CS3319 Foundations of Data Science

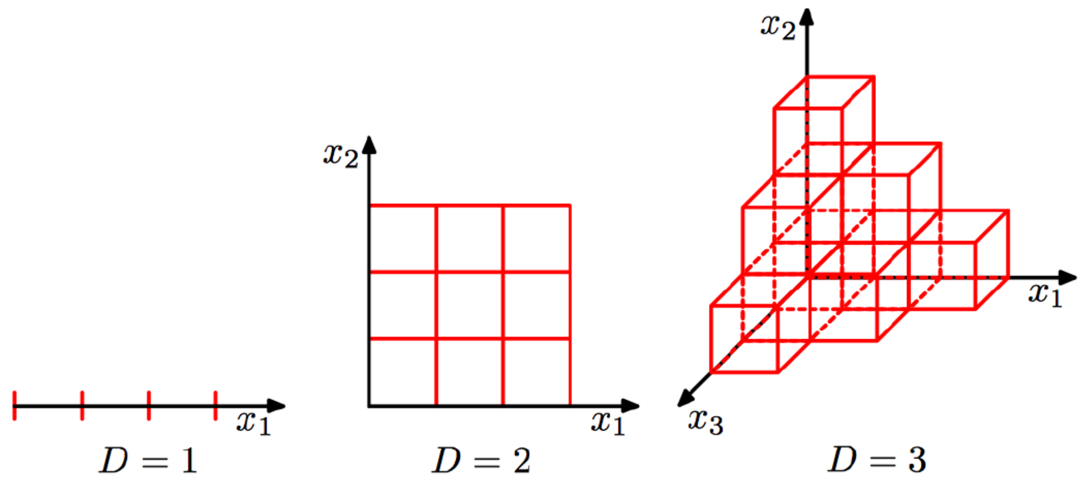
7. Dimensionality Reduction

Jiaxin Ding
John Hopcroft Center

上海交通大学
约翰·霍普克罗夫特计算机科学中心
John Hopcroft Center for Computer Science

Curse of Dimensionality

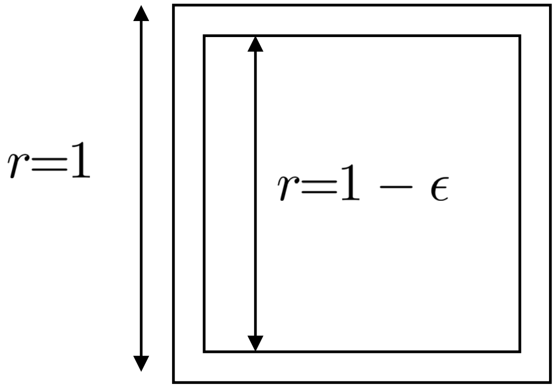
- The high dimensional spaces are empty
  - The sample sizes to cover the space grow exponentially with the dimension

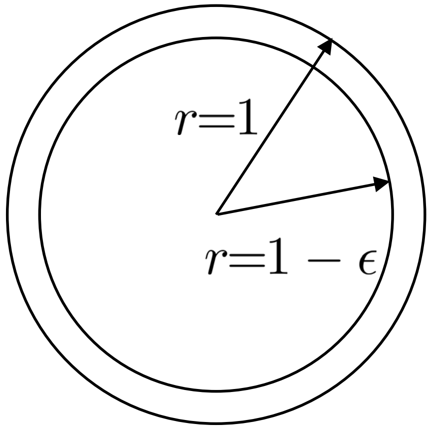


Curse of Dimensionality

- Most of its volume is near the surface in high dimensional spaces
  - Given a d-dim volume, shrink this volume by a small amount ϵ

hypercube $V_d(r) = r^d$ hypersphere $V_d(r) = \frac{2r^d \pi^{\frac{d}{2}}}{d\Gamma(\frac{d}{2})}$

