International Journal of Creative and Open Research in Engineering and Management
ISSN: 3108-1754 (Online)
Volume 02 Issue 09 September-2026 | Impact Factor: 3.5

Matrix Factorization Techniques in Data Science

S.Sathyapriya
Department of Mathematics
Sri Krishna Arts and Science College
13sathya92@gmail.com

How to Cite this Article:

S.Sathyapriya, (2026). Matrix Factorization Techniques in Data Science. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(9), 1-9. https://doi.org/10.55041/ijcope.v2i9.059

https://doi.org/10.55041/ijcope.v2i9.059

Abstract

Matrix factorization is an important mathematical technique used in data science to represent large and complex datasets in a lower-dimensional form. The fundamental idea is to decompose a matrix into two or more matrices whose product approximates or reconstructs the original data matrix. Techniques such as Singular Value Decomposition (SVD), Non-Negative Matrix Factorization (NMF), and QR factorization have applications in dimensionality reduction, recommendation systems, image processing, text mining, and machine learning. This paper presents the mathematical foundations of matrix factorization, discusses major factorization techniques, and examines their applications in data science. The advantages, limitations, and practical significance of these methods are also discussed. The study demonstrates that matrix factorization provides an effective framework for extracting hidden patterns and reducing computational complexity in high-dimensional datasets.

Keywords: Matrix Factorization, Data Science, Singular Value Decomposition, Non-Negative Matrix Factorization, Dimensionality Reduction, Recommendation Systems, Machine Learning

1. Introduction

Data science involves the collection, processing, analysis, and interpretation of large amounts of data. Modern datasets are frequently represented in the form of matrices, where rows may represent observations, users, documents, or images and columns may represent features, products, words, or measurements. As the size and dimensionality of datasets increase, direct analysis of these matrices can become computationally expensive.

Matrix factorization provides a mathematical approach for simplifying such datasets. In general, a matrix Acan be represented as the product of two or more matrices:

$$A \approx UV^T,$$

where $U$and $V$are lower-dimensional factor matrices. Instead of working directly with the original high-dimensional matrix, data scientists can analyze these factor matrices to identify latent or hidden structures in the data.

Matrix factorization is particularly useful when the original dataset contains redundant information. By reducing the number of dimensions while retaining important information, factorization techniques can improve computational efficiency and reveal meaningful relationships within the data.

2. Objectives of the Study

The main objectives of this paper are:

- To introduce the concept and mathematical foundations of matrix factorization.
- To study important matrix factorization techniques used in data science.
- To examine the role of matrix factorization in dimensionality reduction.
- To investigate its applications in recommendation systems, image processing, and text analysis.

International Journal of Creative and Open Research in Engineering and Management
ISSN: 3108-1754 (Online)
Volume 02 Issue 09 September-2026 | Impact Factor: 3.5

- To compare the advantages and limitations of different factorization methods.
- To explain how matrix factorization contributes to efficient data analysis.

3. Mathematical Foundation

Let

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

be a data matrix containing $m$observations and $n$features. Matrix factorization attempts to express Ausing smaller or structured matrices.

A general low-rank approximation can be written as

$$A \approx UV^T,$$

where

$$U \in \mathbb{R}^{m \times k}, V \in \mathbb{R}^{n \times k},$$

and $k \ll \min(m, n)$.

The parameter $k$represents the reduced dimension or rank of the approximation. The objective is generally to minimize the reconstruction error:

$$\min_{U,V} \parallel A - UV^T \parallel_F^2,$$

where $\|\cdot\|_F$denotes the Frobenius norm.

This formulation is fundamental to many matrix factorization methods used in data science.

4. Major Matrix Factorization Techniques

4.1 Singular Value Decomposition

Singular Value Decomposition (SVD) is one of the most important matrix factorization techniques. Any real matrix Acan be decomposed as

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

where:

- $U$is an orthogonal matrix containing left singular vectors.
- $V$is an orthogonal matrix containing right singular vectors.
- $\Sigma$is a diagonal matrix containing non-negative singular values.

The singular values are usually arranged in decreasing order:

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0.$$

A reduced-rank approximation can be obtained by retaining only the first $k$singular values:

$$A_k = U_k \Sigma_k V_k^T.$$

This provides an effective method for dimensionality reduction and noise reduction.

4.2 Non-Negative Matrix Factorization

Non-Negative Matrix Factorization (NMF) decomposes a non-negative matrix Ainto two non-negative matrices:

$$A \approx WH,$$

subject to

$$W \geq 0, H \geq 0.$$

The non-negativity constraint often produces factors that are easier to interpret. NMF has applications in image analysis, document classification, topic modeling, and biological data analysis.

