第三章 度量与投影

第 7 讲 正交性与投影

黄定江

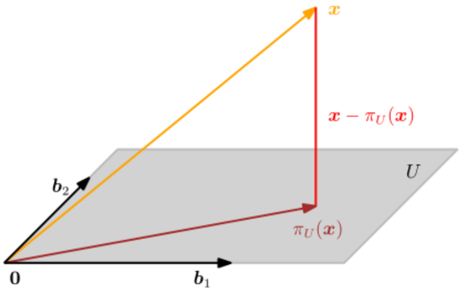
DaSE @ ECNU
djhuang@dase.ecnu.edu.cn

1. 7.1 四个基本子空间
2. 7.2 四个子空间的正交关系
3. 7.3 正交投影
4. 7.4 正交基与 Gram-Schmit

1. 7.1 四个基本子空间
2. 7.2 四个子空间的正交关系
3. 7.3 正交投影
4. 7.4 正交基与 Gram-Schmit

引入

在数据科学的许多工程应用（如信号降噪滤波、数据降维、主成分分析、时间序列分析）中，许多问题的最优求解都可归结为数据在某个子空间的投影问题。本讲我们讲解投影这一在数据分析中极为重要的数学工具。



7.1.1 四个基本子空间

为了更好的理解子空间与投影，我们先讨论四个基本子空间：

1. 列空间:$\mathrm{Col}\,(\boldsymbol{A})$
2. 行空间:$\mathrm{Row}\,(\boldsymbol{A}) = \mathrm{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right)$
3. 零空间:$\mathrm{Null}\,(\boldsymbol{A})$
4. 左零空间:$\mathrm{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right)$

四个基本子空间也是线性代数中非常重要的概念。

7.1.1 四个基本子空间

约定

为方便叙述，对于矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，其 m 个行向量、n 个列向量分别记作

$$\begin{aligned}\boldsymbol{r}_1 &= \left[a_{11}, a_{12}, \cdots, a_{1n}\right]^{\mathrm{T}} \\ \boldsymbol{r}_2 &= \left[a_{21}, a_{22}, \cdots, a_{2n}\right]^{\mathrm{T}} \\ &\cdots \\ \boldsymbol{r}_m &= \left[a_{m1}, a_{m2}, \cdots, a_{mn}\right]^{\mathrm{T}}\end{aligned}$$

$$\boldsymbol{a}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \boldsymbol{a}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \cdots, \boldsymbol{a}_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

即 $\boldsymbol{A} = (\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_m)^{\mathrm{T}} = (\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_n)$

设矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，则它引出下面四个基本子空间：

定义 1

列空间是其列向量 $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n\}$ 的所有线性组合的集合，它是 $\mathbb{R}^m$ 的一个子空间，用符号 $Col(\boldsymbol{A})$ 表示，即有

$$Col(\boldsymbol{A}) = \left\{\boldsymbol{y} \in \mathbb{R}^m | \boldsymbol{y} = \sum_{j=1}^{n} \alpha_j \boldsymbol{a}_j, \alpha_j \in \mathbb{R}\right\} = \mathbf{span}\{\boldsymbol{a}_1, \boldsymbol{a}_2 \cdots \boldsymbol{a}_n\} \tag{1}$$

定义 2

行空间是其行向量 $\{\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_m\}$ 的所有线性组合的集合，它是 $\mathbb{R}^n$ 的一个子空间，用符号 $Row(\boldsymbol{A})$ 表示，也可以用 $Col\left(\boldsymbol{A}^{\mathrm{T}}\right)$ 表示，有

$$Row(\boldsymbol{A}) = Col\left(\boldsymbol{A}^{\mathrm{T}}\right) = \left\{\boldsymbol{y} \in \mathbb{R}^n | \boldsymbol{y} = \sum_{i=1}^{m} \beta_i \boldsymbol{r}_i, \beta_i \in \mathbb{R}\right\} = \mathbf{span}\{\boldsymbol{r}_1, \boldsymbol{r}_2 \cdots \boldsymbol{r}_m\} \tag{2}$$

定义 3

零空间是所有满足齐次线性方程组 $\boldsymbol{A}\boldsymbol{x}=\mathbf{0}$ 的解向量集合，它是 $\mathbb{R}^n$ 的一个子空间，用符号 $Null(\boldsymbol{A})$ 表示，即有

$$Null(\boldsymbol{A})=\{\boldsymbol{x}\in\mathbb{R}^n|\boldsymbol{A}\boldsymbol{x}=\mathbf{0}\} \tag{3}$$

定义 4

左零空间是所有满足齐次线性方程组 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}=\mathbf{0}$ 的解向量集合，它是 $\mathbb{R}^m$ 的一个子空间，用符号 $Null\left(\boldsymbol{A}^{\mathrm{T}}\right)$ 表示，即有

$$Null\left(\boldsymbol{A}^{\mathrm{T}}\right)=\{\boldsymbol{y}\in\mathbb{R}^n|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}=\mathbf{0}\} \tag{4}$$

给定一个矩阵，为了获得其四个基本子空间，我们需要用到以下结论：

定理 1

1. 一系列初等行变换不改变矩阵的行空间。
2. 一系列初等行变换不改变矩阵的零空间。
3. 一系列初等列变换不改变矩阵的列空间。
4. 一系列初等列变换不改变矩阵的左零空间。

证明.

[2] 令 $\boldsymbol{E}_i$ 是对应于矩阵 $\boldsymbol{A}$ 的第 i 次初等行变换的初等矩阵。由初等行变换可逆。于是，

$$\boldsymbol{B}\boldsymbol{x} = (\boldsymbol{E}_k\boldsymbol{E}_{k-1}\cdots\boldsymbol{E}_1\boldsymbol{A})\boldsymbol{x} = \boldsymbol{0} \Leftrightarrow \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$$

即齐次线性方程 $\boldsymbol{B}\boldsymbol{x} = \boldsymbol{0}$ 与 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ 具有相同的解向量，从而 $\boldsymbol{A}$ 经过若干次初等行变换后得到的矩阵 $\boldsymbol{B}$ 与 $\boldsymbol{A}$ 具有相同的零空间。 □

例 1

求 3×3 矩阵

$$\boldsymbol{A}=\begin{pmatrix}1 & 2 & 1\\ -1 & -1 & 1\\ 1 & 4 & 5\end{pmatrix}$$

的行空间、列空间、零空间和左零空间。

解

依次进行初等列变换，得到列简约阶梯型矩阵：

$$\begin{pmatrix}1 & 2 & 1\\ -1 & -1 & 1\\ 1 & 4 & 5\end{pmatrix}\xrightarrow[C_3-C_1]{C_2-2C_1}\begin{pmatrix}1 & 0 & 0\\ -1 & 1 & 2\\ 1 & 2 & 4\end{pmatrix}\xrightarrow[C_3-2C_2]{C_1+C_2}\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 3 & 2 & 0\end{pmatrix}$$

最后的变换结果为：

$$\boldsymbol{A}_C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 2 & 0 \end{pmatrix}$$

由此得到两个线性无关的列向量 $\boldsymbol{c}_1 = (1,0,3)^{\mathrm{T}}, \boldsymbol{c}_2 = (0,1,2)^{\mathrm{T}}$，它们是列空间 $\mathsf{Col}\,(\boldsymbol{A})$ 的基

$$\mathsf{Col}\,(\boldsymbol{A}) = \mathbf{span}\{(1,0,3)^{\mathrm{T}},(0,1,2)^{\mathrm{T}}\}$$

