e-PGPathshala
Subject : Computer Science
Paper: Data Analytics
Module No 29: CS/DA/29 - Dimensionality Reduction: Need for DM and SVD
Quadrant 1 – e-text

1.1 Introduction to Dimensionality Reduction

In machine learning and statistics, dimensionality reduction or dimension reduction is the process of reducing the number of random variables under consideration, via obtaining a set of principal variables. It can be divided into feature selection and feature extraction. Dimensionality reduction1 can also be seen as the process of deriving a set of degrees of freedom which can be used to reproduce most of the variability of a data set. Consider a set of images produced by the rotation of a face through different angles. Clearly only one degree of freedom is being altered and thus the images lie along a continuous one dimensional curve through image space.

It is so easy and convenient to collect data. Data is not collected only for data mining. Data accumulates in an unprecedented speed. Data preprocessing is an important part for effective machine learning and data mining. Dimensionality reduction is an effective approach to downsizing data. Most machine learning and data mining techniques may not be effective for high dimensional data

- Curse of Dimensionality
- Query accuracy and efficiency degrade rapidly as the dimension increases.

1.2 Learning Objectives

- Understand dimensionality reduction and its need

- Know the steps in SVD technique

1.3 Overview

- Dimensionality reduction and its need
- Singular Value Decomposition
- Complexity of SVD

The intrinsic dimension may be small. For example, the number of genes responsible for a certain type of disease may be small.

- **Visualization:** projection of high-dimensional data onto 2D or 3D.
- **Data compression:** efficient storage and retrieval.
- **Noise removal:** positive effect on query accuracy.

Major Techniques of Dimensionality Reduction

- Feature selection
- Feature Extraction (reduction)
- Differences between the two techniques

Compress / reduce dimension:

- 10^6 rows; 10^3 columns; no updates
- Random access to any cell(s); **small error: OK**
-

customer \ day	We 7/10/96	Th 7/11/96	Fr 7/12/96	Sa 7/13/96	Su 7/14/96
ABC Inc.	1	1	1	0	0
DEF Ltd.	2	2	2	0	0
GHI Inc.	1	1	1	0	0
KLM Co.	5	5	5	0	0
Smith	0	0	0	2	2
Johnson	0	0	0	3	3
Thompson	0	0	0	1	1

- The above matrix is really “2-dimensional.” All rows can be reconstructed by scaling [1 1 1 0 0] or [0 0 0 1 1]

Rank of a Matrix

The rank of A is the maximal number of linearly independent column vectors in A, i.e. the maximal number of linearly independent vectors among {a1, a2, . . . , an}. If A = 0, then the rank of A is 0.

- **Q:** What is rank of a matrix A?
- **A:** Number of linearly independent columns of A

- **For example:**
- Matrix **A** = $\begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix}$ has rank **r=2**
- **Why?**
    - The first two rows are linearly independent, so the rank is at least 2, but all three rows are linearly dependent (the first is equal to the sum of the second and third) so the rank must be less than 3.
- Why do we care about low rank?
    - We can write **A** as two “basis” vectors: [1 2 1] [-2 -3 1]
    - And new coordinates of : [1 0] [0 1] [1 1]

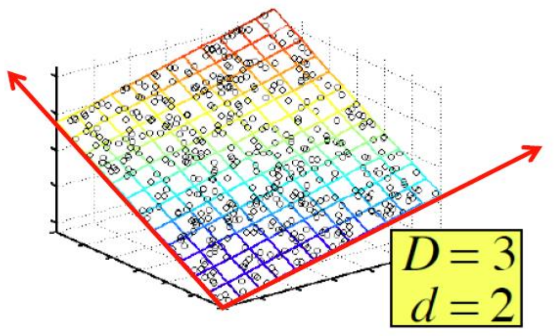
Rank is “Dimensionality”

- **Cloud of points 3D space:**
    - Think of point positions as a matrix:

1 row per point: $\begin{array}{c} \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix} \end{array}\begin{array}{l} \mathbf{A} \\ \mathbf{B} \\ \mathbf{C} \end{array}$

