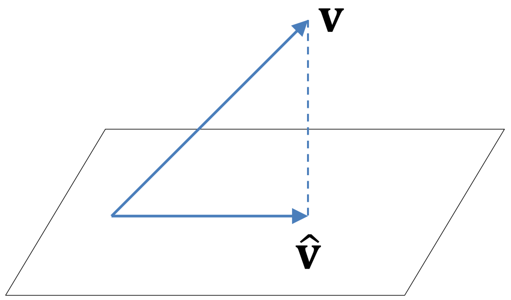
第八讲 正交投影

2023.11.10

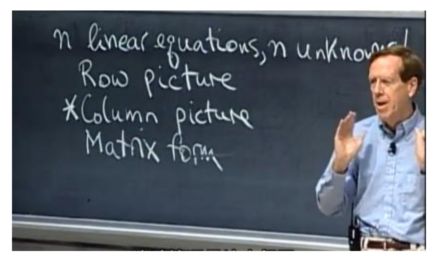


我们首先简单回顾一般向量空间和内积向量空间的投影（可忽略），主要关注下述两种重要特殊情形，尤其是欧氏空间的投影:

- 去相关化是随机变量的投影，条件期望是投影；
- 回归模型的最小二乘法是欧氏空间向量投影；

线性代数教材

1. S. Axler, Linear algebra done right, 3ed. （进阶）
2. G.Strang, Introduction to linear algebra,5ed. （入门）
3. G.Strang, Linear algebra and its applications 4ed. （应用）
4. G.Strang, Linear algebra and learning from data. （深度学习）



主要内容

- 向量空间
  - 函数空间: 欧氏R^n, 实函数L^2, 随机变量$L^2(\mathcal{F})$
- 内积向量空间的投影
- 欧氏空间：矩阵预备知识
  - 四个基本空间：列空间、行空间、核空间
  - 奇异值分解（SVD）、广义逆 A^{-}
- 欧氏空间的投影

下次课

向量空间

法国哲学家、数学家勒内•笛卡尔(Rene Descartes)于1637年创立了笛卡尔坐标系，实现了几何问题代数化，为微积分的建立奠定了基础。R^2和R^3笛卡尔坐标向量的代数运算可进一步拓展到R^n欧氏空间，进而拓展到抽象的向量空间的代数运算，这些拓展极大地扩展了数学研究范围。

向量空间/线性空间

向量空间V是具有线性结构的集合，即集合元素定义了加法和数乘运算、包含加法单位元(零)且满足通常的运算规则（交换律、结合律和分配律）。向量空间中元素称为向量，可以是映射、函数或任何其它数学对象。最常见的向量空间是函数空间（下页）。

线性变换

定义：V, W 是两个向量空间 (比如$V = R^m, W = R^n$).

映射$T: V \to W$称为是线性变换或线性映射，如果对任何 $\mathbf{v}_1, \mathbf{v}_2 \in V$ 和任何实数λ_1, λ_2有

$$T(\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2) = \lambda_1 T(\mathbf{v}_1) + \lambda_2 T(\mathbf{v}_2).$$

所有的$V \to W$线性映射的集合记为$L(V, W)$.

内积向量空间

定义向量空间V中的内积$\langle\cdot,\ \cdot\rangle$满足

- 正性：$\langle\mathbf{v},\mathbf{v}\rangle \geq 0,\ \forall\mathbf{v}\in V$ 且$\langle\mathbf{v},\mathbf{v}\rangle = 0 \Leftrightarrow \mathbf{v} = 0.$
- 对称性：$\langle\mathbf{u},\mathbf{v}\rangle = \langle\mathbf{v},\mathbf{u}\rangle.$
- 双线性：$\langle\lambda_1\mathbf{u} + \lambda_2\mathbf{v},\ \mathbf{w}\rangle = \lambda_1\langle\mathbf{u},\ \mathbf{w}\rangle + \lambda_2\langle\mathbf{v},\ \mathbf{w}\rangle.$

定义了内积的向量空间称为内积向量空间。

定理1.（毕达哥拉斯定理）若$\langle\mathbf{u},\mathbf{v}\rangle=0$，正交，记作$\mathbf{u}\perp\mathbf{v}$。
此时，$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$

定义模长$\|\mathbf{v}\| = \langle\mathbf{v},\mathbf{v}\rangle^{1/2}$，距离$d(\mathbf{u},\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\|$，进而可以研究几何性质和极限、连续等分析性质。

希尔伯特空间

完备的内积向量空间称为希尔伯特空间。完备：对极限封闭。
后续常用的空间都是希尔伯特空间，不再一一提及或验证。

（内积）函数空间R^S

集合S 上所有实函数组成的集合：$R^S = \{f: S \to R\}$，定义R^S上的两个函数的加法以及实数与函数的乘法如下 ($\forall f, g \in R^S$， $s \in S$ ， $\lambda \in R$)