由于一系列初等列变换不改变左零空间，根据 $\boldsymbol{A}_C$，知 $-3\boldsymbol{r}_1 - 2\boldsymbol{r}_2 + \boldsymbol{r}_3 = 0$。
那么我们就可以根据 $\boldsymbol{A}_C$ 的主元位置，矩阵 $\boldsymbol{A}$ 的主元行是第 1 行和第 2 行，即行空间 $\mathsf{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right)$ 可以写作

$$\mathsf{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right) = \mathbf{span}\{(1,2,1)^{\mathrm{T}},(-1,-1,1)^{\mathrm{T}}\}$$

对 $\boldsymbol{A}$ 进行行初等变换

$$\begin{pmatrix} 1 & 2 & 1 \\ -1 & -1 & 1 \\ 1 & 4 & 5 \end{pmatrix} \xrightarrow[R_2+R_1]{R_3-R_1} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \\ 0 & 2 & 4 \end{pmatrix} \xrightarrow[R_1-2R_2]{R_3-2R_2} \boldsymbol{A}_R = \begin{pmatrix} 1 & 0 & -3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

$\boldsymbol{A}$ 的秩为 2。解方程组 $\boldsymbol{A}_R\boldsymbol{x} = \mathbf{0}$ 得到 $\boldsymbol{x} = k(3,-2,1)^{\mathrm{T}}$

$$\mathsf{Null}\,(\boldsymbol{A}) = \mathbf{span}\{(3,-2,1)\}^{\mathrm{T}}$$

所以零空间维数为 1。

类似地，我们求解 $\boldsymbol{A}_C^T\boldsymbol{x} = \mathbf{0}$ 得到 $\boldsymbol{x} = k(3,2,-1)^{\mathrm{T}}$ 所以

$$\mathsf{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right) = \mathbf{span}\{(3,2,-1)^{\mathrm{T}}\}$$

左零空间的维数也是 1。

7.1.2 四个基本子空间的基

我们接下来的目标是: 求四个基本子空间的基和维数。
线性代数的课程中我们学习过矩阵的行秩等于列秩。因此

定理 2

设 $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, 则 $\dim(\mathit{Col}(\boldsymbol{A})) = \dim(\mathit{Row}(\boldsymbol{A})) = \operatorname{rank}(\boldsymbol{A})$

定理 3

设 $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ 则

$$\dim(\mathit{Null}(\boldsymbol{A})) = n - \operatorname{rank}(\boldsymbol{A})$$

证明.

我们令 $r=\operatorname{rank}(\boldsymbol{A})$，根据定义 Null$(\boldsymbol{A})=\{\boldsymbol{x}\in\mathbb{R}^n|\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}\}$ 我们对 $\boldsymbol{A}$ 做行初等变换和列变换，将 $\boldsymbol{A}$ 可以变换为

$$\boldsymbol{A}'=\begin{pmatrix}\boldsymbol{I} & \boldsymbol{B}\\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}=\begin{pmatrix}1 & & & & b_{11} & b_{12} & \dots & b_{1,n-r}\\ & 1 & & & b_{21} & b_{22} & \dots & b_{2,n-r}\\ & & \ddots & & \vdots & \vdots & & \vdots\\ & & & 1 & b_{r1} & b_{r2} & \dots & b_{r,n-r}\\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots\\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0\end{pmatrix}$$

□

显然 $\boldsymbol{A}'\boldsymbol{x}=\boldsymbol{0}$ 有以下 $n-r$ 个解

$$\boldsymbol{x}^{(1)}=\begin{pmatrix} b_{11} \\ \vdots \\ b_{r1} \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \boldsymbol{x}^{(2)}=\begin{pmatrix} b_{12} \\ \vdots \\ b_{r2} \\ 0 \\ -1 \\ \vdots \\ 0 \end{pmatrix} \dots \boldsymbol{x}^{(n-r)}=\begin{pmatrix} b_{1,n-r} \\ \vdots \\ b_{r,n-r} \\ 0 \\ 0 \\ \vdots \\ -1 \end{pmatrix}$$

并且容易看出向量组 $(\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)},\dots,\boldsymbol{x}^{(n-r)})$ 是一个极大线性无关组。

再注意到，如果 $\boldsymbol{x}$ 是方程 $\boldsymbol{A}'\boldsymbol{x}=\boldsymbol{0}$ 的解，那么当 $x_{r+1},x_{r+2},\dots,x_n$ 取定时，可以唯一确定 $\boldsymbol{x}$。换句话说 $\{\boldsymbol{x}\in\mathbb{R}^n|\boldsymbol{A}'\boldsymbol{x}=\boldsymbol{0}\}$ 的维数最大为 $n-r$。

综上 $\boldsymbol{A}'\boldsymbol{x}=\boldsymbol{0}$ 解空间的维数为 $n-r$，即 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$ 解空间的维数为 $n-r$，即

$$\dim(\mathsf{Null}\,(\boldsymbol{A}))=n-r$$

上述的证明过程实际上也就是我们刚刚求解矩阵 $\boldsymbol{A}$ 零空间 Null $(\boldsymbol{A})$ 基底和维数的过程。由此得到秩定理，描述了矩阵的秩与其零空间维数之间的关系。

定理 4

矩阵 $\boldsymbol{A}_{m\times n}$ 的列空间和行空间的维数相等。这个共同的维数就是矩阵 $\boldsymbol{A}$ 的秩 *rank*$(\boldsymbol{A})$, 它与零空间维数之间有下列关系：

$$\dim(Col(\boldsymbol{A})) + \dim[Null(\boldsymbol{A})] = n \tag{5}$$

利用上述定理我们立刻可以得到以下推论

推论 1

设 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 则

$$\dim(Null\left(\boldsymbol{A}^{\mathrm{T}}\right)) = m - \operatorname{rank}(\boldsymbol{A})$$

1. 7.1 四个基本子空间
2. 7.2 四个子空间的正交关系
3. 7.3 正交投影
4. 7.4 正交基与 Gram-Schmit

7.2 四个子空间的正交关系

我们将继续讨论四个基本子空间之间的关系。

设 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，$\boldsymbol{A}$ 的四个基本子空间中，$\text{Col}\,(\boldsymbol{A})\,,\text{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right)$ 都是 $\mathbb{R}^m$ 的子空间，它们是否有交集？$\text{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right),\text{Null}\,(\boldsymbol{A})$ 都是 $\mathbb{R}^n$ 的子空间，它们是否有交集？

定理 5

设 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，

$$\mathit{Col}\,(\boldsymbol{A}) \cap \mathit{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right) = \{\mathbf{0}\}$$

$$\mathit{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right) \cap \mathit{Null}\,(\boldsymbol{A}) = \{\mathbf{0}\}$$

证明.

