

《线性代数》 期末报告 实矩阵的三个等价关系

Saturday 28th May, 2022

目录

I	线性代数的研究历史	3
1	线性代数简史	3
2	线性代数的基本应用	4
2.1	人工智能	4
2.1.1	推导	5
2.1.2	结论	7
II	实矩阵的三个等价关系	8
3	矩阵的等价	8
3.1	等价在生活、数学的运用	9
3.1.1	增广矩阵解线性方程组	9
3.2	实矩阵等价关系下的不变量	10
3.3	矩阵等价的充分必要条件	10
3.4	等价标准形化的方法	11
4	方阵的相似	11
4.1	实方阵相似关系下的不变量	11
5	方阵的可对角化	13
5.1	方阵可对角化的充分必要条件	14
5.2	对角化的方法	15
6	方阵的合同	15
6.1	实对称方阵合同关系下的不变量	16
6.2	实对称方阵合同的充分必要条件	16
6.3	标准形化的方法	16
III	本课程的学习体会	18

Part I

线性代数的研究历史

1 线性代数简史

[1] 线性代数的研究最初出现于对行列式的研究上。行列式当时被用来求解线性方程组。莱布尼茨在 1693 年使用行列式。随后，加布里尔·克拉默在 1750 年推导出求解线性方程组的克莱姆法则。然后，高斯利用高斯消元法发展出求解线性系统的理论。这也被列为大地测量学的一项进展。

现代线性代数的历史可以上溯到 19 世纪中期的英国。1843 年，哈密顿发现四元数。1844 年，赫尔曼·格拉斯曼发表他的著作《线性外代数》(Die lineare Ausdehnungslehre)，包括今日线性代数的一些主题。1848 年，詹姆斯·西尔维斯特引入矩阵 (matrix)，该词是“子宫”的拉丁语。阿瑟·凯莱在研究线性变换时引入矩阵乘法和转置的概念。很重要的是，凯莱使用一个字母来代表一个矩阵，因此将矩阵当做了聚合对象。他也意识到矩阵和行列式之间的联系。

不过除了这些早期的文献以外，线性代数主要是在二十世纪发展的。在抽象代数的环论开发之前，矩阵只有模糊不清的定义。随着狭义相对论的到来，很多开拓者发现线性代数的微妙。进一步的，解偏微分方程的克莱姆法则的例行应用导致大学的标准教育中包括线性代数。例如，E.T. Copson 写到：

“当我在 1922 年到爱丁堡做年轻的讲师的时候，我惊奇的发现不同于牛津的课程。这里包括我根本就不知道的主题如勒贝格积分、矩阵论、数值分析、黎曼几何...”

——E.T. Copson, 《偏微分方程》前言, 1973

1882 年，Hüseyin Tevfik Pasha (英语: Hüseyin Tevfik Pasha) (TR (土耳其语: Hüseyin Tevfik Paşa)) 写了一本书，名为《Linear Algebra》(线性代数)。第一次现代化精确定义向量空间是在 1888 年，由朱塞佩·皮亚诺提出。在 1888 年，弗兰西斯·高尔顿还发起相关系数的应用。经常有多于一个随机变量出现并且它们可以互相关。在多变元随机变量的统计分析中，相关矩阵是自然的工具。所以这种随机向量的统计研究帮助矩阵用途的开发。

到 1900 年，一种有限维向量空间的线性变换理论被提出。在 20 世纪上半叶，许多前几世纪的想法和方法被总结成抽象代数，线性代数第一次有了它的现代形式。矩阵在量子力学、狭义相对论和统计学上的应用帮助线性代数的主题超越纯数学的范畴。计算机的发展导致更多地研究致力于有关高斯消元法和矩阵分解的有效算法上。线性代数成为数字模拟和模型的基本工具。