- 加法：定义$f + g$为一个函数：$(f+g)(s) = f(s) + g(s)$,
- 数乘：定义乘积λf为一个函数：$(\lambda f)(s) = \lambda f(s)$

所有运算法则都满足，R^S是一个向量空间，是最重要的向量空间。

当$S = R$、$\{1, \dots, n\}$、 Ω时， R^S分别是实函数空间、有限维空间R^n和随机变量空间， 分别是泛函、线性代数、概率论的基本框架。

实函数空间R^R

R上的所有实函数构成的函数空间 $R^R = \{f: R \to R\}$ 。

定义内积$\langle f, g \rangle = \int f(x) g(x) dx$，$\|f\| = (\int f(x)^2 dx)^{1/2}$

所有平方可积函数组成空间 $L^2 = \{f: \int f(x)^2 dx < \infty\} \subset R^R$

泛函分析：代数+分析，研究内积(或距离、拓扑)向量空间的泛函（函数的函数）的分析性质，研究函数空间之间的线性变换（如卷积）。

有限集合$S = \{1, \ldots, n\}$上所有实函数集合

$$R^{\{1,\ldots,n\}} = \{f: \{1, \ldots, n\} \to R\}$$

任一$f \in R^{\{1,\ldots,n\}}$以其所有取值表示:

$$(f(i), i = 1, \ldots, n) \text{ 或 } (f_i, i = 1, \ldots, n)\text{或}(f_1, \ldots, f_n)^\top$$

实向量空间R^n

$R^n = \{\mathbf{x} = (x_1, \ldots, x_n)^\top\colon x_i \in R\}$实际上是函数空间$R^{\{1,2,\ldots,n\}}$，习惯上以黑体小写字母表示任一 $\mathbf{x} \in R^n$,有多种写法

$$\mathbf{x} = (x_1, \ldots, x_n)^\top = (x_i, i = 1, \ldots, n) = (x_i)$$

对$\mathbf{x} = (x_1, \ldots, x_n)^\top$，$\mathbf{y} = (y_1, \ldots, y_n)^\top$，“加法”和数乘

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, \ldots, x_n + y_n)^\top, \mathbf{x}\lambda = (\lambda x_1, \ldots, \lambda x_n)^\top$$

任何n维向量空间存在一组正交基(Schmidt正交化)，且等价于/同构于R^n。

欧氏空间：R^n中定义内积$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{y} = \sum x_i y_i$

随机变量空间变R^{Ω}

概率空间$(\Omega, \mathcal{F}, P)$。样本空间Ω上的实函数集合$R^{\Omega} = \{f: \Omega \to R\}$，对任何$x, y \in R^{\Omega}$，定义两个函数的“和”以及数乘为新的函数：

$$(x+y)(\omega) = x(\omega) + y(\omega),\quad (\lambda x)(\omega)=\lambda x(\omega)\ , \omega \in \ \Omega$$

为了能够分析（比如积分），只考虑R^{Ω}中$\{\omega: x(\omega) < t\} \in \mathcal{F}$ (即可计算概率的) 那些函数，称为随机变量。

定义内积$\langle x, y\rangle = E(xy)$，模长 $\|x\| = \sqrt{Ex^2}$

所有二阶矩存在的随机变量组成内积向量空间

$$L^2(\mathcal{F}) = \{x: x\text{关于}\mathcal{F}\text{可测}, Ex^2 < \infty\} \subset R^{\Omega}.$$

随机变量$x(\omega)$通常省略ω, 简写为x，虽然称为随机“变量”，但 x 实际上是函数，是向量空间中的向量。

伴随
adjoint

希尔伯特空间之间连续线性变换（有限维内积向量空间的任何线性变换）$T\in L(V,W)$有唯一的伴随(共轭，adjoint)变换$T^*\in L(V,W)$，使得

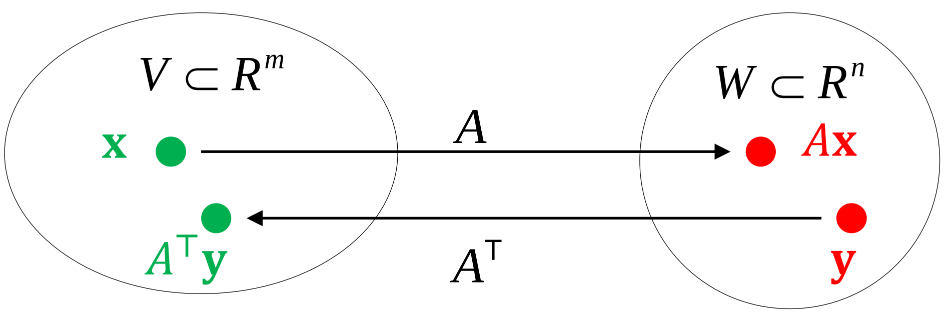
$$(T\mathbf{v},\mathbf{w})=(\mathbf{v},T^*\mathbf{w}),\quad \forall\mathbf{v}\in V,\ \mathbf{w}\in W$$