设 $\boldsymbol{v} \in \operatorname{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right) \cap \operatorname{Null}(\boldsymbol{A})$, 即 $\boldsymbol{v}$ 在 $\boldsymbol{A} = (\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_m)^{\mathrm{T}}$ 的行空间中且 $\boldsymbol{A}\boldsymbol{v} = \mathbf{0}$。
设 $\boldsymbol{v} = a_1\boldsymbol{r}_1 + a_2\boldsymbol{r}_2 + \ldots + a_m\boldsymbol{r}_m$, 则

$$\boldsymbol{A}\boldsymbol{v} = \mathbf{0} \implies \boldsymbol{r}_1^{\mathrm{T}}\boldsymbol{v} = 0, \ldots, \boldsymbol{r}_m^{\mathrm{T}}\boldsymbol{v} = 0 \implies \boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} = 0 \implies \boldsymbol{v} = 0$$

即

$$\operatorname{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right) \cap \operatorname{Null}(\boldsymbol{A}) = \{\mathbf{0}\}.$$

同理 $\operatorname{Col}(\boldsymbol{A}) \cap \operatorname{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right) = \{\mathbf{0}\}$。 □

定义 5

设 $\mathbb{S}$ 和 $\mathbb{T}$ 是 $\mathbb{R}^n$ 的两个子空间。如果

$$\mathbb{S} \cap \mathbb{T} = \{\mathbf{0}\}$$

我们称 $\mathbb{S}$ 和 $\mathbb{T}$ **无交连**。

列空间和左零空间是无交连的，行空间和零空间是无交连的。

定义 6

设 $\mathbb{S}$ 和 $\mathbb{T}$ 是 $\mathbb{R}^n$ 的两个子空间。如果对于 $\forall \boldsymbol{v} \in \mathbb{S}, \forall \boldsymbol{w} \in \mathbb{T}$，均有

$$\boldsymbol{v}^{\mathrm{T}} \boldsymbol{w} = 0$$

我们说 $\mathbb{S}$ **垂直于** $\mathbb{T}$，$\mathbb{T}$ **垂直于** $\mathbb{S}$，记做 $\mathbb{S} \perp \mathbb{T}, \mathbb{T} \perp \mathbb{S}$。
或者说，子空间 $\mathbb{S}$ 和子空间 $\mathbb{T}$ 是**正交**的。

定理 6

正交的两个子空间必定是无交连的。

证明.

假设 $\mathbb{R}^n$ 中的两个子空间 $\mathbb{S}$，$\mathbb{T}$ 不是无交连的

则

$$\exists \boldsymbol{v} \neq \mathbf{0}, \boldsymbol{v} \in \mathbb{S} \cap \mathbb{T}$$

而

$$\boldsymbol{v}^{\mathrm{T}} \boldsymbol{v} \neq 0$$

因而 $\mathbb{S}$ 和 $\mathbb{T}$ 不正交。从而正交的两个子空间必是无交连的。 □

显然，无交连的子空间不一定是正交的。如 $\mathbf{span}\{(1,1)^{\mathrm{T}}\}$ 和 $\mathbf{span}\{(1,0)^{\mathrm{T}}\}$。

那么列空间和左零空间，行空间和零空间是正交的么？

例 2

设 $\boldsymbol{A}$ 是 $m \times n$ 阶阵，则 *Col*$(\boldsymbol{A})$ 和 *Null*$\left(\boldsymbol{A}^{\mathrm{T}}\right)$ 正交，*Col*$\left(\boldsymbol{A}^{\mathrm{T}}\right)$ 和 *Null*$(\boldsymbol{A})$ 正交。

证明.

对 $\forall \boldsymbol{v} \in \mathrm{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right)$，则

$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A} = \mathbf{0} \implies \boldsymbol{v}^{\mathrm{T}}\boldsymbol{a}_1 = 0, \boldsymbol{v}^{\mathrm{T}}\boldsymbol{a}_2 = 0, \cdots, \boldsymbol{v}^{\mathrm{T}}\boldsymbol{a}_n = 0$$

对 $\forall \boldsymbol{w} \in \mathrm{Col}(\boldsymbol{A})$，有 $\boldsymbol{w} = \alpha_1\boldsymbol{a}_1 + \alpha_2\boldsymbol{a}_2 + \cdots + \alpha_n\boldsymbol{a}_n$:

$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w} = \alpha_1\boldsymbol{v}^{\mathrm{T}}\boldsymbol{a}_1 + \alpha_2\boldsymbol{v}^{\mathrm{T}}\boldsymbol{a}_2 + \cdots + \alpha_n\boldsymbol{v}^{\mathrm{T}}\boldsymbol{a}_n = 0$$

因此，$\mathrm{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right) \perp \mathrm{Col}(\boldsymbol{A})$，$\mathrm{Col}(\boldsymbol{A})$ 和 $\mathrm{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right)$ 正交。

将 $\boldsymbol{A}$ 换成 $\boldsymbol{A}^{\mathrm{T}}$，我们得到 $\mathrm{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right) \perp \mathrm{Null}(\boldsymbol{A})$，$\mathrm{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right)$ 和 $\mathrm{Null}(\boldsymbol{A})$ 正交。 □

相对于正交，正交补是两个子空间更强的一种关系。

定义 7

设 $\mathbb{V} \subset \mathbb{R}^n$ 是一个子空间,$\mathbb{V}$ 在 $\mathbb{R}^n$ 中的正交补定义为集合

$$\{\boldsymbol{w} \in \mathbb{R}^n | \boldsymbol{v}^{\mathrm{T}} \boldsymbol{w} = 0, \forall \boldsymbol{v} \in \mathbb{V}\}$$

记作 $\mathbb{V}^{\perp}$。

也就是说 $\mathbb{V}$ 的正交补空间是 $\mathbb{R}^n$ 中所有和 $\mathbb{V}$ 正交的向量构成的集合。
显然一个空间和它的正交补空间是正交的，即 $\mathbb{V} \perp \mathbb{V}^{\perp}$。

显然 $\mathbb{V}$ 与 $\mathbb{V}^{\perp}$ 的和是直和，因此，对于 $\mathbb{R}^n$ 中的任意向量 $\boldsymbol{x}$ 可以唯一的分解成如下形式:

$$\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2$$

其中 $\boldsymbol{x}_1 \in \mathbb{V}, \boldsymbol{x}_2 \in \mathbb{V}^{\perp}$ 并且 $\boldsymbol{x}_1^{\mathrm{T}} \boldsymbol{x}_2 = 0$。这种分解形式叫做向量的**正交分解**。

例 3

证明：$Col\left(\boldsymbol{A}^{\mathrm{T}}\right)^{\perp}=Null(\boldsymbol{A})$，$Col(\boldsymbol{A})^{\perp}=Null\left(\boldsymbol{A}^{\mathrm{T}}\right)$。

我们已经知道，$\text{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right)$ 和 $\text{Null}(\boldsymbol{A})$ 是正交的，也就是说

$$\text{Null}(\boldsymbol{A}) \subseteq \text{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right)^{\perp}$$

对 $\forall \boldsymbol{x} \in \text{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right)^{\perp}$，$\boldsymbol{x}$ 和 $\text{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right)$ 中的任意向量正交，那么：

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{r}_1=0, \boldsymbol{x}^{\mathrm{T}}\boldsymbol{r}_2=0, \cdots, \boldsymbol{x}^{\mathrm{T}}\boldsymbol{r}_m=0$$

即 $\boldsymbol{A}\boldsymbol{x}=\mathbf{0}$。说明 $\boldsymbol{x} \in \text{Null}(\boldsymbol{A})$。也即

$$\text{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right)^{\perp} \subseteq \text{Null}(\boldsymbol{A})$$

因此 $\text{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right)^{\perp}=\text{Null}(\boldsymbol{A})$。同样可以证明 $\text{Col}(\boldsymbol{A})^{\perp}=\text{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right)$。

顾名思义，子空间 $\mathbb{V}$ 在向量空间 $\mathbb{R}^n$ 的正交补空间 $\mathbb{V}^\perp$ 含有正交和补充双重含义：

1. 子空间 $\mathbb{V}^\perp$ 与 $\mathbb{V}$ 正交；
2. 向量空间 $\mathbb{R}^n$ 是子空间 $\mathbb{V}$ 与 $\mathbb{V}^\perp$ 的直和，即 $\mathbb{R}^n = \mathbb{V} \oplus \mathbb{V}^\perp$。这表明，向量空间 $\mathbb{R}^n$ 是由子空间 $\mathbb{V}$ 补充 $\mathbb{V}^\perp$ 而成。