2 线性代数的基本应用

2.1 人工智能

线性代数是 AI 专家必须掌握的知识，这已不再是个秘密。如果不掌握应用数学这个领域，你永远就只能是「门外汉」。当然，学习线性代数道阻且长。数学，尤其是线性代数常与枯燥、复杂和毫无意义的事物联系起来。不过你还可以另辟蹊径。

——Oleksii Kharkovyna, *TowardsDataScience*

人工智能的机器学习中构建神经网络大量运用了线性代数的内容。在最简单的模型中，输入输出的数据集被视为向量，而输入输出的线性变换则是矩阵。但真实情况往往更加复杂：例如图片除了高度宽度两个维度外，还有四个色彩通道（R、G、B、A），所以每张图片都是三维的数据，多张图片则变为了四维，我们称高维的“向量”为“张量”，是线性代数的广义拓展。同理输出和中间的线性变换也都是张量。

例如在吴恩达机器学习课程 [2] 里反向传播算法中，运用了线性代数的知识进行运算：

Definition 2.1.1 (对数似然代价函数).

$$J(\theta) = y \ln h_{\theta}(x) + (1 - y) \ln (1 - h_{\theta}(x))$$

Definition 2.1.2 (对数似然代价函数).

$$h_{\theta}(x) = \sum_i \theta_i x_i = \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Definition 2.1.3 (*Logistic* 激活函数及其导数).

$$g(x) = \frac{1}{1 + e^{-x}}$$

$$g'(x) = g(x) [1 - g(x)]$$

Corollary 2.1.0.1 (*Logistic* 激活函数及其导数).

$$g(x) = \frac{1}{1 + e^{-x}}$$

$$g'(x) = g(x) [1 - g(x)]$$

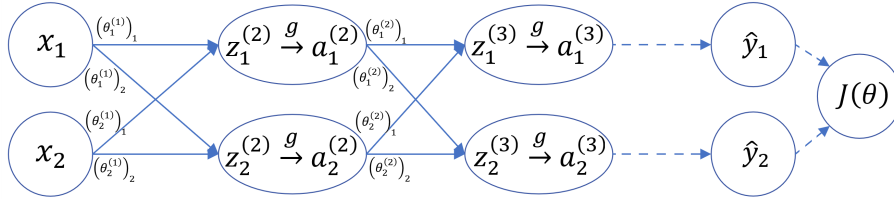


图 1: 反向传播符号含义图示

如图 1 (省略了偏置), 输入数据为 $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, 实际输出为 $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$

这张图上画出了所有运算, 例如:

$$a_1^{(2)} = g(z_1^{(2)})$$

$$z_2^{(2)} = (\theta_1^{(1)})_1 x_1 + (\theta_1^{(1)})_2 x_2$$

同时, 此图认为预测输出为 $\hat{y}_1 = a_1^{(3)}$, 即有误差 (注意此处不是定义而是结论):

$$\delta_1^{(3)} = \hat{y}_1 - y_1 = a_1^{(3)} - y_1$$

2.1.1 推导

下面我们将上列函数改写成对应元素的写法, 先作定义:

Definition 2.1.4 (字母定义). 字母 L, m, n

L : 被 θ 作用的层

m : L 层单元数量, 用 j 进行遍历 (即 $j \in \{1, 2, \dots, m\}$)

n : L 层单元数量, 用 i 进行遍历

综上可得, 若 L 是倒数第二层, 则有:

Corollary 2.1.0.2 (误差 δ 定义).

$$\begin{aligned}
 \delta_i^{(L+1)} &= \frac{\partial J}{\partial z_i^{(L+1)}} \\
 &= \frac{\partial J}{\partial a_i^{(L+1)}} \cdot \frac{\partial a_i^{(L+1)}}{\partial z_i^{(L+1)}} \\
 &= \left(\frac{-y_i}{a_i^{(L+1)}} + \frac{1-y_i}{1-a_i^{(L+1)}} \right) \cdot g' z_i^{(L+1)} \\
 &= \left(\frac{-y_i}{a_i^{(L+1)}} + \frac{1-y_i}{1-a_i^{(L+1)}} \right) \cdot a_i^{(L+1)} (1-a_i^{(L+1)}) \\
 &= a_i^{(L+1)} - y_i
 \end{aligned}$$