若$T=T^*$, 自伴随。

对于欧氏空间， 矩阵A的伴随矩阵是其**转置**矩阵$A^\top$，满足

$$(A\mathbf{x},\mathbf{y})=(\mathbf{x},A^\top\mathbf{y}),\ 即(A\mathbf{x})^\top\mathbf{y}=\mathbf{x}^\top(A^\top\mathbf{y}),$$

注意两端的内积是在不同空间计算的，如下图所示



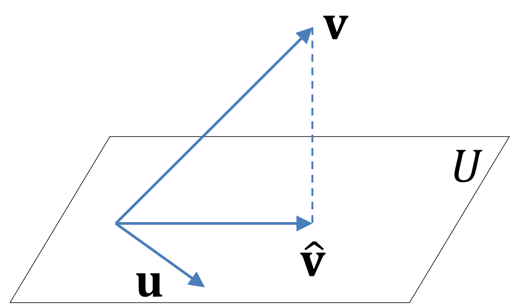
矩阵自伴随/对称：$A^\top=A$

内积向量空间的正交投影

定义：假设V是内积向量空间，$U \subset V$是内积向量子空间，则任何$\mathbf{v} \in V$ 在U上的正交投影$\hat{\mathbf{v}} \in U$ ，满足 $\mathbf{v} - \hat{\mathbf{v}} \perp U$, 即

$$\forall \ \mathbf{u} \in U, \ \langle \mathbf{v} - \hat{\mathbf{v}}, \mathbf{u} \rangle = 0 \Leftrightarrow \langle \mathbf{v}, \mathbf{u} \rangle = \langle \hat{\mathbf{v}}, \mathbf{u} \rangle$$

容易验证投影是一个线性变换/算子，记$\hat{\mathbf{v}} = P_U \mathbf{v}$.



正交投影是极小化误差平方和问题的最优解

定理2. (最小二乘) $P_U \mathbf{v} = \underset{\mathbf{u} \in U}{\operatorname{argmin}} \|\mathbf{u} - \mathbf{v}\|^2$

证明：$\mathbf{u} \in U, \hat{\mathbf{v}} = P_U \mathbf{v} \in U \Rightarrow \mathbf{u} - \hat{\mathbf{v}} \in U$,故$\mathbf{v} - \hat{\mathbf{v}} \perp \mathbf{u} - \hat{\mathbf{v}}$,

由毕达哥拉斯定理

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u} - \hat{\mathbf{v}} + \hat{\mathbf{v}} - \mathbf{v}\|^2 = \|\mathbf{u} - \hat{\mathbf{v}}\|^2 + \|\hat{\mathbf{v}} - \mathbf{v}\|^2 \geq \|\hat{\mathbf{v}} - \mathbf{v}\|^2$$

命题1. 假设V是内积向量空间，$\mathbf{u} \in V$，$U = \{\lambda\mathbf{u}\colon \lambda \in R\} \subset V$，则任何$\mathbf{v} \in V$在$U$上的正交投影 $P_{\mathbf{u}}\mathbf{v} = \hat{\mathbf{v}} = \frac{\langle \mathbf{v},\mathbf{u}\rangle}{\langle \mathbf{u},\mathbf{u}\rangle}\mathbf{u}$，

证：$\hat{\mathbf{v}} = \lambda\mathbf{u}$, $0 = \langle \mathbf{v} - \lambda\mathbf{u}, \mathbf{u}\rangle = 0 \Rightarrow \langle \mathbf{v}, \mathbf{u}\rangle = \lambda\langle \mathbf{u}, \mathbf{u}\rangle, \lambda = \frac{\langle \mathbf{v},\mathbf{u}\rangle}{\langle \mathbf{u},\mathbf{u}\rangle}$

命题2.（Cauchy-Schwarz）假设V内积向量空间，对任何$\mathbf{u}, \mathbf{v} \in V$，

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \le \|\mathbf{u}\|\|\mathbf{v}\|$$

证：$\hat{\mathbf{v}} = P_{\mathbf{u}}\mathbf{v} = \frac{\langle \mathbf{v},\mathbf{u}\rangle}{\langle \mathbf{u},\mathbf{u}\rangle}\mathbf{u}$, $\mathbf{v}^{\perp} = \mathbf{v} - \hat{\mathbf{v}} \perp \hat{\mathbf{v}}$,

$$\|\mathbf{v}\|^2 = \|\hat{\mathbf{v}}\|^2 + \|\mathbf{v}^{\perp}\|^2 \ge \|\hat{\mathbf{v}}\|^2 = |\langle \mathbf{u}, \mathbf{v} \rangle|^2/\|\mathbf{u}\|^2$$

