2022 春高等代数习题课

2022-5-20 第六次习题课

第一题. 9.2 例 3（最小二乘法）求直线 $y = kx + b$ 尽可能接近已知数据点 (x_i, y_i) $(1 \leqslant i \leqslant n)$，也就是使 $d(k,b) = \sum_{i=1}^{n}(kx_i + b - y_i)^2$ 达到最小值。

解. 我们首先从分析的角度来看这个问题。$d(k,b) = \sum_{i=1}^{n}(kx_i + b - y_i)^2$ 关于 k, b 都是（下）凸的，所以满足 $\operatorname{grad} d(k,b) = (0,0)^T$ 的点都是使 $d(k,b)$ 达到最小值的点。计算梯度得

$$\operatorname{grad} d(k,b) = \begin{pmatrix} 2\sum_{i=1}^{n} x_i(kx_i + b - y_i) \\ 2\sum_{i=1}^{n}(kx_i + b - y_i) \end{pmatrix}$$

得方程组

$$\begin{cases} k\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i y_i = 0 \\ k\sum_{i=1}^{n} x_i + nb - \sum_{i=1}^{n} y_i = 0 \end{cases}$$

或者写成矩阵的形式

$$M\begin{pmatrix} k \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n \end{pmatrix}\begin{pmatrix} k \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} y_i \end{pmatrix}$$

只要满足 x_i 不全相等，系数矩阵就非奇异，(k,b) 有唯一解

$$\begin{pmatrix} k \\ b \end{pmatrix} = M^{-1}\begin{pmatrix} \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} y_i \end{pmatrix}$$

令 $A = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}$, $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$，那么 $M = A^T A$，上述解又可以写为

$$(A^T A)^{-1} A^T \mathbf{y}.$$

下面，我们从代数的角度来解这个问题。在最理想的情况下，所有数据点都满足某个线性方程 $y = kx + b$, 写成矩阵的形式即为

$$\begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} k \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

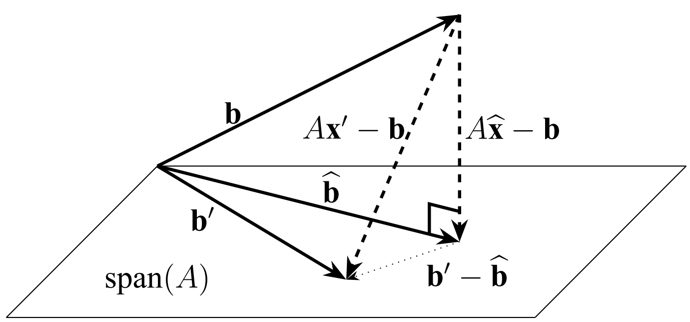
一般情况下，满足上式的 (k, b) 不存在。那么就退而求其次，找一个近似解（最小二乘解）。我们考虑更一般的形式 $A\mathbf{x} = \mathbf{b}$. 例如在这里，$A = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} k \\ b \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$.

我们知道 $A\mathbf{x} = \mathbf{b}$ 有解 $\Longleftrightarrow$ $\mathbf{b} \in \operatorname{span}(A)$（$A$ 的列空间）。当它无解时，$\mathbf{b} \notin \operatorname{span}(A)$. 最小二乘解 $\widehat{\mathbf{x}}$ 是使得 $|A\widehat{\mathbf{x}} - \mathbf{b}|$（向量长度一般用 $\|\cdot\|$ 表示，这里遵从课本的记号，用 $|\cdot|$ 表示）最小的向量。令 $\widehat{\mathbf{b}} = A\widehat{\mathbf{x}} \in \operatorname{span}(A)$, 那么

$$\widehat{\mathbf{b}} = \underset{\mathbf{b}' \in \operatorname{span}(A)}{\arg\min} |\mathbf{b}' - \mathbf{b}|$$

$\mathbf{b}' - \mathbf{b}$ 就是下图虚线（dashed line）代表的向量，它最短当且仅当 $\mathbf{b}'$ 是 $\mathbf{b}$ 到 $\operatorname{span}(A)$ 的投影（之后习题会证明），即 $\widehat{\mathbf{b}} - \mathbf{b} = A\widehat{\mathbf{x}} - \mathbf{b} \in \operatorname{span}(A)^{\perp}$, 这等价于正则方程

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



当 A 列满秩时，在这题里就是 x_i 不全相等时，A^TA 是可逆方阵，上述正则方程有唯一解

$$\widehat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}.$$

当 A 列不满秩时，最小二乘解不唯一，但长度最小的最小二乘解（被称作最优最小二乘解）是唯一的。这部分参阅补充内容，参见slides-pseudoinverse-least-squares.pdf。