将同一层 $\delta^{(L+1)}$ 合并为矩阵得 δ, a, y 都是列向量):

$$\delta^{(L+1)} = a^{(L+1)} - y$$

下面推隐含层, 以第一个单元为例:

$$\begin{aligned}
 \delta_1^{(2)} &= \frac{\partial J}{\partial z_1^{(2)}} \\
 &= \frac{\partial J}{\partial z_1^{(3)}} \cdot \frac{\partial z_1^{(3)}}{\partial a_1^{(2)}} \cdot \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} + \frac{\partial J}{\partial z_2^{(3)}} \cdot \frac{\partial z_2^{(3)}}{\partial a_1^{(2)}} \cdot \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} \\
 &= \delta_1^{(3)} \cdot \left(\theta_1^{(2)} \right)_1 \cdot g' z_1^{(2)} + \delta_2^{(3)} \cdot \left(\theta_1^{(2)} \right)_2 \cdot g' z_1^{(2)}
 \end{aligned}$$

令:

$$\begin{cases} \delta^{(L)} = \begin{bmatrix} \delta_1^{(L)} \\ \delta_2^{(L)} \\ \vdots \\ \delta_n^{(L)} \end{bmatrix} \\ \theta_i^{(L)} = \left[\left(\theta_i^{(L)} \right)_1 \quad \left(\theta_i^{(L)} \right)_2 \quad \cdots \quad \left(\theta_i^{(L)} \right)_n \right] \end{cases}$$

可将上式化为矩阵：

$$\delta_1^{(2)} = \theta_1^{(2)} \delta^{(3)} \cdot g' z_1^{(2)}$$

2.1.2 结论

由上，可写出递推普式：

$$\delta_j^{(L)} = \theta_j^{(L)} \delta^{(L+1)} \cdot g' z_j^{(L)}$$

其中最后一层：

$$\delta^{(Last)} = a^{(Last)} - y$$

Part II

实矩阵的三个等价关系

3 矩阵的等价

Definition 3.0.1 (行等价、列等价和等价).

$$A, B \text{ 行等价} \iff PA = B$$

$$A, B \text{ 列等价} \iff AQ = B$$

$$A, B \text{ 等价} \iff PAQ = B$$

其中:

$$A = A_{m,n}$$

$$B = B_{m,n}$$

$$R(P_{m,m}) = m$$

$$R(Q_{n,n}) = n$$

Theorem 3.0.1 (反身性).

$$A, A \text{ 等价}$$

Proof.

$$EAE = A$$

$$\implies A, A \text{ 等价}$$

□

Theorem 3.0.2 (对称性).

$$A, B \text{ 等价} \implies B, A \text{ 等价}$$

Proof.

$$R(P) = m, R(Q) = n$$

$$\implies \exists P^{-1}, Q^{-1}$$

$$\implies P^{-1}BQ^{-1} = A$$

$$\implies B, A \text{ 等价}$$

□

Theorem 3.0.3 (传递性).

$$A, B \text{ 等价}, B, C \text{ 等价} \implies A, C \text{ 等价}$$

Proof.

$$\begin{aligned} & A, B \text{ 等价}, B, C \text{ 等价} \\ \implies & P_1 A Q_1 = B, P_2 B Q_2 = C \\ \implies & P_2^{-1} (P_1^{-1} A Q_1^{-1}) Q_2^{-1} = C \\ \implies & (P_2^{-1} P_1^{-1}) A (Q_1^{-1} Q_2^{-1}) = C \\ \implies & A, C \text{ 等价} \end{aligned}$$

□

3.1 等价在生活、数学的运用

3.1.1 增广矩阵解线性方程组

若有线性方程组：

$$\begin{cases} \sum_i a_{1i} x_i = b_1 \\ \sum_i a_{2i} x_i = b_2 \\ \vdots \\ \sum_i a_{mi} x_i = b_m \end{cases}$$

