-975805997.png

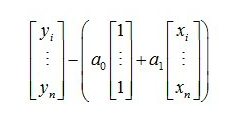
要求解上面的最小化问题，我们可以通过求导的方式得到，最好是转化为矩阵表达形式：AX=b (这里x表示上述的系数a)

求得结果为：

http://images2015.cnblogs.com/blog/585228/201512/585228-20151217104611740-567783292.png

如果通过超定方程的解法http://images2015.cnblogs.com/blog/585228/201512/585228-20151217104613756-1977726673.png，很容易就可以得到上面结果。

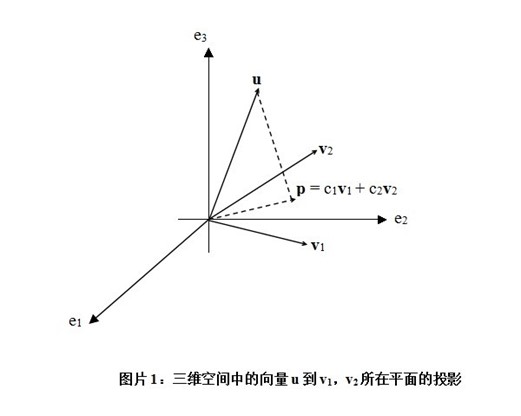
先来说说向量表达形式：



小括号中表示：它是两个向量 [1, ... , 1]T 和 [x1, ... , xn]T 的线性组合，换句话说，它是这两个向量构成的二维子空间（想成一个平面就可以）的任意一点。

那么上面式子的几何含义：表示向量 [y1, ... , yn]T（表示空间中的一点） 到这个二维子空间任意一点的距离；（向量的长度）

最小化上面式子的平方（向量长度的最小化）的几何含义：寻找在 [1, ... , 1]T 和 [x1, ... , xn]T 构成的二维子空间上的一个点，使得向量 [y1, ... , yn]T 到这个点的距离最小。怎么找这个点呢？只要做一个几何投影就好了。（如下图）



如上图所示，在三维空间中给定一个向量 u，以及由向量 v1，v2 构成的一个二维平面，向量 p 为 u 到这个平面的投影，它是 v1，v2 的线性组合:

利用投影的垂直性质，我们可以得到关于系数C的两个方程：

http://images2015.cnblogs.com/blog/585228/201512/585228-20151217104619959-1274494141.png

令 V = [v1, v2], p = c1v1 + c2v2，将上述式子合并并转化为矩阵形式（更容易扩展到高维空间），得到：

http://images2015.cnblogs.com/blog/585228/201512/585228-20151217104621943-1947586720.png http://images2015.cnblogs.com/blog/585228/201512/585228-20151217104624537-1887047884.png

因此系数ｃ的表达式为：

http://images2015.cnblogs.com/blog/585228/201512/585228-20151217104627146-308417244.png

有没有发现很熟悉?和式子 http://images2015.cnblogs.com/blog/585228/201512/585228-20151217104629240-1333895007.png一模一样有木有！！！

好了，我们回到原来的例子，看看几何关系中的投影点和被投影的空间分别代表什么。

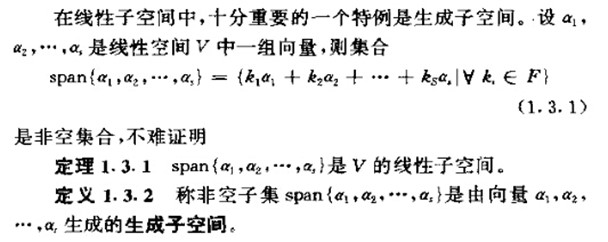
把图中的 u 替换成 [y1, ... , yn]T ，把 v1，v2 分别替换成 [1, ... , 1]T 和 [x1, ... , xn]T, 系数 c1 和 c2 也就是我们要求的 a0，a1。

所以，最小二乘法的几何意义是高维空间的一个向量（由y数据决定）在低维子空间（由x数据以及多项式的次数决定）的投影。

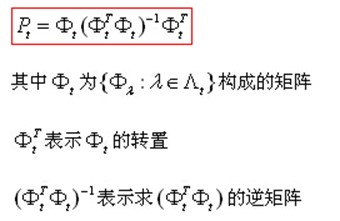
正交投影矩阵

上面提到了最小二乘的几何意义就是空间中的投影，其实投影在线性代数中也是存在其数学公式的，可以联系以下数学知识来理解最小二乘的几何意义。

**张成子空间：**



**张成子空间的投影矩阵：**



**最小二乘的投影解释：**