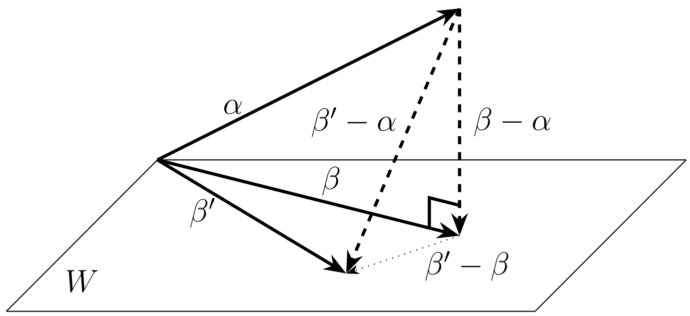
第二题. 习题 9.2 第 8 题. 用向量的内积证明平面外一点到平面的线段长以垂线段最短，并推广到 n 维欧氏空间：

(1) 取平面上任一点为原点 O, 将空间 V 每一点 P 用向量 $\overrightarrow{OP}$ 表示，则平面是一个 2 维子空间 W. 设 A 是空间中给定的任一点，B 是平面内任一点，分别对应于向量 $\alpha = \overrightarrow{OA}, \beta = \overrightarrow{OB}$, 则 $|AB| = |\alpha - \beta|$. 求证：当 $\alpha - \beta \in W^{\perp}$ 时 $|\alpha - \beta|$ 取最小值；

(2) 设 $E(\mathbb{R})$ 是欧氏空间，W 是它的任意子空间，α 是 $E(\mathbb{R})$ 任意给定的向量。求证：当 $\alpha - \beta \in W^{\perp}$ 时，$|\alpha - \beta|$ $(\beta \in W)$ 取最小值；

(3) 设 $E(\mathbb{R})$ 是欧氏空间，$\alpha \in E(\mathbb{R})$, W 是由 $\alpha_1, \ldots, \alpha_k \in E(\mathbb{R})$ 生成的子空间。当 $x_1, \ldots, x_k \in \mathbb{R}$ 满足什么条件时，$|\alpha - (x_1\alpha_1 + \cdots + x_k\alpha_k)|$ 取最小值？

证明：我们还用一下上一题的图，从而有一个比较直观的认识：我们设 $\beta\in W$ 是使得向量 $\beta-\alpha$ 垂直于 W 的向量，$\beta'\in W$ 是空间 W 中任意的另外一个向量，那么 $\beta'-\alpha,\beta-\alpha,\beta'-\beta$ 组成了一个直角三角形，以 $\beta'-\alpha$ 为斜边，垂线段 $\beta-\alpha$ 是一条直角边，必然比斜边短。



下面我们用向量的内积来证明这题的结论。以下，我们用 $\langle\cdot,\cdot\rangle$ 来表示向量内积。

(1) 假设向量 $\beta\in W$ 使得 $\alpha-\beta\in W^{\perp}$, $\beta'\in W$ 是 W 中任意一个向量。那么

$$\begin{aligned}|\alpha-\beta'|^2&=|(\alpha-\beta)+(\beta-\beta')|^2\\&=\langle(\alpha-\beta)+(\beta-\beta'),(\alpha-\beta)+(\beta-\beta')\rangle\\&=|\alpha-\beta|^2+2\langle\alpha-\beta,\beta-\beta'\rangle+|\beta-\beta'|^2\\&=|\alpha-\beta|^2+|\beta-\beta'|^2\\&\geqslant|\alpha-\beta|^2\end{aligned}$$

上式中的 $\geqslant$ 是相等当且仅当 $\beta=\beta'$. 这就证明了当 $\alpha-\beta\in W^{\perp}$ 时 $|\alpha-\beta|$ 取最小值。

(2) 和第 (1) 问的证明相同，因为 (1) 的证明与 W 的维数是无关的。

(3) 由第 (2) 问知需要满足 $\alpha-(x_1\alpha_1+\cdots+x_k\alpha_k)\in W^{\perp}$. 由于 W 是由 $\alpha_1,\ldots,\alpha_k\in E(\mathbb{R})$ 生成的，这又等价于 $\forall 1\leqslant i\leqslant k$, 有

$$0=\langle\alpha_i,\alpha-(x_1\alpha_1+\cdots+x_k\alpha_k)\rangle,$$

即

$$\sum_{j=1}^{k}\langle\alpha_i,\alpha_j\rangle x_j=\langle\alpha_i,\alpha\rangle,$$

