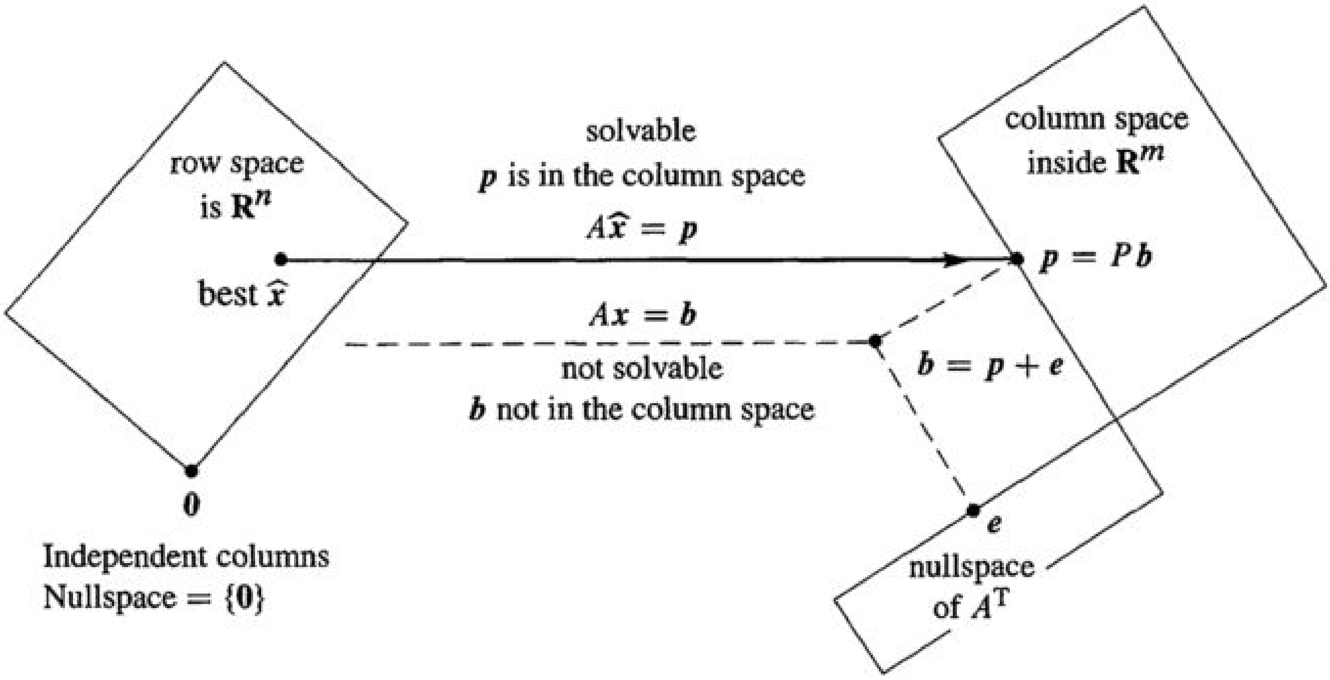

Figure 4.7: The projection $p = A\widehat{x}$ is closest to b, so $\widehat{x}$ minimizes $E = \|b - Ax\|^2$.

信息门下...

Part 1.投影矩阵

先看一个例子， $p_1 = P_1 b = \begin{bmatrix} 0&0&0 \\ 0&0&0 \\ 0&0&1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ z \end{bmatrix}$ ， $p_2 = P_2 b = \begin{bmatrix} 1&0&0 \\ 0&1&0 \\ 0&0&0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$.通过左乘一个矩阵 P ，我们得到了 b 在 z 轴和 xy 平面的投影 p_1 ， p_2 .