正交补空间是一个比正交子空间更严格的概念：当向量空间 $\mathbb{R}^n$ 和子空间 $\mathbb{V}$ 给定之后，和 $\mathbb{V}$ 正交的空间不一定是唯一的，但是 $\mathbb{V}$ 的正交补 $\mathbb{V}^\perp$ 是唯一的。

线性代数基本定理

我们将本节内容总结成线性代数基本定理：

定理 7

(线性代数基本定理) 若 $\boldsymbol{A}$ 是 $m \times n$ 矩阵，

1. *[正交角度]* $\mathit{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right) \perp \mathit{Null}(\boldsymbol{A})$， $\mathit{Col}(\boldsymbol{A}) \perp \mathit{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right)$，
2. *[扩张角度]* $\mathit{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right) \oplus \mathit{Null}(\boldsymbol{A}) = \mathbb{R}^n$， $\mathit{Col}(\boldsymbol{A}) \oplus \mathit{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right) = \mathbb{R}^m$，
3. *[维数角度]* $\dim \mathit{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right) + \dim \mathit{Null}(\boldsymbol{A}) = n$， $\dim \mathit{Col}(\boldsymbol{A}) + \dim \mathit{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right) = m$。

线性代数基本定理

$$\mathsf{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right)^{\perp} = \mathsf{Null}\left(\boldsymbol{A}\right)$$

$$\mathsf{Col}\left(\boldsymbol{A}\right)^{\perp} = \mathsf{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right)$$

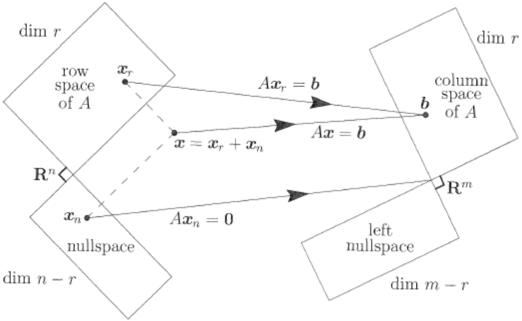


图 1: 四个子空间

1 7.1 四个基本子空间

2 7.2 四个子空间的正交关系

3 7.3 正交投影

4 7.4 正交基与 Gram-Schmit

7.3.1 引言

- 投影是一类重要的线性变换。
- 投影在图形学，编码理论，统计和机器学习中起着重要作用。
- 在机器学习中，我们经常处理高维数据。高维数据通常很难分析或可视化。
- 但是，高维数据通常具有以下属性：只有少数维包含大多数信息，而其它大多数维对于描述数据的关键属性也不是必需的。当我们压缩或可视化高维数据时，我们将丢失信息。
- 为了最大程度地减少这种压缩损失，我们理想地希望在数据中找到最有用的信息维度。然后，我们可以将原始的高维数据投影到低维特征空间上，并在此低维空间中进行操作，以了解有关数据集的更多信息并提取模式。
- 例如机器学习中主成分分析（PCA）、深度学习中深度自动编码器大量采用了降维的想法。
- 下面，我们将专注于正交投影。

投影

定义 8

设 $\mathbb{V}$ 是一向量空间,$\mathbb{U} \subseteq \mathbb{V}$ 是 $\mathbb{V}$ 的一个子空间。如果线性映射 $\pi : \mathbb{V} \to \mathbb{U}$ 满足

$$\pi^2 = \pi \circ \pi = \pi$$

则称 π 为投影。

设 π 对应的矩阵 $\boldsymbol{P}_\pi$，显然 $\boldsymbol{P}_\pi$ 满足 $\boldsymbol{P}_\pi^2 = \boldsymbol{P}_\pi$，称 $\boldsymbol{P}_\pi$ 为投影矩阵。

正交投影

本节所讲的投影指正交投影。

正交投影

即给定定义了标准内积和欧氏距离的向量空间 $\mathbb{R}^n$ 中的向量 $\boldsymbol{x}$，$\mathbb{U}$ 是 $\mathbb{R}^n$ 的子空间，求 $\boldsymbol{y} \in \mathbb{U}$，使得 $\|\boldsymbol{y}-\boldsymbol{x}\|$ 最小，即

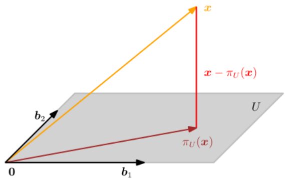
$$\pi_{\mathbb{U}}(\boldsymbol{x}) = \arg\min_{\boldsymbol{y}\in\mathbb{U}} \|\boldsymbol{y}-\boldsymbol{x}\|,$$

称向量 $\boldsymbol{y}$ 为向量 $\boldsymbol{x}$ 在子空间 $\mathbb{U}$ 的正交投影。

因为可以对 $\boldsymbol{x}$ 正交分解，$\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2$，其中 $\boldsymbol{x}_1 \in \mathbb{U}$，$\boldsymbol{x}_2 \in \mathbb{U}^{\perp}$。所以

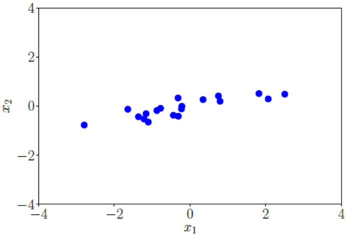
$$\|\boldsymbol{y}-\boldsymbol{x}\|^2 = \|\boldsymbol{y}-(\boldsymbol{x}_1+\boldsymbol{x}_2)\|^2 = \|(\boldsymbol{x}_1-\boldsymbol{y})+\boldsymbol{x}_2\|^2 .$$

而 $\boldsymbol{x}_1-\boldsymbol{y} \in \mathbb{U}$，$\boldsymbol{x}_2 \in \mathbb{U}^{\perp}$，所以 $\|(\boldsymbol{x}_1-\boldsymbol{y})+\boldsymbol{x}_2\|^2 = \|\boldsymbol{x}_1-\boldsymbol{y}\|^2+\|\boldsymbol{x}_2\|^2$。所以我们只需令 $\boldsymbol{y}=\boldsymbol{x}_1$ 即可，那么 $\boldsymbol{x}_2 = \boldsymbol{x}-\boldsymbol{y} = \boldsymbol{x}-\pi_{\mathbb{U}}(\boldsymbol{x}) \in \mathbb{U}^{\perp}$。

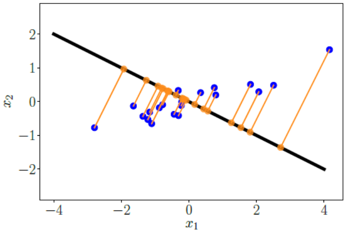


7.3.2 投影到 1 维子空间

接下来，我们看一下如何寻找一个投影矩阵 $\boldsymbol{P}_\pi$ 使得向量投影到某个 1 维子空间上。



(a) Original dataset.



(b) Original data (blue) and their corresponding orthogonal projections (orange) onto a lower-dimensional subspace (straight line).