写成矩阵形式即

$$\begin{pmatrix}\langle\alpha_1,\alpha_1\rangle&\cdots&\langle\alpha_1,\alpha_k\rangle\\\vdots&&\vdots\\\langle\alpha_k,\alpha_1\rangle&\cdots&\langle\alpha_k,\alpha_k\rangle\end{pmatrix}\begin{pmatrix}x_1\\\vdots\\x_k\end{pmatrix}=\begin{pmatrix}\langle\alpha_1,\alpha\rangle\\\vdots\\\langle\alpha_k,\alpha\rangle\end{pmatrix}$$

当 $x_1,\ldots,x_k\in\mathbb{R}$ 是上述线性方程组的解时，$|\alpha-(x_1\alpha_1+\cdots+x_k\alpha_k)|$ 取最小值。

注意，假设我们在 $E(\mathbb{R})$ 中取定了一组基，那么 $\alpha,\alpha_1,\ldots,\alpha_k$ 就有（列向量形式的）坐标表示，这些坐标（列向量）我们还用 $\alpha,\alpha_1,\ldots,\alpha_k$ 来记，并令矩阵 $A=(\alpha_1,\ldots,\alpha_k)$, 那么上述方程就可以写为

$$A^TA\begin{pmatrix}x_1\\\vdots\\x_k\end{pmatrix}=A^T\alpha.$$

第三题. 习题 9.2 第 9 题. (1) 设 W 是 $\mathbb{R}^3$ 中过点 $(0,0,0),(1,2,2),(3,4,0)$ 的平面，求点 $A(5,0,0)$ 到平面 W 的最短距离；

(2) 求方程组

$$\begin{cases}0.39x-1.89y=1\\0.61x-1.80y=1\\0.93x-1.68y=1\\1.35x-1.50y=1\end{cases}$$

的最小二乘解。

(3) 设 $A\in\mathbb{R}^{m\times n}, X=(x_1,\cdots,x_n)^T, \beta\in\mathbb{R}^{m\times 1}$. 如果实系数线性方程组 $AX=\beta$ 无解，我们可以求 X 使 $\mathbb{R}^{m\times 1}$ 中的向量 $\delta=AX-\beta$ 的长度 $|\delta|$ 取最小值。满足这个条件的解 X 称为方程组 $AX=\beta$ 的最小二乘解。设 A 的各列依次为 $\alpha_1,\ldots,\alpha_n$, 则 $AX=x_1\alpha_1+\cdots+x_n\alpha_n$. 求证：

$$|\delta|\text{取最小值}\iff(\delta,\alpha_i)=0(\forall 1\leqslant i\leqslant n)\iff A^TAX=A^T\beta$$

解：设 $O=(0,0,0), P_1=(1,2,2), P_2=(3,4,0)$, 那么 W 正好是过原点的由 $\alpha_1=\overrightarrow{OP_1}=(1,2,2), \alpha_2=\overrightarrow{OP_2}=(3,4,0)$ 张成的线性空间。设 $\beta'=\overrightarrow{OA'}$ 是 $\beta=\overrightarrow{OA}=(5,0,0)$ 在 W 上的投影，则 $|\overrightarrow{AA'}|$ 即是点 A 到平面 W 的最短距离。设 $\beta'=x_1\alpha_1+x_2\alpha_2\in W$, 那么根据上一题的结论，$(x_1,x_2)$ 需要是如下的线性方程组的解

$$\begin{pmatrix}\langle\alpha_1,\alpha_1\rangle & \langle\alpha_1,\alpha_2\rangle\\\langle\alpha_2,\alpha_1\rangle & \langle\alpha_2,\alpha_2\rangle\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}\langle\alpha_1,\beta\rangle\\\langle\alpha_2,\beta\rangle\end{pmatrix}$$

即

$$\begin{pmatrix}9 & 11\\11 & 25\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}5\\15\end{pmatrix}$$

解得

$$\begin{pmatrix}x_1\\x_2\end{pmatrix}=\frac{1}{104}\begin{pmatrix}25 & -11\\-11 & 9\end{pmatrix}\begin{pmatrix}5\\15\end{pmatrix}=\frac{1}{104}\begin{pmatrix}-40\\80\end{pmatrix}=\frac{5}{13}\begin{pmatrix}-1\\2\end{pmatrix}$$

最短距离即等于 $|\beta-(x_1\alpha_1+x_2\alpha_2)|=|(5,0,0)-\frac{5}{13}(5,6,-2)|=\frac{5}{13}|(8,-6,2)|=10\sqrt{\frac{2}{13}}$.