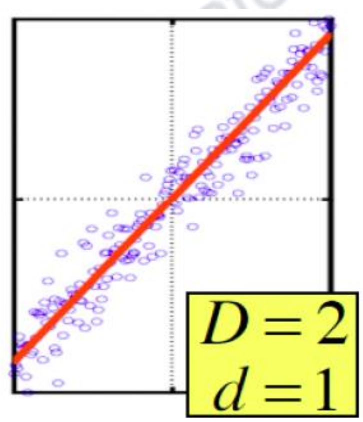
- **We can rewrite coordinates more efficiently!**

- Old basis vectors:[1 0 0] [0 1 0] [0 0 1]
- **New basis vectors: [1 2 1] [-2 -3 1]**
- Then **A** has new coordinates: [1 0]. **B**: [0 1], **C**: [1 1]
  - **Notice:** We reduced the number of coordinates!



1.4 Dimensionality Reduction

- Goal of dimensionality reduction is todiscover the axis of data!
- Rather than representing every point with 2 coordinates we represent each point with 1 coordinate (corresponding to the position of the point on the red line).
- By doing this we incur a bit of **error** as the points do not exactly lie on the line



Why reduce dimensions?

- Discover hidden correlations/topics

        - Words that occur commonly together
    - Remove redundant and noisy features
        - Not all words are useful
    - Interpretation and visualization
    - Easier storage and processing of the data

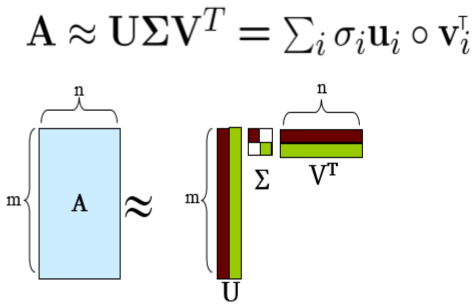
1.5 Singular Value Decomposition:

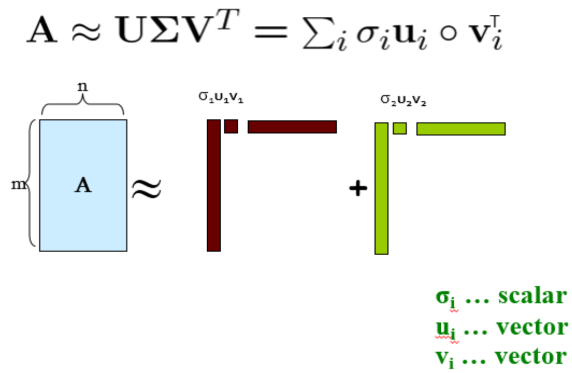
We now take up a second form of matrix analysis that leads to a low-dimensional representation of a high-dimensional matrix. This approach, called singularvalue decomposition (SVD), allows an exact representation of any matrix, and also makes it easy to eliminate the less important parts of that representation to produce an approximate representation with any desired number of dimensions. Of course the fewer the dimensions we choose, the less accurate will be the approximation. We begin with the necessary definitions. Then, we explore the idea that the SVD defines a small number of "concepts" that connect the rows and columns of the matrix. We show how eliminating the least important concepts gives us a smaller representation that closely approximates the original matrix.

SVD - Definition

$A_{[m \times n]} = U_{[m \times r]} \Sigma_{[r \times r]} (V_{[n \times r]})^T$

- A: Input data matrix
    - *m* x *n* matrix (e.g., *m* documents, *n* terms)
- U: Left singular vectors
    - *m* x *r* matrix (*m* documents, *r* concepts)
- Σ: Singular values
    - *r* x *r* diagonal matrix (strength of each 'concept') (*r* : rank of the matrix A)
- V: Right singular vectors
    - *n* x *r* matrix (*n* terms, *r* concepts)