将 2 维空间的点投影到 1 维子空间上。

7.3.2 投影到 1 维子空间

投影到一维子空间

假设给定 $\mathbb{R}^n$ 中一条通过原点的直线（1 维子空间），其具有基向量 $\boldsymbol{b}$，相应的基底矩阵表示为 $\boldsymbol{B}=[\boldsymbol{b}]$，也就是说这组基中仅有一个向量。

这条直线是由 $\boldsymbol{b}$ 张成的一维子空间 $\mathbb{U}=\mathsf{Col}\,(\boldsymbol{B})\subseteq\mathbb{R}^n$。

假设 $\boldsymbol{x}\in\mathbb{R}^n$，当把 $\boldsymbol{x}$ 投影到 $\mathbb{U}$ 时，我们想寻找一个点 $\pi_{\mathbb{U}}(\boldsymbol{x})\in\mathbb{U}$ 最接近 $\boldsymbol{x}$，即

$$\pi_{\mathbb{U}}(\boldsymbol{x})=\arg\min_{\boldsymbol{y}\in\mathbb{U}}\|\boldsymbol{y}-\boldsymbol{x}\|$$

因为 $\pi_{\mathbb{U}}(\boldsymbol{x})\in\mathbb{U}$，又 $\mathbb{U}=\mathsf{Col}\,(\boldsymbol{B})=\mathbf{span}\{\boldsymbol{b}\}$，所以 $\pi_{\mathbb{U}}(\boldsymbol{x})=\lambda\boldsymbol{b}$，$\lambda\in\mathbb{R}$。

我们将结合 $\boldsymbol{x}-\pi_{\mathbb{U}}(x)\in\mathbb{U}^{\perp}$，逐步确定坐标 λ，投影 $\pi_{\mathbb{U}}(\boldsymbol{x})\in\mathbb{U}$ 和 $\pi_{\mathbb{U}}$ 的投影矩阵 $\boldsymbol{P}_{\pi}$。

1. 确定 λ

因为 $\pi_{\mathbb{U}}(\boldsymbol{x}) \in \mathrm{Col}\,(\boldsymbol{B})$ 是 $\boldsymbol{x}$ 的投影，所以 $\boldsymbol{x} - \pi_{\mathbb{U}}(\boldsymbol{x}) \in \mathrm{Col}\,(\boldsymbol{B})^{\perp} = \mathrm{Null}\left(\boldsymbol{B}^{\mathrm{T}}\right)$，有

$$\boldsymbol{b}^{\mathrm{T}}(\boldsymbol{x} - \pi_{\mathbb{U}}) = 0 \iff \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} - \lambda \boldsymbol{b}^{\mathrm{T}}\boldsymbol{b} = 0$$

从而

$$\lambda = \frac{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}}{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}}$$

或者利用内积和范数表示可得

$$\langle \boldsymbol{x}, \boldsymbol{b} \rangle - \lambda \langle \boldsymbol{b}, \boldsymbol{b} \rangle = 0 \iff \lambda = \frac{\langle \boldsymbol{x}, \boldsymbol{b} \rangle}{\langle \boldsymbol{b}, \boldsymbol{b} \rangle} = \frac{\langle \boldsymbol{x}, \boldsymbol{b} \rangle}{||\boldsymbol{b}||^2}.$$

7.3.2 投影到 1 维子空间

2. 确定 $\pi_{\mathbb{U}}(\boldsymbol{x})$

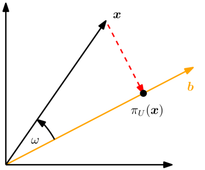
因为 $\pi_{\mathbb{U}}(\boldsymbol{x}) = \lambda \boldsymbol{b}$，由上面的结论可得：

$$\pi_{\mathbb{U}}(\boldsymbol{x}) = \frac{\langle \boldsymbol{x}, \boldsymbol{b} \rangle}{||\boldsymbol{b}||^2} \boldsymbol{b} = \frac{\boldsymbol{b}^{\mathrm{T}} \boldsymbol{x}}{||\boldsymbol{b}||^2} \boldsymbol{b}$$

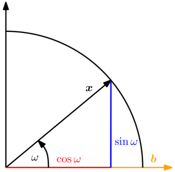
我们可以给出 $\pi_{\mathbb{U}}(\boldsymbol{x})$ 的长度

$$\begin{aligned} ||\pi_{\mathbb{U}}(\boldsymbol{x})|| &= ||\lambda \boldsymbol{b}|| = |\lambda| ||\boldsymbol{b}|| \\ &= |\cos\omega| ||\boldsymbol{x}|| ||\boldsymbol{b}|| \frac{||\boldsymbol{b}||}{||\boldsymbol{b}||^2} \\ &= |\cos\omega| ||\boldsymbol{x}|| \end{aligned}$$

其中 ω 是 $\boldsymbol{x}$ 和 $\boldsymbol{b}$ 之间的夹角，$\cos\omega = \frac{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}}{||\boldsymbol{b}|| ||\boldsymbol{x}||}$。



(a) Projection of $\boldsymbol{x} \in \mathbb{R}^2$ onto a subspace U with basis vector $\boldsymbol{b}$.



(b) Projection of a two-dimensional vector $\boldsymbol{x}$ with $||\boldsymbol{x}|| = 1$ onto a one-dimensional subspace spanned by $\boldsymbol{b}$.

7.3.2 投影到 1 维子空间

3. 确定投影矩阵 $\boldsymbol{P}_\pi$

投影矩阵 $\boldsymbol{P}_\pi$ 是投影 $\pi_{\mathbb{U}}(\boldsymbol{x})$ 对应的变换矩阵, 那么就有 $\pi_{\mathbb{U}}(\boldsymbol{x}) = \boldsymbol{P}_\pi \boldsymbol{x}$, 则有

$$\pi_{\mathbb{U}}(\boldsymbol{x}) = \lambda \boldsymbol{b} = \boldsymbol{b}\lambda = \boldsymbol{b}\frac{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}}{||\boldsymbol{b}||^2} = \frac{\boldsymbol{b}\boldsymbol{b}^{\mathrm{T}}}{||\boldsymbol{b}||^2}\boldsymbol{x}$$

我们立刻可以看出

$$\boldsymbol{P}_\pi = \frac{\boldsymbol{b}\boldsymbol{b}^{\mathrm{T}}}{||\boldsymbol{b}||^2}$$

注:

$\boldsymbol{b}\boldsymbol{b}^{\mathrm{T}}$ 是一个对称矩阵，而 $\boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}$ 则是一个标量。

例 4

确定投影到 $\mathbb{R}^3$ 的子空间 $\mathbf{span}\{\boldsymbol{b}\}$ 上的投影矩阵 $\boldsymbol{P}_\pi$，其中 $\boldsymbol{b}=(1,2,2)^{\mathrm{T}}$。

由上面的结论可得

$$\boldsymbol{P}_\pi=\frac{\boldsymbol{b}\boldsymbol{b}^{\mathrm{T}}}{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}}=\frac{1}{9}\begin{pmatrix}1\\2\\2\end{pmatrix}\begin{pmatrix}1&2&2\end{pmatrix}=\frac{1}{9}\begin{pmatrix}1&2&2\\2&4&4\\2&4&4\end{pmatrix}$$

给定向量 $\boldsymbol{x}=(1,1,1)^{\mathrm{T}}$ 其投影为

$$\pi_{\mathbb{U}}(\boldsymbol{x})=\boldsymbol{P}_\pi(\boldsymbol{x})=\frac{1}{9}\begin{pmatrix}1&2&2\\2&4&4\\2&4&4\end{pmatrix}\begin{pmatrix}1\\1\\1\end{pmatrix}=\frac{1}{9}\begin{pmatrix}5\\10\\10\end{pmatrix}\in \mathit{Col}\left(\begin{pmatrix}1\\2\\2\end{pmatrix}\right)$$

7.3.3 投影到一般子空间

我们将 $\mathbb{R}^m$ 中的向量 $\boldsymbol{x} \in \mathbb{R}^m$ 投影到更高维的子空间 $\mathbb{U} \subseteq \mathbb{R}^m$ 中，其中 $\dim(\mathbb{U}) = n \geq 1$。