命题3. 假设V是内积向量空间，假设 $\mathbf{u}_1, \dots, \mathbf{u}_m \in V$ 相互正交, 它们张成的线性子空间

$$U = \{\lambda_1\mathbf{u}_1 + \cdots + \lambda_m\mathbf{u}_m\colon \lambda_1, \dots, \lambda_m \in R\} \subset V,$$

则任何$\mathbf{v} \in V$在U上的正交投影 $P_U\mathbf{v} = \sum_{i=1}^{m} P_{\mathbf{u}_i}\mathbf{v} = \sum_{i=1}^{m} \frac{\langle \mathbf{v},\mathbf{u}_i\rangle}{\langle \mathbf{u}_i,\mathbf{u}_i\rangle}\mathbf{u}_i$,

由命题1、3，得

命题4 (Gram-Schmidt正交化) 对内积向量空间中任何一组线性无关向量 $\mathbf{u}_1, \dots, \mathbf{u}_m$ ，可将它们正交化：

$$\mathbf{e}_1 = \mathbf{u}_1$$

$$\mathbf{e}_2 = \mathbf{u}_2 - P_{\mathbf{u}_1}\mathbf{u}_2 = \mathbf{u}_2 - \frac{\langle \mathbf{u}_2, \mathbf{e}_1 \rangle}{\langle \mathbf{e}_1, \mathbf{e}_1 \rangle}\mathbf{e}_1$$

$$\mathbf{e}_3 = \mathbf{u}_3 - P_{\mathbf{u}_1,\mathbf{u}_2}\mathbf{u}_3 = \mathbf{u}_3 - P_{\mathbf{e}_1,\mathbf{e}_2}\mathbf{u}_3 = \mathbf{u}_3 - P_{\mathbf{e}_1}\mathbf{u}_3 - P_{\mathbf{e}_2}\mathbf{u}_3$$

$$= \mathbf{u}_3 - \frac{\langle \mathbf{u}_3, \mathbf{e}_1 \rangle}{\langle \mathbf{e}_1, \mathbf{e}_1 \rangle}\mathbf{e}_1 - - \frac{\langle \mathbf{u}_3, \mathbf{e}_2 \rangle}{\langle \mathbf{e}_2, \mathbf{e}_2 \rangle}\mathbf{e}_2$$

…

由命题3、4，得

命题5. 若U是内积向量空间V的有限维子空间，则任何$\mathbf{v} \in V$在U上存在正交投影$P_U\mathbf{v}$.

一般地，若V是希尔伯特内积空间，U是闭线性子空间，投影存在（略）。

条件期望

定义（条件期望）假设概率空间$(\Omega, \mathcal{F}, P)$，假设子σ域 $\mathcal{G} \subset \mathcal{F}$

考虑二阶矩存在的随机变量空间

$$V = L^2(\mathcal{F}) = \{y : y\text{关于}\mathcal{F}\text{可测}, Ey^2 < \infty\},$$

及其子空间（限制r.v.复杂性）

$$U = L^2(\mathcal{G}) = \{u : u\text{关于}\mathcal{G}\text{可测}, Eu^2 < \infty\},$$

随机变量$y \in V$在U上的投影$P_U y = E(y|\mathcal{G})$满足

$$\langle y - E(y|\mathcal{G}), u\rangle = E(y - E(y|\mathcal{G}))u = 0, \forall\ u \in U$$

$P_U y = E(y|\mathcal{G})$称为条件期望。

假设随机向量 $\mathbf{x}$ 的所有函数

$$U = \{f(\mathbf{x}) :\ f\text{可测}, Ef(x)^2 < \infty\}$$

即$U = L^2\ (\sigma(\mathbf{x})) = \ \{u : u\text{关于}\sigma(\mathbf{x})\text{可测}, Eu^2 < \infty\}$，此时条件期望 $P_U y = E(y|\mathbf{x})$是y在所有$\mathbf{x}$的可测函数构成的空间上的投影，是下述最小二乘问题的最优解

$$\min_f \|y - f(\mathbf{x})\|^2 = \min_f E(y - f(\mathbf{x}))^2$$

例1. $V = R^n$. $\mathbf{u}$，$\mathbf{v} \in R^n$， $P_{\mathbf{u}}\mathbf{v} = \frac{\langle \mathbf{v},\mathbf{u} \rangle}{\langle \mathbf{u},\mathbf{u} \rangle}\mathbf{u} = \mathbf{u}\frac{\mathbf{u}^\top \mathbf{v}}{\mathbf{u}^\top \mathbf{u}} = \mathbf{u}(\mathbf{u}^\top \mathbf{u})^{-1}\mathbf{u}^\top \mathbf{v}.$

欧氏空间的投影将在后面详细讨论。

例2. $V = L^2([0,2\pi])$. $1, \sin(x), \cos(x)$张成的自空间

$$U = \{f_{abc}(x) = a + b\sin(x) + c\cos(x)\text{：} a, b, c \in R\} \subset L^2,$$