则对应增广矩阵

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix}$$

经过任意初等行变换之后，仍可以表示为原方程组。

3.2 实矩阵等价关系下的不变量

Theorem 3.2.1 (等价矩阵秩相同).

$$A, B \text{ 等价} \implies R(A) = R(B)$$

3.3 矩阵等价的充分必要条件

Theorem 3.3.1 (矩阵等价的充要条件是同秩).

$$A, B \text{ 等价} \iff R(A) = R(B)$$

Proof. 在 A, B 同型前提下:

将 A, B 化为标准型:

$$\begin{cases} A = P_1 D_1 Q_1 = \begin{bmatrix} E_A & O \\ O & O \end{bmatrix} \\ B = P_2 D_2 Q_2 = \begin{bmatrix} E_B & O \\ O & O \end{bmatrix} \end{cases}$$

充分性: A, B 等价 $\iff R(A) = R(B)$

$$R(A) = R(B) = k$$

$$\implies D_1 = D_2$$

$$\text{取 } P = P_2 P_1^{-1}, Q = Q_1^{-1} Q_2$$

$$\implies PAQ = B$$

$$\implies A, B \text{ 等价}$$

必要性: A, B 等价 $\implies R(A) = R(B)$

$$A, B \text{ 等价}$$

$$\implies PAQ = B$$

$$\implies P(P_1 D_1 Q_1) Q = B$$

$$\implies (PP_1) D_1 (Q_1 Q) = B$$

$$\implies R(A) = R(D_1) = R(B) (= R(D_2))$$

□

3.4 等价标准形化的方法

Theorem 3.4.1 (初等变换). 矩阵初等变换后与之前等价

行: r_i , 列: c_i

1. 对换两行 (列): $r_i \leftrightarrow r_j$
2. k 乘某行 (列): kr_i 或 $r_i \times k (k \neq 0)$
3. 加某行 (列) k 倍: $r_i + kr_j$

4 方阵的相似

Definition 4.0.1 (相似).

$$A \sim B \iff P^{-1}AP = B$$

Theorem 4.0.1 (反身性).

$$A \sim A$$

Theorem 4.0.2 (对称性).

$$A \sim B \implies B \sim A$$

Theorem 4.0.3 (传递性).

$$A \sim B, B \sim C \implies A \sim C$$

4.1 实方阵相似关系下的不变量

Theorem 4.1.1 (相似方阵特征多项式相同).

$$A \sim B \implies \det(A - \lambda_A E) = \det(B - \lambda_B E)$$

Corollary 4.1.1.1 (相似方阵特征值相同).

$$A \sim B \implies \lambda_A = \lambda_B$$

Corollary 4.1.1.2 (相似方阵迹相同).

$$A \sim B \implies \operatorname{tr} A = \operatorname{tr} B$$

Proof.

$$\begin{aligned}
A &\sim B \\
\implies P^{-1}AP &= B \\
\implies \det(B - \lambda E) \\
&= \det(P^{-1}AP - \lambda P^{-1}EP) \\
&= \det(P^{-1}(A - \lambda E)P) \\
&= \det P^{-1} \cdot \det(A - \lambda E) \cdot \det P \\
&= \det(A - \lambda E)
\end{aligned}$$

□

Theorem 4.1.2 (相似方阵行列式相同).

$$A \sim B \implies \det A = \det B$$

Proof.

$$\begin{aligned}
A &\sim B \\
\implies P^{-1}AP &= B \\
\implies \det(P^{-1}AP) &= \det B \\
\implies \det P^{-1} \cdot \det A \cdot \det P &= \det B \\
\implies \det A &= \det B \\
\implies \lambda A &= \lambda B \\
\operatorname{tr} A &= \operatorname{tr} B
\end{aligned}$$

□

Theorem 4.1.3 (相似方阵等价).