投影到一般的子空间中

设 $\boldsymbol{B} = (\boldsymbol{b}_1, ..., \boldsymbol{b}_n)$ 是子空间 $\mathbb{U}$ 的一个有序基底。$\mathbb{U}$ 上的任何投影 $\pi_{\mathbb{U}}(\boldsymbol{x})$ 必须是 $\mathbb{U}$ 中的一个元素。故有

$$\pi_{\mathbb{U}}(\boldsymbol{x}) = \sum_{i=1}^{n} \lambda_i \boldsymbol{b}_i$$

和一维情况一样，我们将逐步确定 $\lambda_1, ..., \lambda_n$，$\pi_{\mathbb{U}}(\boldsymbol{x})$ 和投影矩阵 $\boldsymbol{P}_\pi$。

1. 确定 $\lambda_1, ..., \lambda_n$

设

$$\pi_{\mathbb{U}}(\boldsymbol{x}) = \sum_{i=1}^{n} \lambda_i \boldsymbol{b}_i = \boldsymbol{B}\boldsymbol{\lambda} \in \mathsf{Col}\,(\boldsymbol{B})$$

最接近 $\boldsymbol{x} \in \mathbb{R}^m$，其中 $\boldsymbol{B} = [\boldsymbol{b}_1, ..., \boldsymbol{b}_n] \in \mathbb{R}^{m\times n}, \boldsymbol{\lambda} = [\lambda_1, ..., \lambda_n]^{\mathrm{T}} \in \mathbb{R}^n$。

因为 $\pi_{\mathbb{U}}(\boldsymbol{x})$ 是 $\boldsymbol{x}$ 的投影，所以 $\boldsymbol{x} - \pi_{\mathbb{U}}(\boldsymbol{x}) \in \mathsf{Col}\,(\boldsymbol{B})^{\perp} = \mathsf{Null}\,(\boldsymbol{B}^{\mathrm{T}})$

$$\begin{aligned}
\boldsymbol{b}_1^{\mathrm{T}}(\boldsymbol{x} - \pi_{\mathbb{U}}(\boldsymbol{x})) &= \langle \boldsymbol{b}_1, \boldsymbol{x} - \pi_{\mathbb{U}}(\boldsymbol{x}) \rangle = 0 \\
\boldsymbol{b}_2^{\mathrm{T}}(\boldsymbol{x} - \pi_{\mathbb{U}}(\boldsymbol{x})) &= \langle \boldsymbol{b}_2, \boldsymbol{x} - \pi_{\mathbb{U}}(\boldsymbol{x}) \rangle = 0 \\
&\vdots \\
\boldsymbol{b}_n^{\mathrm{T}}(\boldsymbol{x} - \pi_{\mathbb{U}}(\boldsymbol{x})) &= \langle \boldsymbol{b}_n, \boldsymbol{x} - \pi_{\mathbb{U}}(\boldsymbol{x}) \rangle = 0
\end{aligned}$$

使用矩阵可以将上式改写成

$$\boldsymbol{b}_1^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{B\lambda})=0$$
$$\vdots$$
$$\boldsymbol{b}_n^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{B\lambda})=0$$

故有

$$\begin{pmatrix}\boldsymbol{b}_1^{\mathrm{T}}\\ \vdots \\ \boldsymbol{b}_n^{\mathrm{T}}\end{pmatrix}\Big(\boldsymbol{x}-\boldsymbol{B\lambda}\Big)=\mathbf{0}\iff \boldsymbol{B}^{T}(\boldsymbol{x}-\boldsymbol{B\lambda})=\mathbf{0}\iff \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B\lambda}=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{x}$$

最终的方程我们称之为正规方程。因为 $\boldsymbol{b}_1,\ldots,\boldsymbol{b}_n$ 是 $\mathbb{U}$ 的基。因此 $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$ 是可逆的 $(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{By}=\mathbf{0}\implies \boldsymbol{y}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{By}=0\implies \boldsymbol{By}=\mathbf{0})$。也就是说

$$\boldsymbol{\lambda}=(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{x}$$

2. 确定 $\pi_{\mathbb{U}}(\boldsymbol{x})$

$$\boldsymbol{\lambda} = (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{x}$$

$\boldsymbol{\lambda}$ 也就是 $\pi_{\mathbb{U}}(\boldsymbol{x})$ 在有序基底 $\boldsymbol{B}$ 下的坐标。

$$\pi_{\mathbb{U}}(\boldsymbol{x}) = \boldsymbol{B}\boldsymbol{\lambda} = \boldsymbol{B}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{x}$$

3. 确定 $\boldsymbol{P}_{\pi}$

由上面的讨论容易看出

$$\boldsymbol{P}_{\pi} = \boldsymbol{B}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}$$

例 5

已知 $\mathbb{R}^3$ 中的子空间 $\mathbb{U} = \mathbf{span}\left\{\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}0\\1\\2\end{pmatrix}\right\}$ 和向量 $\boldsymbol{x} = \begin{pmatrix}6\\0\\0\end{pmatrix}$,

确定 $\boldsymbol{x}$ 投影到 $\mathbb{U}$ 上的坐标 $\boldsymbol{\lambda}$ 和投影点 $\pi_{\mathbb{U}}(\boldsymbol{x})$ 和投影矩阵 $\boldsymbol{P}_\pi$

解

首先确定 $\boldsymbol{B} = \begin{pmatrix}1 & 0\\1 & 1\\1 & 2\end{pmatrix}$

其次计算

$$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} = \begin{pmatrix}1 & 1 & 1\\0 & 1 & 2\end{pmatrix}\begin{pmatrix}1 & 0\\1 & 1\\1 & 2\end{pmatrix} = \begin{pmatrix}3 & 3\\3 & 5\end{pmatrix}, \quad \boldsymbol{B}^{\mathrm{T}}\boldsymbol{x} = \begin{pmatrix}1 & 1 & 1\\0 & 1 & 2\end{pmatrix}\begin{pmatrix}6\\0\\0\end{pmatrix} = \begin{pmatrix}6\\0\end{pmatrix}$$

然后只需要解方程 $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{\lambda}=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{x}$ 得到 $\boldsymbol{\lambda}$,

$$\begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix}\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}=\begin{pmatrix} 6 \\ 0 \end{pmatrix} \iff \boldsymbol{\lambda}=\begin{pmatrix} 5 \\ -3 \end{pmatrix}$$

故投影点 $\pi_{\mathbb{U}}(\boldsymbol{x})=\boldsymbol{B}\boldsymbol{\lambda}=\begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix}$。最后

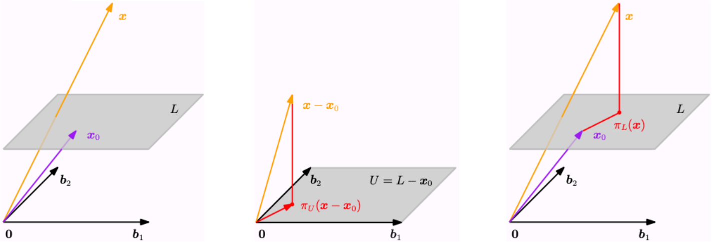
$$\boldsymbol{P}_\pi=\boldsymbol{B}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}=\frac{1}{6}\begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}$$

我们还可以验证 $\boldsymbol{P}_\pi^2=\boldsymbol{P}_\pi$

- 投影使我们可以研究线性系统 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 没有解的情况。
- 回想一下，方程无解意味着 $\boldsymbol{b}$ 不在 $\boldsymbol{A}$ 的范围内，即向量 $\boldsymbol{b}$ 不在 $\boldsymbol{A}$ 的列张成的子空间中。
- 鉴于线性方程无法精确求解，我们可以找到一个近似解。
- 想法是在 $\boldsymbol{A}$ 的列张成的子空间中找到最接近 $\boldsymbol{b}$ 的向量，即，我们计算 $\boldsymbol{b}$ 到 $\boldsymbol{A}$ 列张成的子空间上的正交投影。
- 此问题在实践中经常出现，并且解被称为超定系统的最小二乘解。

7.3.4 投影到仿射子空间

到目前为止，我们讨论了如何将向量投影到低维子空间 $\mathbb{U}$ 上。
下面，我们将讨论如何将向量投影到仿射子空间上。



(a) Setting.