对任何$f \in V$，$P_U f = P_1 f + P_{\sin} f + P_{\cos} f$, 其中 $P_1 f = \frac{\langle f,1 \rangle}{\langle 1,1 \rangle} = \frac{1}{2\pi}\int_0^{2\pi} f(x)\,dx$，

$P_{\sin} f = \frac{\langle f,\sin \rangle}{\langle \sin,\sin \rangle}\sin(x) = \left(\frac{1}{\pi}\int_0^{2\pi} f(x)\sin(x)dx\right)\sin(x)$， $P_{\cos} f$类似。

$$P_U f = argmin \int_0^{2\pi} [f(x) - f_{abc}(x)]^2\,dx$$

例3. $V = L^2(\mathcal{F})$. 对任何r.v. x和常数1 $\in L^2(\mathcal{F})$,

$$P_1 x = \frac{\langle x,1 \rangle}{\langle 1,1 \rangle}1 = E(x),$$

令$x^\perp = x - P_1 x = x - E(x)$，则

$$\mathrm{var}(x) = E(x - E(x))^2 = E(x^\perp)^2 = \|x^\perp\|^2$$

这是为什么我们常说方差是随机变量的长度平方的原因。

例4. $V = L^2(\mathcal{F})$. 随机变量 $x_1, \dots, x_p, 1 \in V = L^2(\mathcal{F})$ 张成的子空间

$$U = \{a + \mathbf{b}^\top \mathbf{x}:\ a \in R, \mathbf{b} \in R^p\} \subset L^2(\mathcal{F}),$$

其中随机向量 $\mathbf{x} = (x_1, \dots, x_p)^\top$。

假设r.v. y在U上的投影 $P_U y = a + \mathbf{b}^\top \mathbf{x}$, 则$y - P_U y = y - a - \mathbf{b}^\top \mathbf{x}$与$x_i, i = 1, \dots, p$和1正交：

$$\langle y - a - \mathbf{b}^\top \mathbf{x}, 1\rangle = E(y - a - \mathbf{b}^\top \mathbf{x}) = 0$$

$$\langle y - a - \mathbf{b}^\top \mathbf{x}, x_i\rangle = E x_i(y - a - \mathbf{b}^\top \mathbf{x}) = 0, i = 1, \dots, p$$

$$\Leftrightarrow E\mathbf{x}(y - a - \mathbf{b}^\top \mathbf{x}) = 0$$

解得$\mathbf{b}^\top = \Sigma_{y\mathbf{x}} \Sigma_{\mathbf{xx}}^{-1}$，$a = Ey - \mathbf{b}^\top E\mathbf{x}$，因此投影

$$P_U y = Ey + \Sigma_{y\mathbf{x}} \Sigma_{\mathbf{xx}}^{-1}(\mathbf{x} - E\mathbf{x})$$

投影$P_U y$是下述最小二乘问题的最优解

$$\min_{a,\mathbf{b}} \|y - a - \mathbf{b}^\top \mathbf{x}\|^2 = \min_{a,\mathbf{b}} E(y - a - \mathbf{b}^\top \mathbf{x})^2$$

注意这是y在 $\mathbf{x}$,1线性组合子空间上的投影，不是条件期望。条件期望$E(y|\mathbf{x})$是y在 $\mathbf{x}$的所有函数构成的子空间上的投影。

注：例4说明： y关于$\mathbf{x}$去相关化 $\Leftrightarrow$ y向$\mathbf{x}$正交补空间投影。

令

$$\varepsilon \triangleq y - P_U y = y - \Sigma_{y\mathbf{x}}\Sigma_{\mathbf{x}}^{-1}\mathbf{x} - (Ey - \mathbf{b}^\top E\mathbf{x}) \qquad (*)$$

ε与x不相关，改写（*）即线性模型

$$y = (Ey - \mathbf{b}^\top E\mathbf{x}) + \Sigma_{y\mathbf{x}}\Sigma_{\mathbf{x}}^{-1}\mathbf{x} + \varepsilon = a + \mathbf{b}^\top \mathbf{x} + \varepsilon$$

ε中略去常数项即去相关化：

$$y^\perp \triangleq y - \Sigma_{y\mathbf{x}}\Sigma_{\mathbf{x}\mathbf{x}}^{-1}\mathbf{x} \quad \text{与}\mathbf{x}\text{不相关}$$

投影的性质

命题6. 若U是内积向量空间V的子空间， U对应的投影算子(如果存在) 记作 $P = P_U$.