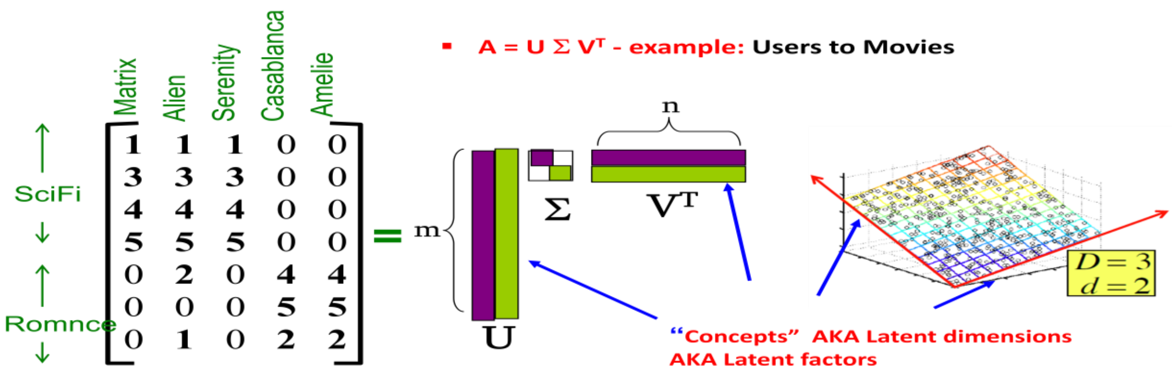
Basic ideas behind SVD:

These are the basic ideas behind SVD: taking a high dimensional, highly variable set of data points and reducing it to a lower dimensional space that exposes the substructure of the original data more clearly and orders it from most variation to the least. What makes SVD practical for NLP applications is that you can simply ignore variation below a particular threshold to massively reduce your data but be assured that the main relationships of interest have been preserved.

Properties

- It is **always** possible to decompose a real matrix $\boldsymbol{A}$ into $\boldsymbol{A} = \boldsymbol{U}\,\Sigma \boldsymbol{V}^T$, where
  - $\boldsymbol{U}, \Sigma, \boldsymbol{V}$: unique
  - $\boldsymbol{U}, \boldsymbol{V}$: column orthonormal
    - $\boldsymbol{U}^T \boldsymbol{U} = \boldsymbol{I}$; $\boldsymbol{V}^T \boldsymbol{V} = \boldsymbol{I}$ ($\boldsymbol{I}$: identity matrix)
    - (Columns are orthogonal unit vectors)
  - Σ: diagonal
    - Entries (**singular values**) are positive, and sorted in decreasing order ($\boldsymbol{\sigma_1 \geq \sigma_2 \geq \ldots \geq 0}$)

Example: Users-to-Movies



$$
\begin{array}{c|c}
 & \text{Matrix} \quad \text{Alien} \quad \text{Serenity} \quad \text{Casablanca} \quad \text{Amelie} \\
\begin{array}{c} \uparrow \\ \text{SciFi} \\ \downarrow \\ \uparrow \\ \text{Romnce} \\ \downarrow \end{array} &
\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix}
\end{array}
=
\begin{bmatrix} \mathbf{0.13} & 0.02 & -0.01 \\ \mathbf{0.41} & 0.07 & -0.03 \\ \mathbf{0.55} & 0.09 & -0.04 \\ \mathbf{0.68} & 0.11 & -0.05 \\ 0.15 & \mathbf{-0.59} & \mathbf{0.65} \\ 0.07 & \mathbf{-0.73} & \mathbf{-0.67} \\ 0.07 & \mathbf{-0.29} & \mathbf{0.32} \end{bmatrix}
\times
\begin{bmatrix} \mathbf{12.4} & 0 & 0 \\ 0 & \mathbf{9.5} & 0 \\ 0 & 0 & \mathbf{1.3} \end{bmatrix}
\times
\begin{bmatrix} \mathbf{0.56} & \mathbf{0.59} & \mathbf{0.56} & 0.09 & 0.09 \\ 0.12 & -0.02 & 0.12 & \mathbf{-0.69} & \mathbf{-0.69} \\ 0.40 & \mathbf{-0.80} & 0.40 & 0.09 & 0.09 \end{bmatrix}
$$