$$A \sim B \implies A, B \text{ 等价}$$

Corollary 4.1.3.1 (相似方阵秩相同).

$$A \sim B \implies R(A) = R(B)$$

Proof.

$$\begin{aligned}
 & A \sim B \\
 \implies & P^{-1}AP = B \\
 & P, P^{-1} \text{满秩} \\
 \implies & A, B \text{等价} \\
 \implies & R(A) = R(B)
 \end{aligned}$$

□

Corollary 4.1.3.2 (相似方阵逆也相似 (如果有逆)).

$$A \sim B, \exists A^{-1}, B^{-1} \implies A^{-1} \sim B^{-1}$$

Proof.

$$\begin{aligned}
 & \exists A^{-1}, B^{-1} \\
 \implies & B^{-1} \\
 & = (P^{-1}AP)^{-1} \\
 & = PA^{-1}P^{-1} \\
 \implies & A^{-1} \sim B^{-1}
 \end{aligned}$$

□

5 方阵的可对角化

Definition 5.0.1 (可对角化).

$$A_n \sim \Lambda$$

Corollary 5.0.0.1 (对角矩阵 Λ 和特征向量 \vec{p} 、矩阵 P).

$$A \sim \Lambda \implies P^{-1}AP = \Lambda \left\{ \begin{array}{l} \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \\ P = \begin{bmatrix} \vec{p}_1 & \vec{p}_2 & \cdots & \vec{p}_n \end{bmatrix} \end{array} \right.$$

5.1 方阵可对角化的充分必要条件

Theorem 5.1.1 (方阵可对角化充要条件是 A 有 n 个线性无关特征向量).

$$A \sim \Lambda \iff \text{全体线性无关}\vec{p}\text{数} = n$$

若设 $r = \lambda$ 重数,

又 $R(A - \lambda E) = \lambda$ 对应线性无关 \vec{p} 数,

即:

$$A \sim \Lambda \iff R(A - \lambda E) = n - r$$

Proof. [4] 由5定义

充分性: $A \sim \Lambda \iff \text{全体线性无关}\vec{p}\text{数} = n$

$$\begin{aligned} AP &= \begin{bmatrix} A\vec{p}_1 & A\vec{p}_2 & \cdots & A\vec{p}_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1\vec{p}_1 & \lambda_2\vec{p}_2 & \cdots & \lambda_n\vec{p}_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \vec{p}_1 & \vec{p}_2 & \cdots & \vec{p}_n \end{bmatrix} \\ &= \Lambda P \\ &\text{又 } \begin{bmatrix} \vec{p}_1 & \vec{p}_2 & \cdots & \vec{p}_n \end{bmatrix} \text{ 线性无关} \\ &\implies \exists P^{-1} \\ &\implies P^{-1}AP = \Lambda \\ &\implies A \sim \Lambda \end{aligned}$$

必要性: $A \sim \Lambda \implies$ 全体线性无关 \vec{p} 数 $= n$

$$A \sim \Lambda \implies AP = \Lambda P$$

$$\begin{aligned} \implies AP &= \begin{bmatrix} A\vec{p}_1 & A\vec{p}_2 & \cdots & A\vec{p}_n \end{bmatrix} \\ &= \Lambda P = \begin{bmatrix} \lambda_1 \vec{p}_1 & \lambda_2 \vec{p}_2 & \cdots & \lambda_n \vec{p}_n \end{bmatrix} \end{aligned}$$

$$\text{又} \exists P^{-1}$$

$$\begin{aligned} \implies \begin{bmatrix} \vec{p}_1 & \vec{p}_2 & \cdots & \vec{p}_n \end{bmatrix} &\text{线性无关且为特征向量} \\ \begin{bmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_n \end{bmatrix} &\text{是} A \text{的特征值} \end{aligned}$$