(b) Reduce problem to projection π_U onto vector subspace.

(c) Add support point back in to get affine projection π_L.

图 2: 投影到仿射空间

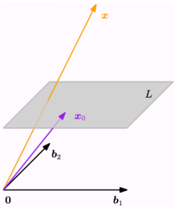
考虑图 (a)。给定一个仿射空间 $\mathbb{L} = \boldsymbol{x}_0 + \mathbb{U}$，其中 $\boldsymbol{b}_1$ $\boldsymbol{b}_2$ 是 $\mathbb{U}$ 的基向量。为了确定 $\boldsymbol{x}$ 在 $\mathbb{L}$ 上的正交投影 $\pi_{\mathbb{L}}(\boldsymbol{x})$。

我们将问题转化为我们知道如何解决的问题：投影到向量子空间上。为此，我们从 $\boldsymbol{x}$ 和 $\mathbb{L}$ 中减去支撑点 $\boldsymbol{x}_0$，所以 $\mathbb{L} - \boldsymbol{x}_0 = \mathbb{U}$ 恰好是向量子空间 $\mathbb{U}$。

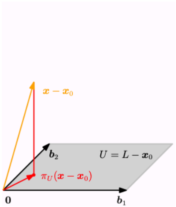
现在，我们可以用前面讨论过的在子空间上的正交投影，来获得投影 $\pi_{\mathbb{U}}(\boldsymbol{x} - \boldsymbol{x}_0)$，如图 (b) 所示。

最后我们通过添加 $\boldsymbol{x}_0$ 将该投影转换回 $\mathbb{L}$，这样我们就可以得出仿射空间 $\mathbb{L}$ 上的正交投影为

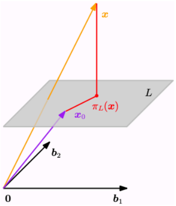
$$\pi_{\mathbb{L}}(\boldsymbol{x}) = \boldsymbol{x}_0 + \pi_{\mathbb{U}}(\boldsymbol{x} - \boldsymbol{x}_0)$$



(a) Setting.



(b) Reduce problem to projection π_U onto vector subspace.



(c) Add support point back in to get affine projection π_L.

1 7.1 四个基本子空间

2 7.2 四个子空间的正交关系

3 7.3 正交投影

4 7.4 正交基与 Gram-Schmit

7.4.1 标准正交基

线性代数中已经学过，线性空间中的向量可以由该空间的一组基表示。

[回忆：标准正交基]

设 n 维向量 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_r$ 是向量空间 $\mathbb{V}(\mathbb{V} \subset \mathbb{R}^n)$ 的一个基，如果 $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_r$ 两两正交，且都是单位向量，即对于 $\forall i, j = 1, \cdots, r$，有

$$\langle \boldsymbol{e}_i, \boldsymbol{e}_j \rangle = \begin{cases} 0, i \neq j \\ 1, i = j \end{cases}$$

则称 $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_r$ 是 $\mathbb{V}$ 的一个规范 (标准) 正交基，有时也简称做正交基。

为什么需要标准正交基的概念？用标准正交基表示向量有什么好处呢？

7.4.1 标准正交基

若 $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_r$ 是 $\mathbb{V}$ 的一个规范正交基，那么 $\mathbb{V}$ 中任意向量 $\boldsymbol{a}$ 可以由 $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_r$ 线性表示，设表示为

$$\boldsymbol{a} = \lambda_1 \boldsymbol{e}_1 + \lambda_2 \boldsymbol{e}_2 + \cdots + \lambda_r \boldsymbol{e}_r,$$

为求其中的系数 $\lambda_i (i = 1, \cdots, r)$，可以计算 $\boldsymbol{e}_i$ 与 $\boldsymbol{a}$ 的内积，有

$$\langle \boldsymbol{e}_i, \boldsymbol{a} \rangle = \langle \boldsymbol{e}_i, \lambda_1 \boldsymbol{e}_1 + \lambda_2 \boldsymbol{e}_2 + \cdots + \lambda_r \boldsymbol{e}_r \rangle = \lambda_1 \langle \boldsymbol{e}_i, \boldsymbol{e}_1 \rangle + \lambda_2 \langle \boldsymbol{e}_i, \boldsymbol{e}_2 \rangle + \cdots + \lambda_r \langle \boldsymbol{e}_i, \boldsymbol{e}_r \rangle = \lambda_i$$

即

$$\lambda_i = \langle \boldsymbol{a}, \boldsymbol{e}_i \rangle$$

利用这个公式能方便地求得向量的坐标。

因此，我们给向量空间取基时常常取标准正交基。

接下来我们应用投影的思想，确定 $\mathsf{Col}\,(\boldsymbol{A})$ 中的一组标准正交基。

投影与基的正交化

设 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_r$ 是向量空间 $\mathbb{V}$ 的一个基：我们的目的是找到一组正交基 $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_r$ 使得

$$\mathbf{span}\{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_r\} = \mathbf{span}\{\boldsymbol{a}_1, \cdots, \boldsymbol{a}_r\}$$

我们可以这样做，我们先取 $\boldsymbol{a}_1$ 作为一个基，记为 $\boldsymbol{b}_1$。那么 $\boldsymbol{a}_2$ 可以正交分解

$$\boldsymbol{a}_2 = \boldsymbol{a}_2^{(1)} + \boldsymbol{a}_2^{(2)},$$

其中 $\boldsymbol{a}_2^{(1)} \in \mathsf{Col}\left((\boldsymbol{a}_1)\right), \boldsymbol{a}_2^{(2)} \in \mathsf{Null}\left((\boldsymbol{a}_1)^{\mathrm{T}}\right)$。利用投影公式：

$$\boldsymbol{a}_2^{(1)} = \frac{\langle \boldsymbol{b}_1, \boldsymbol{a}_2 \rangle}{\langle \boldsymbol{b}_1, \boldsymbol{b}_1 \rangle} \boldsymbol{b}_1$$

$$\boldsymbol{a}_2^{(2)} = \boldsymbol{a}_2 - \frac{\langle \boldsymbol{b}_1, \boldsymbol{a}_2 \rangle}{\langle \boldsymbol{b}_1, \boldsymbol{b}_1 \rangle} \boldsymbol{b}_1$$

我们记 $\boldsymbol{a}_2^{(2)}$ 为 $\boldsymbol{b}_2$。并把 $\boldsymbol{b}_2$ 添加到正交基中，$\mathbf{span}\{\boldsymbol{a}_1, \boldsymbol{a}_2\} = \mathbf{span}\{\boldsymbol{b}_1, \boldsymbol{b}_2\}$。注意这里 $\boldsymbol{b}_1, \boldsymbol{b}_2$ 还不是标准正交基。