1.5.1 Steps in SVD

1. Consider a matrix A.

if

$$A = \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}$$

2. Computing A^T, AA^T, and A^TA

then

$$A^T = \begin{bmatrix} 4 & 3 \\ 0 & -5 \end{bmatrix}$$

"left matrix"

$$AA^T = \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}\begin{bmatrix} 4 & 3 \\ 0 & -5 \end{bmatrix}$$

$$AA^T = \begin{bmatrix} 16 & 12 \\ 12 & 34 \end{bmatrix}$$

"right matrix"

$$A^TA = \begin{bmatrix} 4 & 3 \\ 0 & -5 \end{bmatrix}\begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}$$

$$A^TA = \begin{bmatrix} 25 & -15 \\ -15 & 25 \end{bmatrix}$$

3. Compute the Eigenvalues

"left matrix"

$$AA^T = \begin{bmatrix} 16 & 12 \\ 12 & 34 \end{bmatrix}$$

$$AA^T - cI = \begin{bmatrix} 16 - c & 12 \\ 12 & 34 - c \end{bmatrix}$$

$$| AA^T - cI | = (16 - c)(34 - c) - (12)(12) = 0$$

"right matrix"

$$A^TA = \begin{bmatrix} 25 & -15 \\ -15 & 25 \end{bmatrix}$$

$$A^TA - cI = \begin{bmatrix} 25 - c & -15 \\ -15 & 25 - c \end{bmatrix}$$

$$| A^TA - cI | = (25 - c)(25 - c) - (-15)(-15) = 0$$

characteristic equation $\longrightarrow$ $c^2 - 50c + 400 = 0$

The quadratic equation gives two values.
In decreasing order, these are $\longrightarrow$ $| 40 | > | 10 |$

eigenvalues $\longrightarrow$ $c_1 = 40 \quad c_2 = 10$

for $c_1 = 40$

$$A^TA - cI = \begin{bmatrix} 25-40 & -15 \\ -15 & 25-40 \end{bmatrix} = \begin{bmatrix} -15 & -15 \\ -15 & -15 \end{bmatrix}$$

$$(A^TA - cI)\,\mathbf{x}_1 = 0$$

$$\begin{bmatrix} -15 & -15 \\ -15 & -15 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$-15x_1 + -15x_2 = 0$$
$$-15x_1 + -15x_2 = 0$$

Solving for x_2 for either equation: $x_2 = -x_1$

$$\mathbf{x}_1 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ -x_1 \end{bmatrix}$$

Dividing by its length,

$$L = \sqrt{x_1^2 + x_2^2} = x_1\sqrt{2}$$

$$\mathbf{x}_1 = \begin{bmatrix} x_1/L \\ -x_1/L \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 0.7071 \\ -0.7071 \end{bmatrix}$$

for $c_2 = 10$

$$A^TA - cI = \begin{bmatrix} 25-10 & -15 \\ -15 & 25-10 \end{bmatrix} = \begin{bmatrix} 15 & -15 \\ -15 & 15 \end{bmatrix}$$

$$(A^TA - cI)\,\mathbf{x}_2 = 0$$

$$\begin{bmatrix} 15 & -15 \\ -15 & 15 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$15x_1 + -15x_2 = 0$$
$$-15x_1 + 15x_2 = 0$$

Solving for x_2 for either equation: $x_2 = x_1$

$$\mathbf{x}_2 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_1 \end{bmatrix}$$

Dividing by its length,

$$L = \sqrt{x_1^2 + x_2^2} = x_1\sqrt{2}$$

$$\mathbf{x}_2 = \begin{bmatrix} x_1/L \\ x_1/L \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 0.7071 \\ 0.7071 \end{bmatrix}$$

1.5.2 Construct V and S matrix

- Construct V by placing vectors along its columns and compute V^T

$$V = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} 0.7071 & 0.7071 \\ -0.7071 & 0.7071 \end{bmatrix}$$

$$V^T = \begin{bmatrix} 0.7071 & -0.7071 \\ 0.7071 & 0.7071 \end{bmatrix}$$

- C

$$s_1 = (c_1)^{1/2} = (40)^{1/2} = 6.32\ldots$$
$$s_2 = (c_2)^{1/2} = (10)^{1/2} = 3.16\ldots$$

- Construct matrix S by pl

$$S = \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix} = \begin{bmatrix} 6.32 & 0 \\ 0 & 3.16 \end{bmatrix}$$

lescending order along its main diagonal

1.5.3Computing "left" eigenvectors and U

- Use $U = AVS^{-1}$
- A and V are known. So compute S^{-1}

$$S^{-1} = \begin{bmatrix} \frac{1}{s_1} & 0 \\ 0 & \frac{1}{s_2} \end{bmatrix}$$

$$\frac{1}{s_1} = \frac{1}{6.3245} = 0.1581 \qquad \frac{1}{s_2} = \frac{1}{3.1622} = 0.3162$$

