

Multiplication among matrices, vectors

Four ways of understanding matrix multiplication

A, B and C are $m \times n$, $n \times p$ and $m \times p$ respectively. The following four ways of treating matrix multiplication are useful different scenarios.

Inner product

$$C = AB \iff C = \{c_{ij}\} = \{a_i^T b_j\} = \left\{ \sum_{k=1}^n a_{ik} b_{kj} \right\}$$

Outer product

$$C = AB \iff C = \sum_{i=1}^n a_i b_i^T,$$

where a_i is column vector of A , and b_i^T is row vector of B . The outer product of two vectors has rank one. b_i^T **is not the transpose of column vector b_i of B .**

Column picture

$$\begin{aligned} Ax &= x_1 A_{:,1} + x_2 A_{:,2} + x_3 A_{:,3} + \dots \\ AB &= \begin{bmatrix} AB_{:,1} & AB_{:,2} & AB_{:,3} + \dots \end{bmatrix}, \end{aligned}$$

where $A_{:,i}, B_{:,i}$ are column vectors of matrices A, B .

Row picture

$$\begin{aligned} x^T A &= x_1 A_{1,:} + x_2 A_{2,:} + x_3 A_{3,:} + \dots \\ AB &= \begin{bmatrix} A_{1,:} B \\ A_{2,:} B \\ A_{3,:} B \\ \dots \end{bmatrix}, \end{aligned}$$

where $A_{i,:}$ is row vector.

Examples

- $Ax = b$. Linear combinations of columns of A with coefficients x gives b .
- $x^T A = b^T$. Linear combinations of rows of A with coefficients x gives a row b^T ;
- $AB = C$ can be taken as right multiplying B to A to linear combine the columns of A . Each column of B is a set of coefficients and will combine columns of A to give a column in C .
- $AB = C$ can also be taken as left multiplying A to B to linear combine the rows of B . Each row of A is a set of coefficients and will combine rows of B to give a row in C .

Multiplications of matrix and vectors

A is a $m \times n$ matrix.

$$z = Ax \iff z_i = \sum_{k=1}^n a_{ik} x_k$$

$$z^T = y^T A \iff z_i = \sum_{k=1}^m y_k A_{ki}$$

$$\alpha = y^T Ax \iff \alpha = \sum_{i=1}^m \sum_{j=1}^n y_i A_{ij} x_j$$

$$C = AB \iff c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$$

- From left to right, keep the order of symbols in the LHS and let the summing indices adjacent. Use the dimension of either matrix or vector to determine the limit of summing index.
- From right to left, rearrange the identical indices adjacent if they are not. If the vector index is in the left, need transpose sign. Another way is not rearranging, but using transpose. For example,

$$\sum_{i=1}^n a_j^{(i)} b_k^{(i)} = \sum_{i=1}^n a_{ij} b_{ik} = \sum_{i=1}^n (a^T)_{ji} b_{ik} = a^T b$$

Basic matrix factorization / decomposition

Matrix factorization / decomposition

LU decomposition

- LU decomposition or factorization is done from ROW operation, or called elimination operations. The elimination operation is by left multiplying a matrix E . The matrix E can also be written as the multiplication of a series of elementary matrices. This process can be written as:

$$E(A, I) = (EA, EI) = (U, E)$$

where U is a upper triangle matrix. All the row operations (eliminations) to A is recorded in the E matrix through the operations on I . Here the E arises from many elementary matrix multiplication, and thus we need to apply these operations on I to record them. From above, we have,

$$EA = U \implies A = E^{-1}U \equiv LU$$

- In the above LU decomposition, L does not stand for 'left', but 'lower'.
- The time complexity of LU factorization is $O(n^3)$.
- LU decomposition can be used to calculate inverse, solve system of equation, determinant, etc. However, the time complexity is low.

$E^{-1}I$ decomposition / factorization

- This is a special name coined here. It is actually the essence of Gauss-Jordan way of calculating inverse matrix by elimination. So it is still row operation by left multiplying a matrix. It is similar to LU factorization except the target here is not an upper triangle matrix U , but an identity matrix I .

$$E(A, I) = (EA, EI) = (I, E)$$

Following the similar reasoning we have,

$$EA = I \implies A = E^{-1}I$$

- Here the Gauss-Jordan way of calculating inverse matrix is just similar to LU decomposition except the target of row operations (eliminations) is different. LU decomposition, if in another way, can be called $E^{-1}U$ decomposition.
- The time complexity of Gauss-Jordan way of finding inverse matrix should be same as the LU factorization, which is $O(n^3)$.