假设我们已经有了一组有序正交基底 $\boldsymbol{B}_k = (\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_k)$，那么 $\boldsymbol{a}_{k+1}$ 可以正交分解

$$\boldsymbol{a}_{k+1} = \boldsymbol{a}_{k+1}^{(1)} + \boldsymbol{a}_{k+1}^{(2)},$$

其中 $\boldsymbol{a}_{k+1}^{(1)} \in \mathsf{Col}\left(\boldsymbol{B}_k\right), \boldsymbol{a}_{k+1}^{(2)} \in \mathsf{Null}\left(\boldsymbol{B}_k^{\mathrm{T}}\right)$。利用投影公式：

$$\begin{aligned}\boldsymbol{a}_{k+1}^{(1)} &= \pi_{\mathsf{Col}(\boldsymbol{B}_k)}(\boldsymbol{a}_{k+1}) = \boldsymbol{B}_k(\boldsymbol{B}_k^{\mathrm{T}}\boldsymbol{B}_k)^{-1}\boldsymbol{B}_k^{\mathrm{T}}\boldsymbol{a}_{k+1} \\ &= \left(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_k\right)\begin{pmatrix}\langle \boldsymbol{b}_1, \boldsymbol{b}_1\rangle & \cdots & \langle \boldsymbol{b}_1, \boldsymbol{b}_k\rangle \\ \cdots & \ddots & \cdots \\ \langle \boldsymbol{b}_k, \boldsymbol{b}_1\rangle & \cdots & \langle \boldsymbol{b}_k, \boldsymbol{b}_k\rangle\end{pmatrix}^{-1}\begin{pmatrix}\langle \boldsymbol{b}_1, \boldsymbol{a}_{k+1}\rangle \\ \vdots \\ \langle \boldsymbol{b}_k, \boldsymbol{a}_{k+1}\rangle\end{pmatrix}\end{aligned}$$

而 $\{\boldsymbol{b}_1, \cdots, \boldsymbol{b}_k\}$ 是相互正交的。也就是说若 $i \neq j$，$\langle \boldsymbol{b}_i, \boldsymbol{b}_j\rangle = 0$。

所以

$$
\begin{aligned}
\boldsymbol{a}_{k+1}^{(1)} &= (\boldsymbol{b}_1, \cdots, \boldsymbol{b}_k) \begin{pmatrix} \langle \boldsymbol{b}_1, \boldsymbol{b}_1 \rangle & & \\ & \ddots & \\ & & \langle \boldsymbol{b}_k, \boldsymbol{b}_k \rangle \end{pmatrix}^{-1} \begin{pmatrix} \langle \boldsymbol{b}_1, \boldsymbol{a}_{k+1} \rangle \\ \vdots \\ \langle \boldsymbol{b}_k, \boldsymbol{a}_{k+1} \rangle \end{pmatrix} \\
&= \frac{\langle \boldsymbol{b}_1, \boldsymbol{a}_{k+1} \rangle}{\langle \boldsymbol{b}_1, \boldsymbol{b}_1 \rangle} \boldsymbol{b}_1 + \frac{\langle \boldsymbol{b}_2, \boldsymbol{a}_{k+1} \rangle}{\langle \boldsymbol{b}_2, \boldsymbol{b}_2 \rangle} \boldsymbol{b}_2 + \cdots + \frac{\langle \boldsymbol{b}_k, \boldsymbol{a}_{k+1} \rangle}{\langle \boldsymbol{b}_k, \boldsymbol{b}_k \rangle} \boldsymbol{b}_k
\end{aligned}
$$

而 $\boldsymbol{a}_{k+1}^{(2)} = \boldsymbol{a}_{k+1} - \boldsymbol{a}_{k+1}^{(1)}$，我们记 $\boldsymbol{a}_{k+1}^{(2)}$ 为 $\boldsymbol{b}_{k+1}$，并把 $\boldsymbol{b}_{k+1}$ 添加到正交基中，$\mathbf{span}\{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_{k+1}\} = \mathbf{span}\{\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_{k+1}\}$。
以此类推，我们可以得到 $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_r\}$ 使得

$$
\mathbf{span}\{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_r\} = \mathbf{span}\{\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_r\}.
$$

只需要再把这组基单位化即可。

7.4.2 Gram-Schmidt 正交化

总结之前的过程，可以通过以下方法求得 $\mathbb{V}$ 的一个规范正交基 $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_r$。这种方法称为 Gram-Schmidt 正交化。

取

$$\boldsymbol{b}_1 = \boldsymbol{a}_1;$$

$$\boldsymbol{b}_2 = \boldsymbol{a}_2 - \frac{\langle \boldsymbol{b}_1, \boldsymbol{a}_2 \rangle}{\langle \boldsymbol{b}_1, \boldsymbol{b}_1 \rangle} \boldsymbol{b}_1;$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$\boldsymbol{b}_r = \boldsymbol{a}_r - \frac{\langle \boldsymbol{b}_1, \boldsymbol{a}_r \rangle}{\langle \boldsymbol{b}_1, \boldsymbol{b}_1 \rangle} \boldsymbol{b}_1 - \frac{\langle \boldsymbol{b}_2, \boldsymbol{a}_r \rangle}{\langle \boldsymbol{b}_2, \boldsymbol{b}_2 \rangle} \boldsymbol{b}_2 - \cdots - \frac{\langle \boldsymbol{b}_{r-1}, \boldsymbol{a}_r \rangle}{\langle \boldsymbol{b}_{r-1}, \boldsymbol{b}_{r-1} \rangle} \boldsymbol{b}_r$$

然后把它们单位化，取

$$\boldsymbol{e}_1 = \frac{1}{\|\boldsymbol{b}_1\|} \boldsymbol{b}_1, \boldsymbol{e}_2 = \frac{1}{\|\boldsymbol{b}_2\|} \boldsymbol{b}_2, \cdots, \boldsymbol{e}_r = \frac{1}{\|\boldsymbol{b}_r\|} \boldsymbol{b}_r$$

就是 $\mathbb{V}$ 的一个规范正交基。

例 6

求向量组 $\boldsymbol{a}_1=(3,1,1)^T, \boldsymbol{a}_2=(2,2,0)^T$ 的生成子空间的标准正交基。

取

$$\boldsymbol{b}_1=(3,1,1)^T$$

$$\boldsymbol{b}_2=\boldsymbol{a}_2-\frac{\boldsymbol{b}_1^{\mathrm{T}}\boldsymbol{a}_2}{\boldsymbol{b}_1^{\mathrm{T}}\boldsymbol{b}_1}\boldsymbol{b}_1=(2,2,0)^T-\frac{8}{11}(3,1,1)^T=\frac{-2}{11}(1,-7,4)^T$$

$$\boldsymbol{e}_1=\frac{1}{\sqrt{11}}(3,1,1)^T$$

$$\boldsymbol{e}_2=\frac{1}{\sqrt{66}}(1,-7,4)^T$$

故标准正交基为 $\boldsymbol{e}_1, \boldsymbol{e}_2$：即，

$$\frac{1}{\sqrt{11}}(3,1,1)^T, \frac{1}{\sqrt{66}}(1,-7,4)^T$$

本讲小结

四个基本子空间

- 行空间
- 列空间
- 零空间
- 左零空间
- 正交关系

正交投影

- 1 维子空间
- 一般子空间
- 仿射子空间
- 正交基与 Gram-Schmit

正交和投影是基础性概念，与超定系统的最小二乘解，并与机器学习中的降维、分类或回归都有紧密联系，我们在后续章节中进一步给出。