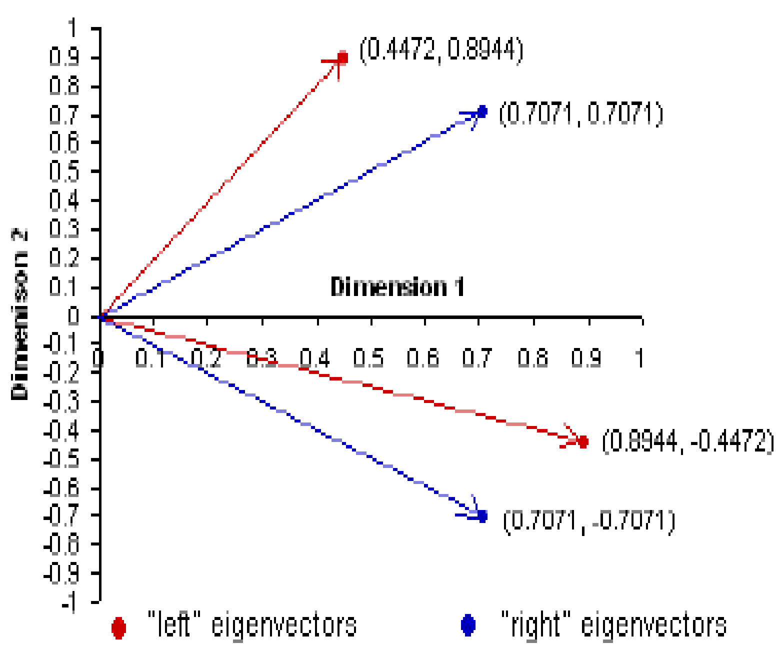
$$U = AVS^{-1} = \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix} \begin{bmatrix} 0.7071 & 0.7071 \\ -0.7071 & 0.7071 \end{bmatrix} \begin{bmatrix} 0.1581 & 0 \\ 0 & 0.3162 \end{bmatrix}$$

$$U = AVS^{-1} = \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix} \begin{bmatrix} 0.1118 & 0.2236 \\ -0.1118 & 0.2236 \end{bmatrix}$$

$$U = AVS^{-1} = \begin{bmatrix} 0.4472 & 0.8944 \\ 0.8944 & -0.4472 \end{bmatrix}$$

Orthogonal Nature

- Dot products between column vectors
- All dot products are equal to zero



Computing the Full SVD

$$A = USV^T = \begin{bmatrix} 0.4472 & 0.8944 \\ 0.8944 & -0.4472 \end{bmatrix} \begin{bmatrix} 6.3245 & 0 \\ 0 & 3.1622 \end{bmatrix} \begin{bmatrix} 0.7071 & -0.7071 \\ 0.7071 & 0.7071 \end{bmatrix}$$

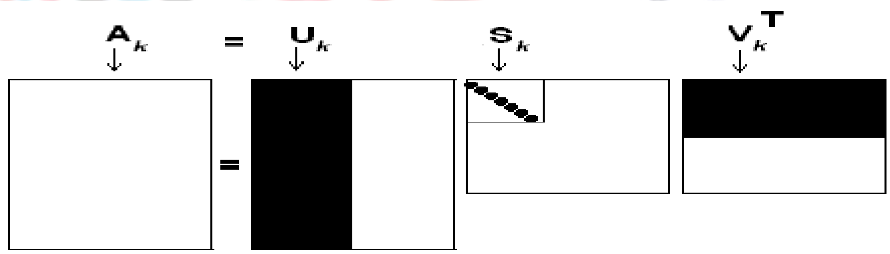
$$A = USV^T = \begin{bmatrix} 0.4472 & 0.8944 \\ 0.8944 & -0.4472 \end{bmatrix} \begin{bmatrix} 4.4721 & -4.4721 \\ 2.2360 & 2.2360 \end{bmatrix}$$

$$A = USV^T = \begin{bmatrix} 3.9998 & 0 \\ 2.9999 & -4.9997 \end{bmatrix} \approx \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}$$

1.6 Dimensionality Reduction with SVD

Suppose we want to represent a very large matrix M by its SVD components U, Σ, and V , but these matrices are also too large to store conveniently. The best way to reduce the dimensionality of the three matrices is to set the smallest of the singular values to zero. If we set the s smallest singular values to 0, then we can also eliminate the corresponding s columns of U and V .