> **定义一：**若投影 p ，向量 b ，矩阵 P 满足 $p = Pb$ ，则称 P 为投影矩阵

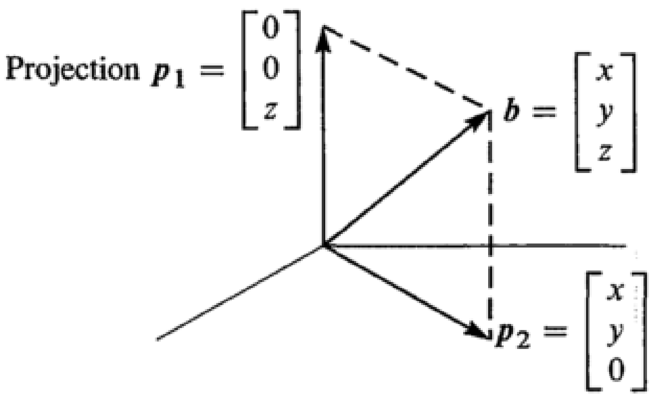

Figure 4.4: The projections $p_1 = P_1 b$ and $p_2 = P_2 b$ onto the z axis and the xy plane.

1.1投影到直线上

要将 一个向量 b 投影到过向量 a 的直线 l 上，关键是作出 b 点（在坐标系中向量与点是一一对应的）到直线 l 的垂线 e ，假设 b 在 l 上的投影为 $p=\hat{x}a$ ，容易知道： $e=b-p$.

> $\because e\perp a \therefore ea=(b-\hat{x}a)a=0$
>
> 故
> $$\hat{x}=\frac{a^Tb}{a^Ta} \tag{1}$$
>
> 有
> $$p=\hat{x}a=\frac{a^Tb}{a^Ta}a=Pb \tag{2}$$
>
> 根据投影矩阵的定义
> $$P=\frac{aa^T}{a^Ta} \tag{3}$$

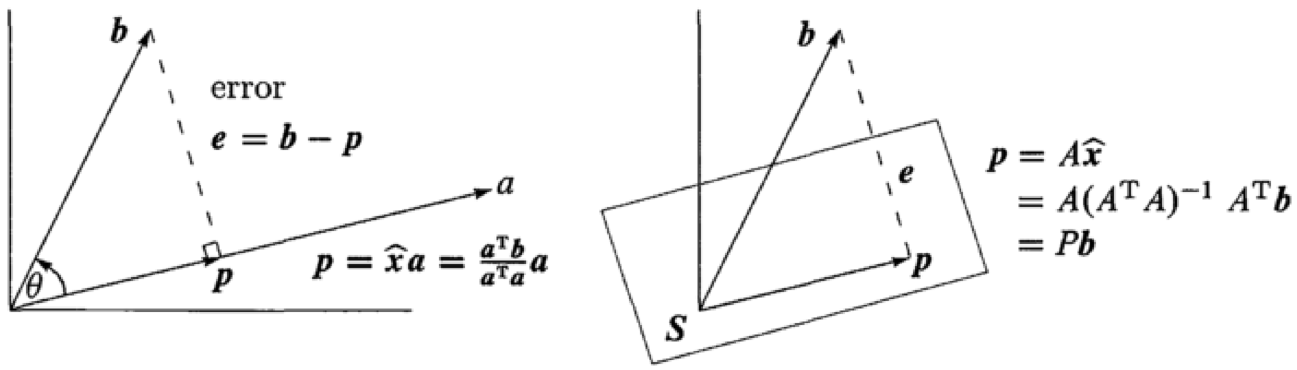

Figure 4.5: The projection p of b onto a line and onto S = column space of A.

1.2投影到子空间上

假设向量 $a_1,\dots,a_n$ 张成 $\mathbb{R}^m$ 中的子空间 $C(A)\subset\mathbb{R}^n$ （ $m>n$),现在我们要将 b 投影到 子空间 S 中.与**1.1**类似,我们先作出向量 b 到子空间 $C(A)$ （ A 的列空间)的垂线 $e=b-p$ ，设投影 $p=\hat{x_1}a_1+\cdots+\hat{x_n}a_n=A\hat{x}(A=[a_1\dots a_n],\hat{x}=\begin{bmatrix}\hat{x_1}\\ \vdots\\ \hat{x_n}\end{bmatrix})$