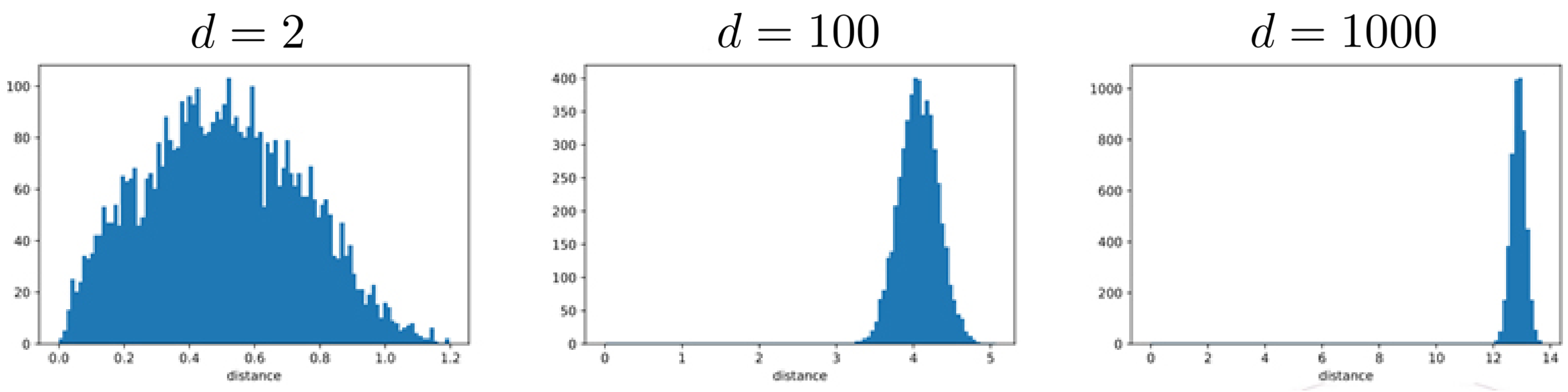




$$\lim_{d \to \infty} \frac{V_d(1-\epsilon)}{V_d(1)} = \lim_{d \to \infty} (1-\epsilon)^d = 0$$

Curse of Dimensionality

- Points are isolated in high dimensional spaces
  - x, y are two independent variables, with uniform distribution on $[0,1]^d$, their mean Euclidean distance satisfies $E\left(\|x-y\|_{\ell_2}\right)=\sqrt{\frac{d}{6}}$

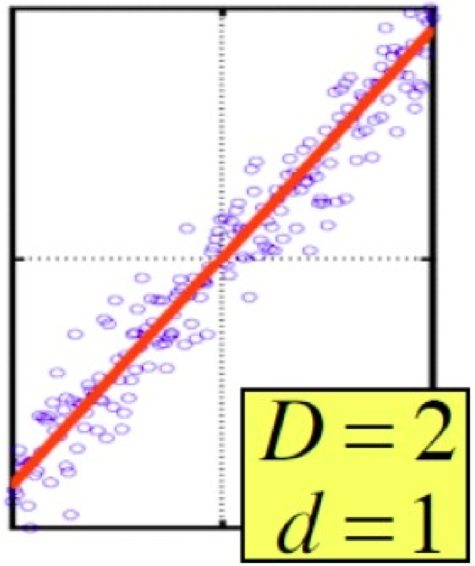


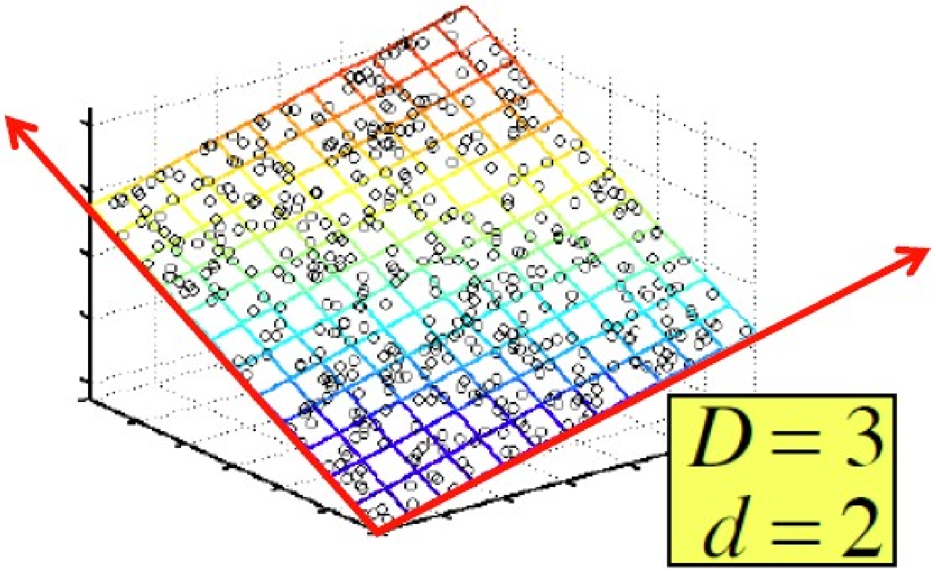
Curse of dimensionality

- Sample space is large
- Almost all points are near the surface
- Distance metric starts losing their effectiveness