The Reduced SVD



The shaded areas indicate the part of the matrices retained

$$\mathbf{A}_k = \mathbf{U}_k \mathbf{S}_k \mathbf{V}^{\mathbf{T}}_k$$

Interpretation

The key to understanding what SVD offers is in viewing the r columns of U, Σ, and V as representing concepts that are hidden in the original matrix M. In Example 11.8, these concepts are clear; one is “science fiction” and the other is “romance.” Let us think of the rows of M as people and the columns of M as movies. Then matrix U connects people to concepts.

In general, the concepts will not be so clearly delineated. There will be fewer 0's in U and V , although Σ is always a diagonal matrix and will always have 0's off the diagonal. The entities represented by the rows and columns of M (analogous to people and movies in our example) will partake of several different concepts to varying degrees. In fact, the decomposition of Example 11.8 was especially simple, since the rank of the matrix M was equal to the desired number of columns of U, Σ, and V . We were therefore able to get an exact decomposition of M with only two columns for each of the three matrices U, Σ, and V ; the product UΣV T, if carried out to infinite precision, would be exactly M. In practice, life is not so simple. When the rank of M is greater than the number of columns we want for the matrices U, Σ, and V , the decomposition is not exact. We need to eliminate from the exact decomposition those columns of U and V that correspond to the smallest singular values, in order to get the best approximation

How exactly is dim. reduction done?

$$\begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & 0 & 0 \\ \mathbf{3} & \mathbf{3} & \mathbf{3} & 0 & 0 \\ \mathbf{4} & \mathbf{4} & \mathbf{4} & 0 & 0 \\ \mathbf{5} & \mathbf{5} & \mathbf{5} & 0 & 0 \\ 0 & \mathbf{2} & 0 & \mathbf{4} & \mathbf{4} \\ 0 & 0 & 0 & \mathbf{5} & \mathbf{5} \\ 0 & \mathbf{1} & 0 & \mathbf{2} & \mathbf{2} \end{bmatrix} = \begin{bmatrix} \mathbf{0.13} & 0.02 & -0.01 \\ \mathbf{0.41} & 0.07 & -0.03 \\ \mathbf{0.55} & 0.09 & -0.04 \\ \mathbf{0.68} & 0.11 & -0.05 \\ 0.15 & \mathbf{-0.59} & \mathbf{0.65} \\ 0.07 & \mathbf{-0.73} & \mathbf{-0.67} \\ 0.07 & \mathbf{-0.29} & \mathbf{0.32} \end{bmatrix} \times \begin{bmatrix} \mathbf{12.4} & 0 & 0 \\ 0 & \mathbf{9.5} & 0 \\ 0 & 0 & \mathbf{1.3} \end{bmatrix} \times \begin{bmatrix} \mathbf{0.56} & \mathbf{0.59} & \mathbf{0.56} & 0.09 & 0.09 \\ 0.12 & -0.02 & 0.12 & \mathbf{-0.69} & \mathbf{-0.69} \\ 0.40 & \mathbf{-0.80} & 0.40 & 0.09 & 0.09 \end{bmatrix}$$

- **More details**
    - Q: How exactly is dim. reduction done?
    - A: Set smallest singular values to zero