□

5.2 对角化的方法

Theorem 5.2.1 (特征多项式). 令特征多项式值为 0, 解得 λ 值即为特征值, 根重数为对应特征值重数

$$\det(A - \lambda E) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} - \lambda \end{vmatrix}$$

6 方阵的合同

Definition 6.0.1 (合同).

$$A, B \text{ 合同} \iff B = C^T A C, \exists C^{-1}$$

Theorem 6.0.1 (反身性).

$$A, A \text{ 合同}$$

Theorem 6.0.2 (对称性).

$$A, B \text{ 合同} \implies B, A \text{ 合同}$$

Theorem 6.0.3 (传递性).

$$A, B \text{ 合同}, B, C \text{ 合同} \implies A, C \text{ 合同}$$

6.1 实对称方阵合同关系下的不变量

Theorem 6.1.1 (合同方阵都是对称矩阵).

$$A, B \text{ 合同} \implies B = B^T$$

Corollary 6.1.1.1 (合同方阵秩相同).

$$A, B \text{ 合同} \implies R(A) = R(B)$$

Proof.

$$\begin{aligned} A, B \text{ 合同} &\implies B = C^T A C \\ &\implies B^T = (C^T A C)^T \\ &= C^T A^T C \\ &= B \\ &\text{又} \exists C^{-1} \\ &\implies A, B \text{ 等价} \\ &\implies R(A) = R(B) \end{aligned}$$

□

Theorem 6.1.2 (合同方阵惯性指数相同).

$$A, B, C \text{ 合同} \implies A, B \text{ 正负惯性指数相同}$$

6.2 实对称方阵合同的充分必要条件

Theorem 6.2.1.

$$A, B \text{ 合同} \iff \text{实对称方阵 } A, B \text{ 正负惯性指数相同}$$

6.3 标准形化的方法

Theorem 6.3.1 (正交变换法).

$$\text{二次型 } A \rightarrow \text{标准型 } B$$

1. 令 $\det(A - \lambda E) = 0$, 解得 n 个特征值 $\{\lambda_n\}$
2. 令 $(A - \lambda_i E) \vec{p} = \vec{0}$, 解得线性无关特征向量组 $\{\vec{p}_n\}$

3. 用格拉姆-施密特正交单位化, 解得正交单位特征向量组 $\{\vec{e}_n\}$
4. 用正交单位特征向量组构建正交矩阵 $P = \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \cdots & \vec{e}_n \end{bmatrix}$
5. $\vec{x} = P\vec{y}$ 即可将 A 化为标准型 B

Part III

本课程的学习体会

在每次提出新概念的时候我都很难理解，只能靠后续学习或者多做题巩固。在学习行列式的时候，通过平行四边形、平行六面体等概念理解了一些。但后续更多知识点也许需要更高阶的知识才能感性理解，让我意识到总是寻求感性理解是不现实的。尤其工科更重视运用的学科更难刨根问底。

我学习的时候还养成了看完定理喜欢阅读证明的习惯，虽然可以增强理解，但会很耗精力。尤其是有些定理的良好定义也许要到更高阶的学习后才能理解。一味寻求理解并不现实，还是要循序渐进才行。

在学习过程中经常遇到的问题是初等教育和大学、大学学科之间数学符号冲突让我很头疼，希望能有统一的标准，方便查阅使用。还有又许多常用的定理会以例题的形式出现在课本中，查询十分麻烦，如果可以将定理、证明和例题分列几处就更好了。

参考文献

- [1] Wikipedia. Linear Algebra: History.
- [2] [美] 吴恩达. Machine Learning.
- [3] [美] Sheldon Axler. Linear Algebra Done Right [M]. 北京: 人民邮电出版社, 1996.
- [4] [美] David C.Lay. Linear Algebra and Its Applications [M]. 北京: 机械工业出版社, 2005.
- [5] 唐烁, 朱士信. 线性代数 [M]. 北京: 机械工业出版社, 2018.

A 附录 1

<https://github.com/Poker-sang/Notes>

<https://www.bilibili.com/read/cv16091164>