Matrix Transformation
04/14/2025

Matrix transformation has broad applications in linear algebra and data science analyses. We treat the transformation for vectors and matrices in separate sections below.

Vector Transformation

Vector transformation is mapping of a vector to another vector using matrix transformation. This is a linear transformation so that the common algebraic properties are preserved, except that the product of matrices is not commutative ($AB \neq BA$).

Markov Chain: Markov chain transforms a vector in its initial state (S_0) to the n-th state (S_n) after successively left multiplying the stochastic transition matrix P.

$$S_n = P^n S_0$$

At the steady steady, S_S, it is the non-trivial solution of the following homogeneous equation:

$$(I - P)S_S = 0$$

Change of Basis: Let B be the matrix whose columns consist of the basis vectors of basis B and B' matrix of basis vectors of basis B'. A coordinate vector in basis B', $[\mathbf{x}]_{B'}$, is represented in basis B as below:

$$B[\mathbf{x}]_B = B'[\mathbf{x}]_{B'}$$

$$[\mathbf{x}]_B = B^{-1}B'[\mathbf{x}]_{B'}$$

Therefore, the change of basis is done with left multiplying the coordinate vector by the matrix $B^{-1}B'$. If columns of B' are the vectors of a standard basis, it becomes the identity matrix. As such, B is also called the transition matrix; its invertibility is assured by linear independence of the basis vectors.

Solving Linear Systems: If the the coefficient matrix (A) of a linear system equation is consistent and invertible, the solution of the linear system is a vector transformation of the constant vector (b) by multiplying the inverse of the coefficient matrix. When the matrix is invertible, the matrix transformation is *explicit*. Otherwise, the transformation is *implicit*, e.g., for the solutions of a system of homogeneous equations.

$$\mathbf{x} = A^{-1}\mathbf{b}$$

If A is consistent but not invertible, the solution ($\mathbf{x}$) is the sum of the solution of the homogeneous equation ($\mathbf{x}_h$) and the particular solution of the non-homogeneous equation ($\mathbf{x}_p$).

$$\mathbf{x} = \mathbf{x}_h + \mathbf{x}_p$$

If A is inconsistent, a solution is not possible, but a best approximation solution ($\hat{\mathbf{x}}$) of the least square problem exists:

$$A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$$

If A is invertible, so is $A^T A$, the solution is explicit:

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T\mathbf{b}$$

Eigenvectors: For a given eigenvalue λ_i of matrix A, the corresponding eigenvector ($\mathbf{v}_i$) is the non-trivial solution of the following homogeneous equation.

$$(A - \lambda_i I)\mathbf{v}_i = \mathbf{0}$$

<u>Matrix Transformation</u>

Similar Matrices: Analogous to change of basis for vectors, if there is an invertible matrix P such that $A = P^{-1}A'P$, then matrix A is similar to matrix A'. Similar matrices have the same dimensions and the same eigenvalues:

$$\begin{aligned} &|\lambda I - A| \\ &= |\lambda I - P^{-1}A'P| = |P^{-1}\lambda IP - P^{-1}A'P| \\ &= |P^{-1}\lambda I - A'P| = |P^{-1}P||\lambda I - A'| = |\lambda I - A'| \end{aligned}$$

It is reasonable to regard similar matrices as representations of the same eigenvalues, but the eigenvectors are in different subspaces with different bases. P, therefore, is the transition matrix between the bases.

Elementary Matrices: An elementary matrix (E) is constructed by performing a single elementary row operation on an identity matrix. The same row operation can be accomplished for a matrix by left multiplying the elementary matrix. The resulting matrix is regarded as “row equivalent” to the original matrix.