> $\because e\perp C(A)\therefore A^Te=\begin{bmatrix}a_1^T\\ \vdots\\ a_n^T\end{bmatrix}(b-A\hat{x})=0$
>
> 故 $A^TA\hat{x}=A^Tb$
>
> **wrong result:** $p=A\hat{x}=b$
>
> **correct result:**
> $$\hat{x}=(A^TA)^{-1}A^Tb \tag{4}$$
>
> 因为 $A_{m\times n}$ 行数小于列数，所以 A (其 n 个列向量均线性无关)不是可逆矩阵,但 $(A^TA)_{n\times n}$

> 是可逆的,具体证明在附录[1]
> $$p=A\hat{x}=A(A^TA)^{-1}A^Tb \tag{5}$$
> 所以
> $$P=A(A^TA)^{-1}A^T \tag{6}$$

现在我们阐述关于投影矩阵的一个性质。先想象这样的场景：我们将一束光线 $\vec{x}$ 投影到墙面上得到 $$\overrightarrow{x_{p1}}=P\vec{x} \tag{a1}$$,我们再对 $\overrightarrow{x_{p1}}$ 进行投影来得到 $\overrightarrow{x_{p2}}=P\overrightarrow{x_{p1}}$ 。$\overrightarrow{x_{p2}}$ **和** $\overrightarrow{x_{p1}}$ **其实是同一个向量。 因为本身就在墙面内的向量** $\overrightarrow{x_{p1}}$ **在墙面内的投影就是它自身。所以有：**
$$\overrightarrow{x_{p1}}(=\overrightarrow{x_{p2}}=P\overrightarrow{x_{p1}})=PP\vec{x} \tag{a2}$$

由**(a1)(a2)得：** $$P=P^2 \tag{7}$$

Part2.最小二乘法

2.1最小二乘法的统计意义

2.1.1 回归的统计意义

设随机变量 Y **(因变量)与普通变量** x **（自变量）之间存在相关关系，对于每一个确定的** x **，** Y **有一个对应的分布** $F(y|x)$ **表示当** x **取确定值时对应的** Y **的分布函数。若能掌握每一个** x **取值对应的** $F(y|x)$ **，我们就完全掌握了** Y **与** x **的关系，但这样做会很复杂，所以作为一种近似，我们去考察期望** $E(Y)$ **关于** x **的函数** $E(Y)=\mu(x)$ **.** $\mu(x)$ **称为** Y **关于** x **的回归函数。**

我们用回归函数 $\mu(x)$ **作为** Y **的近似，来表达** Y **与** x **的关系**

为什么我们用 $E(Y)$ (即 $\mu(x)$)作为 Y 的近似呢?

> **Lemma1:对于随机变量** Z **，** $E(Z-c)$ **在** $c=E(Z)$ **时，** $E(Z-c)$ **最小**

根据Lemma1， x **的一切函数中** $\mu(x)=E(Y)$ **能使均方误差** $E(Y-\mu(x))$ **最小。**

2.1.2最小二乘法与线性回归

回归函数 $\mu(x)$ **需要我们通过估计的方法得到。**现在我们先用极大似然估计法来分析一个二次曲线的拟合问题，并借此阐明一些概念。

> **Problem:** 自变量 x 取一组不完全相同的值 $x_1,x_2,...,x_m$ ，设 $Y_1,Y_2,...,Y_m$ 分别是在 $x_1,x_2,...,x_m$ 处对 Y 的**独立**观察结果，$(x_1,Y_1),(x_2,Y_2),...,(x_m,Y_m)$ 是**一个样本**，对应的

> 样本值由一些样本点组成
> $$(x_1,y_1),(x_2,y_2),\dots,(x_m,y_m) \quad \text{(c1)}$$
> **现用一条二次曲线** $y=\mu(x)=C+Dx+Ex^2$ 去拟合这些样本点，使**经验回归函数** $\mu(x)$ 在 $x_i(i=1,2,\dots,m)$ 各处的函数值与各处的观察值 $y_i(i=1,2,\dots,m)$ 的差值平方和
> $$\sum_{i=1}^{m}(y_i-\mu(x_i))^2 \quad \text{(c2)}$$
> 最小，**试求参数 C，D，E 的值。**

这种令样本值和经验回归函数差值平方和最小的限定条件就是最小二乘法的特点，有此特点的回归函数参数估计方法被称为最小二乘法。注意，此时我们并未考虑这个差值所服从的分布，而在下面我们会先假设这个误差服从正态分布，再讨论当差值服从的分布未知的情况。