QR decomposition / factorization

This is related to Gram-Schmidt orthonormalization. I think it is because the process is for a column operation, we can then have $(A, I)E = (AE, E)$, where E stands for all the column operators to make matrix A to be an orthogonal matrix Q : $AE = Q$. Thus we have $A = QE^{-1} \equiv QR$.

Determinants

The determinant is defined purely on square matrix.

Three fundamental properties of determinant

- (1) $\det I = 1$
- (2) Determinant changes sign when two rows are exchanged. Two applications of properties 1 and 2. Examples:

$$\det P = \begin{cases} 1 & \text{even} \\ -1 & \text{odd} \end{cases}, \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1, \quad \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1.$$

- (3) a. $\begin{vmatrix} ta & tb \\ tc & td \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix}$

b. $\begin{vmatrix} a + a' & b + b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$ Note the above property does not suggest $\det(A + B) = \det A + \det B$.

Three fundamental properties \Rightarrow Seven more properties of determinant

- (4). If two rows are equal (or proportional to each other?), then determinant is zero.
- (5). Elimination (row operation) does not change the determinant of a matrix. So we can always transform a matrix to a triangular matrix and then calculate its determinant.

$$\begin{vmatrix} a & b \\ c - la & d - lb \end{vmatrix} \stackrel{3.b}{=} \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a & b \\ -la & -lb \end{vmatrix} \stackrel{3.a}{=} \begin{vmatrix} a & b \\ c & d \end{vmatrix} - l \begin{vmatrix} a & b \\ a & b \end{vmatrix} \stackrel{4}{=} \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

- (6). If one row is zero, then its determinant is zero. May prove with 3.a or (5, 4).

• (7). $U = \begin{vmatrix} d_1 & * & \cdots & * \\ 0 & d_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{vmatrix}$, then $\det U = d_1 d_2 \cdots d_n$. Using property 5 to first obtain $D = \begin{vmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{vmatrix}$, then using

property 3 to obtain $d_n d_{n-1} \cdots d_1 \begin{vmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{vmatrix}$.

- (8). If A is singular, then $\det A = 0$. If and only if A is invertible, we have $\det A \neq 0$. If a matrix is invertible, then after transforming it to an upper triangular matrix, there are non-zero pivots for each row. And thus determinant will be the product of these pivots and is not zero.

An example $\begin{vmatrix} a & b \\ c & d \end{vmatrix} \xrightarrow{\text{elimination}} \begin{vmatrix} a & b \\ 0 & d - \frac{c}{a}b \end{vmatrix} = ad - bc$. This is first proof of how to calculate the determinant of a 2x2 matrix.

- (9). $\det AB = (\det A)(\det B)$, $\det I = \det A^{-1}A = \det A^{-1} \det A$, $\det A^{-1} = \frac{1}{\det A}$. We also have $\det A^2 = (\det A)^2$ and $\det 2A = 2^n \det A$.
- (10). $\det A^T = \det A$. This indicates that all the properties applied to rows also applied to columns.
- $|A^T| = |A| \rightarrow |U^T L^T| = |LU| \rightarrow |U^T| |L^T| = |L| |U|$

Three fundamental properties \Rightarrow cofactor

Using the three basic properties to derive a new way of calculating determinant using cofactor.

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & 0 \\ c & 0 \end{vmatrix} + \begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} + \begin{vmatrix} 0 & b \\ 0 & d \end{vmatrix} = ad - bc$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & 0 & a_{23} \\ 0 & a_{32} & 0 \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & 0 \\ 0 & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & 0 \\ 0 & 0 & a_{23} \\ a_{31} & 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ a_{21} & 0 & 0 \\ 0 & a_{32} & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 \\ 0 & a_{22} \end{vmatrix} + \begin{vmatrix} 0 & 0 \\ a_{31} & 0 \end{vmatrix}$$

$$= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \quad (1)$$

Introducing the concept of cofactor and rewrite (1) as

$$a_{11}(a_{22}a_{33} - a_{23}a_{32}) + a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$$

$$\begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & a_{23} \\ a_{31} & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 0 \end{vmatrix}$$

Thus we can define the cofactor of a_{ij} to be C_{ij} , which is the determinant of the matrix obtained by taking out the row i and column j of the original matrix. When $i + j$ is an even number, it has a positive sign. Otherwise, a negative sign.