(2) 将问题写成矩阵的形式

$$A\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}0.39 & -1.89\\0.61 & -1.80\\0.93 & -1.68\\1.35 & -1.50\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}1\\1\\1\\1\end{pmatrix}=\beta$$

那么这个问题的最小二乘解为 $(A^TA)^{-1}A^T\beta$, 编程解如下

```
>>> import numpy as np

>>> A = np.array([[0.39,-1.89], [0.61,-1.80], [0.93,-1.68], [1.35,-1.50]])
>>> beta = np.array([1,1,1,1])

>>> sol = np.linalg.inv(A.T @ A) @ A.T @ beta
>>> sol
array([ 0.19722003, -0.48807896])
```

(3) 这问实际上我们在第一题中已经证明完毕。

第四题. 习题 9.3 第 6 题. 给定 $0 \neq \alpha \in E_n(\mathbb{R})$. 定义 $E_n(\mathbb{R})$ 中的线性变换 $\tau_\alpha : \beta \mapsto \beta - \dfrac{2\langle \beta, \alpha \rangle}{\langle \alpha, \alpha \rangle}\alpha$, 求证:

(1) τ_α 是正交变换;

(2) τ_α 在适当的标准正交基下的矩阵为 $\operatorname{diag}(-1, 1, \cdots, 1)$.

证明. (1). 要证 τ_α 是正交变换，我们只要证明任取 $\beta_1, \beta_2 \in E_n(\mathbb{R})$ 都有 $\langle \tau_\alpha(\beta_1), \tau_\alpha(\beta_2) \rangle = \langle \beta_1, \beta_2 \rangle$. 我们来验证这个结论。

$$
\begin{aligned}
&\langle \tau_\alpha(\beta_1), \tau_\alpha(\beta_2) \rangle \\
=&\langle \beta_1 - \frac{2\langle \beta_1, \alpha \rangle}{\langle \alpha, \alpha \rangle}\alpha, \beta_2 - \frac{2\langle \beta_2, \alpha \rangle}{\langle \alpha, \alpha \rangle}\alpha \rangle \\
=&\langle \beta_1, \beta_2 \rangle - \frac{2\langle \beta_2, \alpha \rangle\langle \beta_1, \alpha \rangle}{\langle \alpha, \alpha \rangle} - \frac{2\langle \beta_1, \alpha \rangle\langle \alpha, \beta_2 \rangle}{\langle \alpha, \alpha \rangle} + \frac{4\langle \beta_1, \alpha \rangle\langle \beta_2, \alpha \rangle\langle \alpha, \alpha \rangle}{\langle \alpha, \alpha \rangle^2} \\
=&\langle \beta_1, \beta_2 \rangle
\end{aligned}
$$

(2) 对于线性变换 $\tau_\alpha : \beta \mapsto \beta - \dfrac{2\langle \beta, \alpha \rangle}{\langle \alpha, \alpha \rangle}\alpha$, 我们容易发现如下结论

- $\tau_\alpha(\alpha) = \alpha - \dfrac{2\langle \alpha, \alpha \rangle}{\langle \alpha, \alpha \rangle}\alpha = -\alpha$.
- 任取 $\beta \in \operatorname{span}\{\alpha\}^\perp$, 有 $\tau_\alpha(\beta) = \beta$.

那么我们从向量 $\alpha_1 = \dfrac{\alpha}{|\alpha|}$ 出发，扩充成 $E_n(\mathbb{R})$ 的一组标准正交基 $\alpha_1, \alpha_2, \ldots, \alpha_n$, 那么

$$
\tau_\alpha(\alpha_1) = \tau_\alpha(\frac{\alpha}{|\alpha|}) = \frac{1}{|\alpha|}\tau_\alpha(\alpha) = \frac{-\alpha}{|\alpha|} = -\alpha_1
$$

此外，$\alpha_2, \ldots, \alpha_n \in \operatorname{span}\{\alpha\}^\perp$, 故有 $\tau_\alpha(\alpha_i) = \alpha_i, \forall 2 \leqslant i \leqslant n$. 那么在标准正交基 $\alpha_1, \alpha_2, \ldots, \alpha_n$ 下，τ_α 的矩阵为 $\operatorname{diag}(-1, 1, \cdots, 1)$。

这里我们容易看到，τ_α 实际上就是关于以 α 为法向量的超平面 (由一个线性方程确定的 $n-1$ 维子空间) 的对称变换。