> **Solution：**
> **a.从统计角度**
>
> 因为m>3,所以根据样本点列出的方程数大于未知量 C，D，E 的个数。这时候如果用任意三个样本点列三个方程解出 C，D，E 的值而得到一个曲线方程，不难猜到其他m-3个样本点很可能不会落在这条曲线上而仅仅是落在曲线附近，于是 Y 与 y 之间就存在一个误差量。**注意：x,x_i,x_i^* 均可由样本值得到，在估计未知参数 C,D,E 的过程中 属于已知的常数或常数向量。 假设这个误差量服从正态分布**，有 $Y=C+Dx+Ex^2+\varepsilon,\varepsilon\sim N(0,\sigma^2)$,相当于
> $$Y\sim N(C+Dx+Ex^2,\sigma^2) \quad (8)$$
> 记 $w=(C,D,E)$ ，$x_i^*=\begin{bmatrix}1\\x_i\\x_i^2\end{bmatrix}$ 则有
> $$Y_i\sim N(w^Tx_i^*,\sigma^2) \quad (9)$$
> 式子**(9)暗含一个条件**：在各个 x 取值 $x_1,x_2,\dots,x_m$ 处对Y 的独立观察结果 $Y_i=C+Dx_i^*+E\cdot(x_i^*)^2+\varepsilon_i$ ，$\varepsilon_i\sim N(0,\sigma^2)$,即**各处的偏差均独立地服从同一**正态分布 $N(0,\sigma^2)$
>
> 于是
> $$p(y_i|w,\sigma,x_i^*)=\frac{1}{\sigma\sqrt{2\pi}}exp[\frac{-(y_i-w^Tx_i^*)^2}{2\sigma^2}] \quad (10)$$
> 则 $Y_1,\dots,Y_m$ 的联合概率密度为
> $$L=\prod_{i=1}^{m}p(y_i|w,\sigma,x_i^*)=(1/\sigma\sqrt{2\pi})^m exp[-\frac{\sum_{i=1}^{m}(y_i-w^Tx_i^*)^2}{2\sigma^2}] \quad (11)$$
> 用极大似然估计法来估计 $w=(C,D,E)$ ，需要让 L 取得极大值，容易知道**式(11)在右边括号中平方和部分最小，即令**
> $$Q(w)=\sum_{i=1}^{m}(y_i-w^Tx_i^*)^2 \quad (12)$$
> **最小。式(12)与式(c2)实际上是同一个式子。接下来，$Q(w)$ 对 w 求导并令导数为0可找出极值点** $\hat{w}=(\hat{C},\hat{D},\hat{E})$ 。我们这里先不求出 $\hat{w}$ 的具体表达式，先观察这个式子 $Q(w)$,**有没有感觉跟距离公式很类似**？设有矩阵 $X=(x_1^*,x_2^*,\dots,x_m^*)^T=\begin{bmatrix}1&x_1&x_1^2\\1&x_2&x_2^2\\&\dots&\\1&x_m&x_m^2\end{bmatrix}$ ，向量 $y=\begin{bmatrix}y_1\\y_2\\\vdots\\y_m\end{bmatrix}$,则

$$Q(w)=(Xw-y)^T(Xw-y) \tag{13}$$

式(13)表明 $Q(w)$ 是向量 Xw 与 y 的欧氏距离。

---

b.下面从线性代数的角度出发，再次分析这个拟合问题

分析题意，其实是要我们用已知的 m 个样本点来解关于 $w=(C,D,E)$ 的线性方程组

$$Xw=y \Leftrightarrow \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ & \dots & \\ 1 & x_m & x_m^2 \end{bmatrix} \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \tag{14}$$

此处的 X,y 与**式(13)中的含义相同。**

X **是一个方阵,**易知 $Xw=y$ **无解[2]，即** $y \notin C(X)$,还记得**投影吗？虽然向量** y **不在** $C(X)$ **中,但我们可以把** y **投影到** $C(X)$ **中。然后我们用通过投影得到** p **来近似替代** y ，**并解方程**

$$X\hat{w}=p \tag{15}$$

之所以用投影 p 来近似 y 是因为这样能使误差向量的模 $||e||=||y-p||=||y-X\hat{w}||$ 最小，意味着误差最小，这样符合我们的直觉。