1) 若P存在，必唯一.
2) 对任何$\mathbf{u} \in U, P\mathbf{u} = \mathbf{u}$.
3) P 幂等，即 $P^2 = P$.
4) P自伴随, 即对任何$\mathbf{v}, \mathbf{w} \in V, \langle P\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{v}, P\mathbf{w}\rangle$
5) 若算子$T \neq 0 \in L(V,V)$是幂等且自伴随, 则T是投影算子。

证： 1) 假设$\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2$都是$\mathbf{v}$在$U$上的投影，则对任何 $\mathbf{u} \in U$

$$\langle \mathbf{v} - \hat{\mathbf{v}}_1, \mathbf{u}\rangle = 0，\langle \mathbf{v} - \hat{\mathbf{v}}_2, \mathbf{u}\rangle = 0，$$

所以 $\langle \hat{\mathbf{v}}_1 - \hat{\mathbf{v}}_2, \mathbf{u}\rangle = 0$，取$\mathbf{u} = \hat{\mathbf{v}}_1 - \hat{\mathbf{v}}_2 \Rightarrow \|\hat{\mathbf{v}}_1 - \hat{\mathbf{v}}_2\| = 0$，$\hat{\mathbf{v}}_1 = \hat{\mathbf{v}}_2$。

3) 对任何$\mathbf{v} \in V$, 记$\mathbf{u} = P\mathbf{v} \in U$，则由2), $\mathbf{u} = P\mathbf{u}$, 所以
$P\mathbf{v} = \mathbf{u} = P\mathbf{u} = P^2\mathbf{v}$，$P^2 = P$ 。

4)对任何$\mathbf{v}, \mathbf{w} \in V$, $P\mathbf{w} \in U \Rightarrow \langle \mathbf{v} - P\mathbf{v}, P\mathbf{w}\rangle = \mathbf{0}$,即 $\langle \mathbf{v}, P\mathbf{w}\rangle = \langle P\mathbf{v}, P\mathbf{w}\rangle$,
同理，因为$P\mathbf{v} \in U$，所以$\langle P\mathbf{v}, \mathbf{w} - P\mathbf{w}\rangle = \mathbf{0}$,即 $\langle P\mathbf{v}, P\mathbf{w}\rangle = \langle P\mathbf{v}, \mathbf{w}\rangle$,
所以$\langle \mathbf{v}, P\mathbf{w}\rangle = \langle P\mathbf{v}, \mathbf{w}\rangle$。

5) 记$U = TV$,任取$\mathbf{u} \in U$，假设$\mathbf{u} = T\mathbf{w} = T^2\mathbf{w} = T(T\mathbf{w}) = T\mathbf{u}$,

任取$\mathbf{v} \in V$， $\langle T\mathbf{v}, \mathbf{u}\rangle \overset{\text{自伴随}}{=} \langle \mathbf{v}, T\mathbf{u}\rangle = \langle \mathbf{v}, \mathbf{u}\rangle$, 所以 $T\mathbf{v}$是$\mathbf{v}$在U上的投影。

欧氏空间：矩阵预备知识

矩阵$A_{n\times m}=(a_{ij})$,

矩阵通常列向量表示：$A=\begin{pmatrix} | & & | \\ \mathbf{a}_1, & \cdots & \mathbf{a}_m \\ | & & | \end{pmatrix}$, 第$j$列为向量$\mathbf{a}_j$.

也以行向量表示：$A=\begin{pmatrix} -\boldsymbol{\alpha}_1^{\mathsf{T}}- \\ \vdots \\ -\boldsymbol{\alpha}_n^{\mathsf{T}}- \end{pmatrix}$, 第$i$行为行向量$\boldsymbol{\alpha}_i^{\mathsf{T}}$

列空间与行空间，A与A^{T}（伴随、共轭、对偶）

矩阵$A_{n\times m}=(\mathbf{a}_1,\ldots,\mathbf{a}_m)$的列向量张成的线性空间$C(A)$（或$L(A)$，$\mathrm{Img}(A)$）：

$$C(A)=C(\mathbf{a}_1,\ldots,\mathbf{a}_m)=\left\{\sum_{i=1}^{m}\mathbf{a}_i x_i \mid x_1,\ldots,x_m\in R\right\}=\left\{A\mathbf{x}\mid \mathbf{x}\in R^m\right\}.$$

$A_{n\times m}$的行空间为A的所有行向量张成的空间，即A^{T}的列空间：

$$C(A^{\mathsf{T}})=\left\{A^{\mathsf{T}}\mathbf{y}\mid \mathbf{y}\in R^n\right\}$$

定义：列秩$=\dim C(A)=\mathrm{rank}(A)$；行秩$=\dim C(A^{\mathsf{T}})=\mathrm{rank}(A^{\mathsf{T}})$

定理3：对任一实数矩阵A，$\text{rank}(A^{\mathsf{T}})=\text{rank}(A)$，称为矩阵$A$的秩.