Expanding the determinant according to the first row of A gives

$$\det A = a_{11} C_{11} + a_{12} C_{12} + \cdots + a_{1n} C_{1n}$$

Application of determinant

Find inverse matrix

The proof of the following formula involves the cofactor introduced earlier. Note this is one of ways of finding inverse. See other place for the summary of inverse matrix calculation.

$$A^{-1} = \frac{1}{\det A} C^T \quad (1)$$

Note the difference between cofactor and cofactor matrix. The dimensions of A and the cofactor matrix are same, but the matrix used to calculate cofactor is with lower dimensions.

The adjugate, classical adjoint, or adjunct of a square matrix is the transpose of its cofactor matrix. The adjugate has sometimes been called the "adjoint", but today the "adjoint" of a matrix normally refers to its corresponding adjoint operator, which is its conjugate transpose. From this definition, we have $A^{-1} = \frac{1}{|A|} \text{adj}(A)$. This is rearranged as $\text{adj}(A) = |A| A^{-1}$. Finally $(\text{adj}(A))^T = |A| A^{-T}$. This is the same form as in other notes.

Solve $Ax = b$

$$Ax = b \quad x = A^{-1}b = \frac{1}{\det A} C^T b \quad \text{Cramer's rule}$$

This method is pretty but is not efficient. Also we can only handle equations where A is a square matrix. Otherwise we cannot calculate the determinant.

Area, Volume

In 3D case, taking the first row of the 3×3 matrix A , (a_1, a_2, a_3) , as the coordinates of a point A_1 in 3D space. Similar, second and third rows are the coordinates of points A_2, A_3 . Connecting the three points and the origin gives a parallelepiped. The volume of this parallelepiped is just given by the determinant of the corresponding matrix.

In 2D case, the determinant gives the area of a parallelogram formed by two points given by two rows of the 2D matrix and another point of origin. Normally when we calculate the area of a parallelogram formed by four points $(0, 0), (a, b), (c, d), (a + c, b + d)$, it is necessary to calculate the base length and height. Now we only need $\det A = ad - bc$. For a triangle formed by $(0, 0), (a, b), (c, d)$, its area will be $\frac{1}{2}ad - bc$.

For a triangle formed by points $(x_1, y_1), (x_2, y_2), (x_3, y_3)$, the area is $\frac{1}{2} \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}$. Subtracting first row from 2nd and third rows

(essentially moving triangle to origin) gives $\frac{1}{2} \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 - x_1 & y_2 - y_1 & 0 \\ x_3 - x_1 & y_3 - y_1 & 0 \end{vmatrix}$. Expanding according to third column gives

$$\frac{(x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1)}{2}.$$

Calculation tricks in Linear algebra

Inverse matrix

- Elimination based Gauss-Jordan way. Time complexity is $O(n^3)$. Better ways can achieve complexity of $O(n^{2.3})$, etc. Check Wikipedia on the complexity of calculating inverse matrix of many different ways.
- Using determinant way for 2×2 matrix Using $A^{-1} = \frac{1}{\det A} C^T$, or memorize directly $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

Eigenvalues, pivots, determinant

$$\sum_{i=1}^n \lambda_i = \sum_{i=1}^n a_{ii} = \text{tr}(A)$$

$$\prod_{i=1}^n \lambda_i = \det A$$

$$\prod_{i=1}^n p_i = \prod_{i=1}^n \lambda_i,$$

where p_i are *ith* pivots. The above properties can be used to quickly calculate the eigenvalues, pivots, or determinant of low-dimensional matrix. We can also use them to determine whether they are positive definite or not in a quick way. Also note that the concept of positive definite connects all the three concepts together. See details on how to use eigenvalues, pivots, and determinant to check the positive definiteness of a matrix.

Connecting linear algebra and statistics

- A matrix column (or row) is used to describe the possible values of a random variable. Calculating expectation value or variance are thus often in the forms of vector or matrix calculation. For example, by diagonalizing a matrix, we can calculate the principal axes etc., of a 2D normal distribution.
- As linear transformation can always be decomposed using SVD, which corresponds to rotating, scaling and rotating again. Thus a linear transformation could modify the correlation among the random variables described by a matrix. Rotation of the bases will change correlation while translational action will not. For example, the big correlation among stock prices (not return) should be from the different shifting in two axes, but not due to translational action.