现在，问题变成如何找出 y 在 $C(X)$ 中的投影 p ，**即如何将 y 投影到 $C(X)$ 中**

参考1.2可知投影矩阵 $P=X(X^TX)^{-1}X^T$ 能将 y 投影到 $C(X)$ 中，于是

$$p=X(X^TX)^{-1}X^Ty \tag{16}$$

$$\hat{w}=(X^TX)^{-1}X^Ty$$

由(15)(16) 有 $X\hat{w}=X(X^TX)^{-1}X^Ty$,两边同时左乘 X^T 可得：

$$X^TX\hat{w}=X^Ty \tag{17}$$

当方程 $A_{m\times n}x=b=p+e(m>n)$ 无解时，我们转而求解方程 $A^TA\hat{x}=A^Tb$ ，并以解得得 $\hat{x}$ 作为 x 的估计值，这种做法就是最小二乘估计，它能在 $C(A)$ 中找到离点 b 最近的点 p ，并以 p 近似 b ，使误差向量 $e=b-p$ 模最小。

实际上最小二乘的思想也可以用下面这张图表达： p 来自 $C(A)$,而 e 来自与 $C(A)$ 垂直的 $N(A^T)$

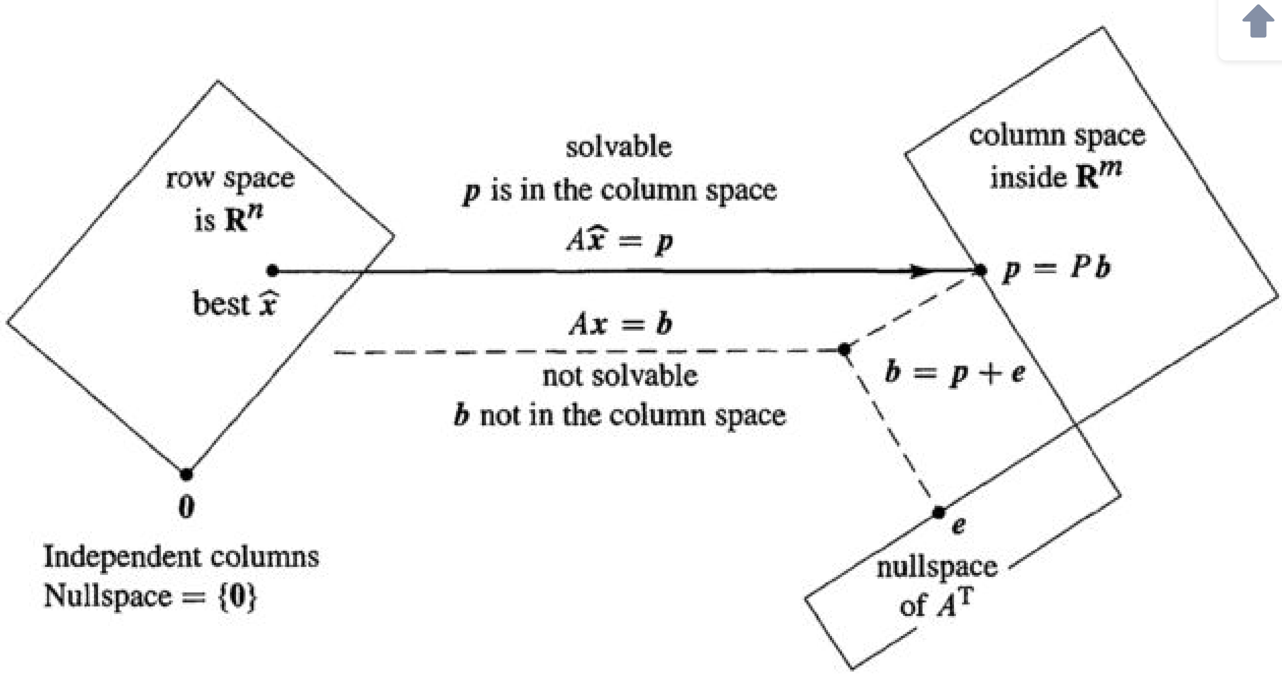

Figure 4.7: The projection $p = A\hat{x}$ is closest to b, so $\hat{x}$ minimizes $E = \|b - Ax\|^2$.