引理1：（1）若$\mathbf{x}_1,\ldots,\mathbf{x}_s$为行空间$C(A^{\mathsf{T}})$的一组线性无关向量，则$A\mathbf{x}_1,\ldots,A\mathbf{x}_s$是列空间$C(A)$中的一组线性无关向量。（2）反之，列空间的一组线性无关向量组的A^{T}变换，在行空间也是线性无关的。

引理1的证明：

假设$s=\text{rank}(A^{\mathsf{T}}),\mathbf{x}_1,\ldots,\mathbf{x}_s\in C(A^{\mathsf{T}})$是一个线性无关组，下面证明$A\mathbf{x}_1,\ldots,A\mathbf{x}_s\in C(A)$也一定是线性无关的。假设存在实数$c_1,\ldots,c_s$，使得

$$\mathbf{0}=A\mathbf{x}_1c_1+\ldots+A\mathbf{x}_sc_s\triangleq A\mathbf{v},\ \text{其中}\ \mathbf{v}\triangleq\mathbf{x}_1c_1+\ldots+\mathbf{x}_sc_s\text{。}$$

$A\mathbf{v}=\mathbf{0}$表明$\mathbf{v}$与A的各行正交，即$\mathbf{v}\perp C(A^{\mathsf{T}})$.

但$\mathbf{v}\in C(A^{\mathsf{T}})\Rightarrow\mathbf{v}\perp\mathbf{v}\Rightarrow\mathbf{v}=\mathbf{x}_1c_1+\ldots+\mathbf{x}_sc_s=\mathbf{0}$,

因为$\mathbf{x}_1,\ldots,\mathbf{x}_s$线性无关，所以必有$c_1=\ldots=c_s=0$，所以$A\mathbf{x}_1,\ldots,A\mathbf{x}_s$线性无关。

这说明了$\text{rank}(A)\geq s=\text{rank}(A^{\mathsf{T}})$.

反之，若$\mathbf{y}_1,\ldots,\mathbf{y}_t\in C(A)$线性无关，则类似地可证明$A^{\mathsf{T}}\mathbf{y}_1,\ldots,A^{\mathsf{T}}\mathbf{y}_t$在行空间中是线性无关的。这说明$rank(A^{\mathsf{T}})\geq rank(A)$.

至此我们证明了引理1，也证明了定理3。

定理4. $C(A)=C(AA^{\top})$，$C(A^{\top})=C(A^{\top}A)$,
特别地, $\mathrm{rank}(AA^{\top})=\mathrm{rank}(A)$

证明1：只需证$C(A)=C(AA^{\top})$

（1）$C(AA^{\top})\subset C(A)\Rightarrow rank(AA^{\top})\leq rank(A)$.

（2）由引理1，若$\mathbf{x}_1,\ldots,\mathbf{x}_s\in C(A)$线性无关$\Rightarrow A^{\top}\mathbf{x}_1,\ldots,A^{\top}\mathbf{x}_s\in C(A^{\top})$线性无关。

$\Rightarrow AA^{\top}\mathbf{x}_1,\ldots,AA^{\top}\mathbf{x}_s\in C(AA^{\top})$线性无关, $\Rightarrow rank(AA^{\top})=\dim(C(AA^{\top}))\geq rank(A)$

所以$C(A)=C(AA^{\top})$

证明2：只需证$C(A)=C(AA^{\top})$

(1)$C(AA^{\top})\subset C(A)$。

(2)若$\mathbf{x}\in C(AA^{\top})^{\perp}$，$\mathbf{x}$与$AA^{\top}$的列正交，$AA^{\top}\mathbf{x}=0\Rightarrow \mathbf{x}^{\top}AA^{\top}\mathbf{x}=0$

即$\|A^{\top}\mathbf{x}\|=0\Rightarrow A^{\top}\mathbf{x}=0$即$\mathbf{x}$与$A^{\top}$的行正交，即$\mathbf{x}$与$A$的列正交，

所以$\mathbf{x}\in C(A)^{\perp}$,所以$C(AA^{\top})^{\perp}\subset C(A)^{\perp}\Rightarrow C(A)\subset C(AA^{\top})$

零/核空间

矩阵$A_{n\times m}$的核空间(kernel space)或零空间(null space):

$$N(A)=\ker(A)=\{\mathbf{x}\in R^m \mid A\mathbf{x}=\mathbf{0}\},$$

$N(A)$为与A的各行正交的向量构成的子空间。

核空间的维数称为$\text{nullity}(A)=\dim(N(A))$。

同样$N(A^{\mathsf{T}})=\{\mathbf{y}\in R^n : A^{\mathsf{T}}\mathbf{y}=\mathbf{0}\}$是与$A$的列正交的核空间。

定理5. $N(A)=C(A^{\mathsf{T}})^{\perp}$， $C(A^{\mathsf{T}})\oplus N(A)=R^m$

$N(A^{\mathsf{T}})=C(A)^{\perp}$， $C(A)\oplus N(A^{\mathsf{T}})=R^n$

$\text{rank}(A_{n\times m})=m-\text{nullity}(A)=n-\text{nullity}(A^{\mathsf{T}})$