$$
\begin{bmatrix}
\mathbf{1} & \mathbf{1} & \mathbf{1} & 0 & 0 \\
\mathbf{3} & \mathbf{3} & \mathbf{3} & 0 & 0 \\
\mathbf{4} & \mathbf{4} & \mathbf{4} & 0 & 0 \\
\mathbf{5} & \mathbf{5} & \mathbf{5} & 0 & 0 \\
0 & \mathbf{2} & 0 & \mathbf{4} & \mathbf{4} \\
0 & 0 & 0 & \mathbf{5} & \mathbf{5} \\
0 & \mathbf{1} & 0 & \mathbf{2} & \mathbf{2}
\end{bmatrix}
=
\begin{bmatrix}
\mathbf{0.13} & 0.02 & -0.01 \\
\mathbf{0.41} & 0.07 & -0.03 \\
\mathbf{0.55} & 0.09 & -0.04 \\
\mathbf{0.68} & 0.11 & -0.05 \\
0.15 & \mathbf{-0.59} & \mathbf{0.65} \\
0.07 & \mathbf{-0.73} & \mathbf{-0.67} \\
0.07 & \mathbf{-0.29} & \mathbf{0.32}
\end{bmatrix}
\times
\begin{bmatrix}
\mathbf{12.4} & 0 & 0 \\
0 & \mathbf{9.5} & 0 \\
0 & 0 & \cancel{\mathbf{1.3}}
\end{bmatrix}
\times
\begin{bmatrix}
\mathbf{0.56} & \mathbf{0.59} & \mathbf{0.56} & 0.09 & 0.09 \\
0.12 & -0.02 & 0.12 & \mathbf{-0.69} & \mathbf{-0.69} \\
0.40 & \mathbf{-0.80} & 0.40 & 0.09 & 0.09
\end{bmatrix}
$$

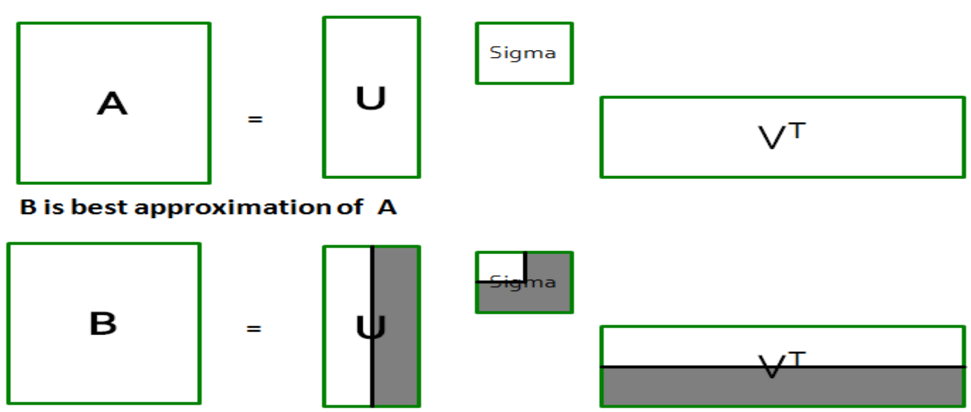
$$
\begin{bmatrix}
\mathbf{1} & \mathbf{1} & \mathbf{1} & 0 & 0 \\
\mathbf{3} & \mathbf{3} & \mathbf{3} & 0 & 0 \\
\mathbf{4} & \mathbf{4} & \mathbf{4} & 0 & 0 \\
\mathbf{5} & \mathbf{5} & \mathbf{5} & 0 & 0 \\
0 & \mathbf{2} & 0 & \mathbf{4} & \mathbf{4} \\
0 & 0 & 0 & \mathbf{5} & \mathbf{5} \\
0 & \mathbf{1} & 0 & \mathbf{2} & \mathbf{2}
\end{bmatrix}
\approx
\begin{bmatrix}
\mathbf{0.13} & 0.02 & -0.01 \\
\mathbf{0.41} & 0.07 & -0.03 \\
\mathbf{0.55} & 0.09 & -0.04 \\
\mathbf{0.68} & 0.11 & -0.05 \\
0.15 & \mathbf{-0.59} & \mathbf{0.65} \\
0.07 & \mathbf{-0.73} & \mathbf{-0.67} \\
0.07 & \mathbf{-0.29} & \mathbf{0.32}
\end{bmatrix}
\times
\begin{bmatrix}
\mathbf{12.4} & 0 & 0 \\
0 & \mathbf{9.5} & 0 \\
0 & 0 & \cancel{\mathbf{1.3}}
\end{bmatrix}
\times
\begin{bmatrix}
\mathbf{0.56} & \mathbf{0.59} & \mathbf{0.56} & 0.09 & 0.09 \\
0.12 & -0.02 & 0.12 & \mathbf{-0.69} & \mathbf{-0.69} \\
0.40 & \mathbf{-0.80} & 0.40 & 0.09 & 0.09
\end{bmatrix}
$$