Matrix Inversion: If matrix A is invertible and row equivalent to the identity matrix through k row operations, or left multiplying by k elementary matrices ($E_k \dots E_2 E_1 A$), the following gives the inverse of matrix A that is a matrix transformation of the identity matrix.

$$A^{-1} = E_k E_{k-1} \dots E_1 I$$

LU Factorization: If matrix A is row reduced to a upper triangular matrix U by left multiplying k elementary matrices ($U = E_k \dots E_2 E_1 A$), then A= LU is an LU-factorization of A, where L is a lower triangular matrix, $(E_k \dots E_2 E_1)^{-1}$. The lower triangular matrix L is assured by the elementary matrices that only replace a row by adding a multiple of an *upper* row (no row interchange, row scaling or replacing a row by multiple of a lower row). It is more efficient in solving linear system equations by using LU-factorization than by using A^{-1}.

Cholesky Factorization: A symmetric matrix A is row reduced to an upper triangular matrix U' with positive diagonal entries and each row operation only adds a multiple of a row to a lower row (equivalently left multiplying a replacement elementary matrix). Further, U is obtained by dividing (scaling) each row by the square root of the diagonal entry of the same row. Finally, A is factored as $U^T U$.

QR Factorization: For an m x n matrix A with linearly independent columns, it can be factored as A=QR, where Q is an m x n matrix whose columns form an orthonormal basis for Col A and R is an n x n upper triangular invertible matrix with positive entries on its diagonal. The orthonormal basis is constructed using the Gram-Schmidt procedure and forms the columns of matrix Q. The R matrix is constructed as follows:

$$\begin{bmatrix} \mathbf{v}_1\mathbf{q}_1 & \mathbf{v}_2\mathbf{q}_1 & \cdots & \mathbf{v}_n\mathbf{q}_1 \\ 0 & \mathbf{v}_2\mathbf{q}_2 & \cdots & \mathbf{v}_n\mathbf{q}_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{v}_m\mathbf{q}_n \end{bmatrix}$$

Diagonalization: If matrix A has n linearly independent eigenvectors and matrix P forms the transition matrix whose columns consist of these eigenvectors, i.e.,

$$P = [\mathbf{p}_1 \ \mathbf{p}_2 \ \cdots \ \mathbf{p}_n]$$

The diagonal matrix $D = P^{-1}AP$ has the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ on its main diagonal and zeros elsewhere. However, if the eigenvalues have multiplicities but the corresponding number of eigenvectors is less than the multiplicity, the matrix is not diagonalizable. If A is symmetric, $A = PDP^T$ is also called the eigenvalue decomposition.

However, if A is not symmetric, we have the Hessenberg decomposition $A = PHP^T$ where H is in the upper Hessenberg form. If A has real eigenvalues, then it is a Schur decomposition $A = PSP^T$ and S is upper triangular.

Singular Value Decomposition (SVD): If A is an m x n matrix of rank k, then A is factored as:

$$A = U\Sigma V^T$$

where U, Σ, and V are m x m, m x n, and n x n matrices, respectively. Further, the column vectors of V are eigenvectors of $A^T A$ ordered based on the magnitude of the positive eigenvalues. The diagonal entries of Σ are the square roots of the positive eigenvalues and arranged in the same order. The first k columns of matrix U are:

$$\mathbf{u}_i = \frac{A\mathbf{v}_i}{||A\mathbf{v}_i||} = \frac{A\mathbf{v}_i}{\sqrt{\lambda_i}}$$

These column vectors form an orthonormal basis for Col A that is extended to form an orthonormal basis for $\mathbb{R}^m$.

Concluding Remarks

Several important vector computations can be accomplished using simple matrix multiplication when the matrix is invertible. Examples are Markov state vector evolution, change of basis, and solving linear system equations. However, if the matrix is not invertible, the vectors must be solved using row reduction (Gaussian elimination) of the homogeneous equations. Therefore the solution is implicit.

Similar cases are found with matrix decompositions. These transformations are often made using matrix row reduction; therefore, these procedures are implicit in most cases, especially when eigenvalues, or orthonormal column vectors are required in constructing the matrix decomposition.