附录

[1]

> **Theorem:如果 A 的所有列均线性无关，则 A^TA 必为可逆矩阵 Proof: Lemma1:**对方阵 F 而言， $Fx=0$ 只有零解(即 $N(F)=\{\vec{0}\}$)时 F 可逆。
> **Proof:** 略
>
> 现在我们想办法证明 $N(A^TA)=\{\vec{0}\}$ ，实际上我们可以先证明一个更广泛的结论
> **Lemma2: 对任意矩阵 A，有 $N(A^TA)=N(A)$ Proof:**
> 先证明 $N(A)\subseteq N(A^TA)$：
> $\forall x\in N(A), Ax=0$,两边左乘 A^T 得: $A^TAx=0$,所以
> $$N(A)\subseteq N(A^TA) \qquad \text{(b1)}$$
> 再证明 $N(A^TA)\subseteq N(A)$ ：
> $\forall x\in N(A^TA), A^TAx=0$,两边左乘 x^T 得:
> $x^TA^TAx=(Ax)^TAx=||Ax||^2=0$,即 $Ax=0$ ，所以
> $$N(A^TA)\subseteq N(A) \qquad \text{(b2)}$$ 由**(b1)(b2)得:**
> $$N(A^TA)=N(A) \qquad (8)$$

> 回到Theorem，**如果 A 的所有列均线性无关，则** $N(A)=\{\vec{0}\}$,根据**(8)可得：** $N(A^TA)=N(A)=\{\vec{0}\}$,根据**lemma1知：** A^TA **可逆**

[2]

> **For a linear system equation:**
>
> $$A_{m\times n}x=b$$
> $$rank(A)=r$$
> $$r<m,r=n$$
>
> $r<m$ means the actual column space of A is not m dimensional but has r dimensions.
> $r=n$ means the matrix A actually transform a n dimensional space to another n dimensional one(though seemingly a m dimensional space). ------------**Theory 1**
> Since A seemingly has m dimensions, b with m components is possible to be not in the actual n dimensional space. I mean **if the rest $m-r$ components of b are not all 0, b will be in a higher dimension space(m dimensions)**. In this case, there is no solutions for $Ax=b$.
> **In contrast**, **when b has only the upper r none-zero components**, b is in the actual column space (n dimensions) of A . On the basis of **Theory 1** we can see a situation similar to the **first case(** $r=n=m$) thus only one solution to $Ax=b$.

13 条评论 ⇌ 切换为时间排序

写下你的评论...

杨超越 2018-09-27

全知乎搜投影矩阵, 就你的回答最靠谱. 其他都是什么shit玩意儿.

2

信息门下勃狗 (作者) 回复 杨超越 2018-09-27

谢谢滋磁(*•̀ᴗ•́*)ʃ

赞

再见孙悟空 2020-12-02

感谢一下

赞

惊彦 2020-06-

被高代搞到崩溃，终于找到一个靠谱的讲解

赞

掀霖 2020-05-29

太感谢了，上课时老师莫名其妙说出投影阵的概念感觉云里雾里的，都是没接触过的，来恶补一下

赞

信息门下勃狗 (作者) 回复 掀霖 2020-05-30

能有点帮助就好

赞

simpcode 2020-01-01

什么时候能像您这么厉害

赞

晨曦 2019-04-08

投影到子空间的时候，不是假设m>n吗？怎么后面说，行数小于列数啊？

赞

信息门下勃狗 (作者) 回复 晨曦 2019-04-08

笔误了，其实意思就是非方阵没有逆

赞

请叫我张先森 2018-12-25

赞

MagicA 2018-11-07

写得很棒

赞

信息门下勃狗 (作者) 回复 MagicA 2018-11-07

谢谢滋磁(*•̀ᴗ•́*)ʃ

赞