For example, an image can be represented as a matrix of pixel intensities. NMF can identify additive components that correspond to meaningful visual patterns.

4.3 QR Factorization

QR factorization expresses a matrix as

$$A = QR,$$

where Qhas orthonormal columns and $R$is an upper triangular matrix.

QR factorization is widely used in numerical linear algebra and is particularly important for solving least-squares problems and computing eigenvalues.

For a system

$$Ax \approx b,$$

QR factorization gives

$$QRx \approx b.$$

Since $Q$is orthogonal,

$$Rx = Q^T b,$$

which can be solved efficiently using back substitution.

4.4 LU Factorization

LU factorization decomposes a matrix into a lower triangular matrix $L$and an upper triangular matrix U:

$$A = LU.$$

It is commonly used for solving systems of linear equations. For example,

$$Ax = b$$

can be solved through the two systems

$$Ly = b$$

and

$$Ux = y.$$

LU factorization is useful when multiple systems involving the same coefficient matrix need to be solved.

5. Applications in Data Science

5.1 Recommendation Systems

One of the most important applications of matrix factorization is collaborative filtering.

Suppose $R$is a user-item rating matrix:

$$R \in \mathbb{R}^{m\times n}.$$

Rows represent users and columns represent items such as movies or products. Since many entries may be missing, matrix factorization can approximate the rating matrix as

$$R \approx UV^T.$$

Here, Urepresents latent user characteristics and Vrepresents latent item characteristics.

The predicted rating can be calculated as

$$\hat{r}_{ij} = u_i^T v_j.$$

This approach allows recommendation systems to predict a user's preferences for items they have not previously rated.

5.2 Dimensionality Reduction

High-dimensional datasets may contain hundreds or thousands of variables. Matrix factorization can reduce the number of dimensions while preserving important information.

SVD, for example, can transform

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

into a lower-dimensional representation involving only the most significant singular vectors.

Dimensionality reduction can:

- Reduce computational cost.
- Remove redundant information.
- Improve visualization.
- Reduce noise.
- Facilitate machine-learning algorithms.

5.3 Image Compression

Digital images can be represented as matrices of pixel values. An image matrix Acan be approximated using a truncated SVD:

$$A_k = U_k \Sigma_k V_k^T.$$

If $k$is significantly smaller than the original rank, the image can be represented using fewer numerical values while maintaining much of its visual information.

Thus, SVD provides a mathematical basis for lossy image compression.

5.4 Text Mining and Topic Modeling

Documents can be represented using a term-document matrix. Each row may represent a document and each column a word.

SVD can be applied to this matrix to obtain a lower-dimensional representation. This approach is closely related to Latent Semantic Analysis (LSA), where hidden relationships between documents and terms are identified.

Matrix factorization can therefore assist in:

- Document classification.
- Information retrieval.
- Search engines.
- Topic discovery.
- Text clustering.

5.5 Machine Learning

Matrix factorization appears throughout machine learning. Feature matrices are often factorized or approximated to reduce dimensionality and computational requirements.

For example, Principal Component Analysis (PCA) can be computed using the eigen decomposition of a covariance matrix or through SVD of a centered data matrix.

Therefore, matrix factorization forms an important mathematical foundation for many dimensionality-reduction techniques.

6. Comparison of Major Techniques

Technique	Basic Form	Main Application	Important Feature
SVD	$A = U\Sigma V^T$	Dimensionality reduction, compression	Excellent low-rank approximation
NMF	$A \approx WH$	Image and text analysis	Non-negative and interpretable factors
QR	$A = QR$	Least squares, numerical computation	Orthogonal factorization
LU	$A = LU$	Linear systems	Efficient solution of equations

Each method has a different mathematical structure and is suitable for different types of data-science problems.

7. Conclusion

Matrix factorization is a fundamental mathematical tool with extensive applications in data science. Techniques such as SVD, NMF, QR factorization, and LU factorization provide different approaches for decomposing matrices and simplifying complex datasets.

Among these methods, SVD is particularly valuable for dimensionality reduction, low-rank approximation, and image compression, while NMF provides interpretable non-negative representations. Matrix factorization is also central to collaborative filtering and recommendation systems, where latent relationships between users and items can be discovered.

The combination of mathematical theory and computational techniques makes matrix factorization an important area of study in modern data science. As datasets continue to grow in size and complexity, efficient matrix factorization methods are expected to remain important in machine learning, artificial intelligence, information retrieval, and other data-driven fields.

References

1. Strang, G. (2016). *Introduction to Linear Algebra*. Wellesley-Cambridge Press.

2. Golub, G. H., & Van Loan, C. F. (2013). *Matrix Computations*. Johns Hopkins University Press.

3. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.

4. Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8), 30–37.

5. Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401, 788–791.

6. Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A*, 374.