证明：设$\forall\mathbf{x}\in N(A), A\mathbf{x}=\mathbf{0}$. 对$\forall\mathbf{y}=A^{\mathsf{T}}\mathbf{b}\in C(A^{\mathsf{T}})$,

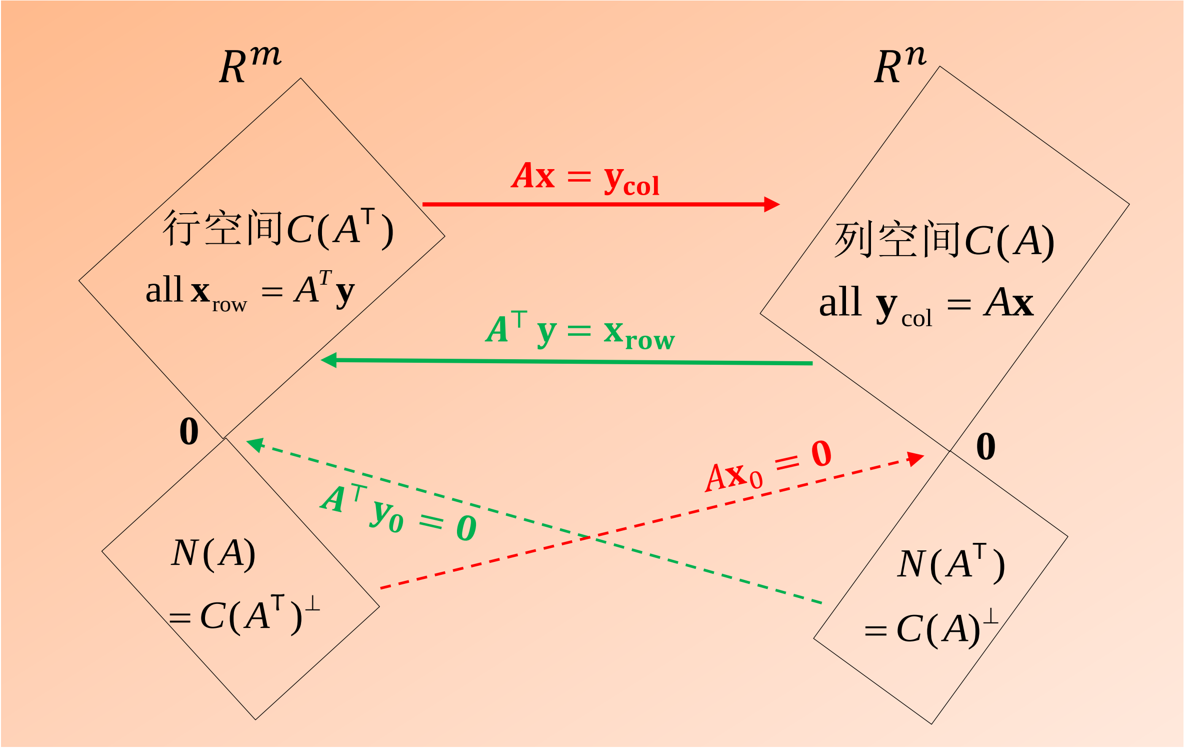
$(\mathbf{x},\mathbf{y})=(\mathbf{x},A^{\mathsf{T}}\mathbf{b})=(A\mathbf{x},\mathbf{b})=(0,\mathbf{b})=0$, 所以$\mathbf{x}\in C(A^{\mathsf{T}})^{\perp}$， $N(A)\subset C(A^{\mathsf{T}})^{\perp}$.

反之，若$\mathbf{x}\in C(A^{\mathsf{T}})^{\perp}, \mathbf{x}\perp C(A^{\mathsf{T}})$, 特别地正交于$A^{\mathsf{T}}$的每一列即$A$的每一行$\mathbf{a}_i$,

所以$A\mathbf{x}=\begin{pmatrix}\mathbf{a}_1{}^{\mathsf{T}}\mathbf{x}\\ \\ \mathbf{a}_n{}^{\mathsf{T}}\mathbf{x}\end{pmatrix}=\mathbf{0}$, 所以$\mathbf{x}\in N(A)$, 所以$C(A^{\mathsf{T}})^{\perp}\subset N(A)$， $C(A^{\mathsf{T}})^{\perp}=N(A)$

所以 $N(A)\oplus C(A^{\mathsf{T}})=R^m$， $m=\dim(N(A))+\dim(C(A^{\mathsf{T}}))=\text{nullity}(A)+\text{rank}(A)$.

四个基本空间（The big picture， Strang）



$R^m = C(A^\top) \oplus + N(A)$ $\qquad$ $R^n = C(A) \oplus + N(A^\top)$