$$
\begin{bmatrix}
\mathbf{1} & \mathbf{1} & \mathbf{1} & 0 & 0 \\
\mathbf{3} & \mathbf{3} & \mathbf{3} & 0 & 0 \\
\mathbf{4} & \mathbf{4} & \mathbf{4} & 0 & 0 \\
\mathbf{5} & \mathbf{5} & \mathbf{5} & 0 & 0 \\
0 & \mathbf{2} & 0 & \mathbf{4} & \mathbf{4} \\
0 & 0 & 0 & \mathbf{5} & \mathbf{5} \\
0 & \mathbf{1} & 0 & \mathbf{2} & \mathbf{2}
\end{bmatrix}
\approx
\begin{bmatrix}
\mathbf{0.13} & 0.02 & \cancel{-0.01} \\
\mathbf{0.41} & 0.07 & \cancel{-0.03} \\
\mathbf{0.55} & 0.09 & \cancel{-0.04} \\
\mathbf{0.68} & 0.11 & \cancel{-0.05} \\
0.15 & \mathbf{-0.59} & \cancel{\mathbf{0.65}} \\
0.07 & \mathbf{-0.73} & \cancel{\mathbf{-0.67}} \\
0.07 & \mathbf{-0.29} & \cancel{\mathbf{0.32}}
\end{bmatrix}
\times
\begin{bmatrix}
\mathbf{12.4} & 0 & 0 \\
0 & \mathbf{9.5} & 0 \\
0 & 0 & \cancel{\mathbf{1.3}}
\end{bmatrix}
\times
\begin{bmatrix}
\mathbf{0.56} & \mathbf{0.59} & \mathbf{0.56} & 0.09 & 0.09 \\
0.12 & -0.02 & 0.12 & \mathbf{-0.69} & \mathbf{-0.69} \\
\cancel{0.40} & \cancel{\mathbf{-0.80}} & \cancel{0.40} & \cancel{0.09} & \cancel{0.09}
\end{bmatrix}
$$

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix} \approx \begin{bmatrix} \mathbf{0.13} & 0.02 \\ \mathbf{0.41} & 0.07 \\ \mathbf{0.55} & 0.09 \\ \mathbf{0.68} & 0.11 \\ 0.15 & \mathbf{-0.59} \\ 0.07 & \mathbf{-0.73} \\ 0.07 & \mathbf{-0.29} \end{bmatrix} \times \begin{bmatrix} \mathbf{12.4} & 0 \\ 0 & \mathbf{9.5} \end{bmatrix} \times \begin{bmatrix} \mathbf{0.56} & \mathbf{0.59} & \mathbf{0.56} & 0.09 & 0.09 \\ 0.12 & -0.02 & 0.12 & \mathbf{-0.69} & \mathbf{-0.69} \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix} \approx \begin{bmatrix} \mathbf{0.92} & \mathbf{0.95} & \mathbf{0.92} & 0.01 & 0.01 \\ \mathbf{2.91} & \mathbf{3.01} & \mathbf{2.91} & -0.01 & -0.01 \\ \mathbf{3.90} & \mathbf{4.04} & \mathbf{3.90} & 0.01 & 0.01 \\ \mathbf{4.82} & \mathbf{5.00} & \mathbf{4.82} & 0.03 & 0.03 \\ 0.70 & \mathbf{0.53} & 0.70 & \mathbf{4.11} & \mathbf{4.11} \\ -0.69 & 1.34 & -0.69 & \mathbf{4.78} & \mathbf{4.78} \\ 0.32 & \mathbf{0.23} & 0.32 & \mathbf{2.01} & \mathbf{2.01} \end{bmatrix}$$

Best Low Rank Approx.



Complexity to compute SVD:

- **O(nm^2)** or **O(n^2m)** (whichever is less)
- **But** Less work,

        - if we just want singular values
        - if we want first *k* singular vectors
        - if the matrix is sparse
- **Implemented in** linear algebra packages like
    - LINPACK, Matlab, SPlus, Mathematica ...

Summary

- Dimensionality reduction is essential for any kind of data analysis involving big data
- Dimensionality reductions need basic knowledge of mathematics
- Explained with the example of SVD