Solution: **dimensionality reduction**

Dimensionality Reduction

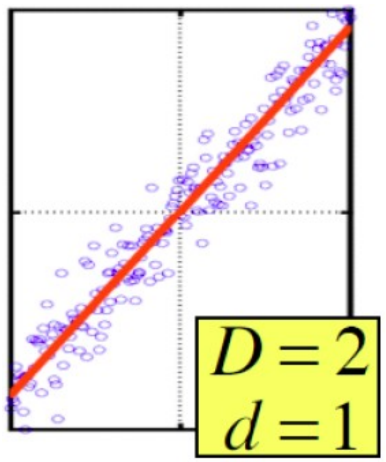


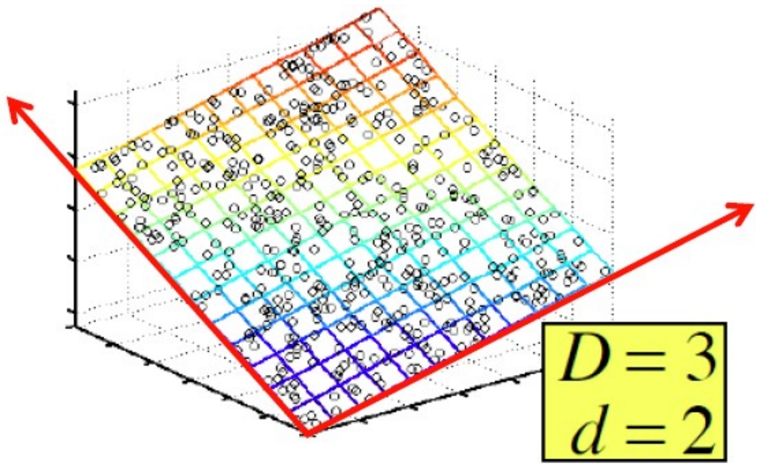


- There are hidden, or **latent factors, latent dimensions** that – to a close approximation – explain why the values are as they appear in the data matrix.
- **Goal of dimensionality reduction is to discover the axes of data!**

Dimensionality Reduction

- **The axes of these dimensions can be chosen by:**
  - The first dimension is the direction in which the points exhibit the **greatest variance**.
  - The second dimension is the direction, **orthogonal** to the first, in which points show the $\mathbf{2^{nd}}$ **greatest variance**.
  - And so on…, until you have enough dimensions that variance is really low.





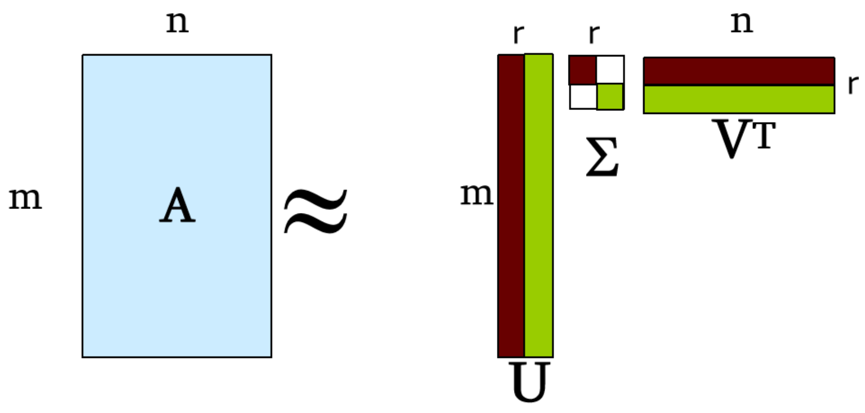
SVD: Singular Value Decomposition

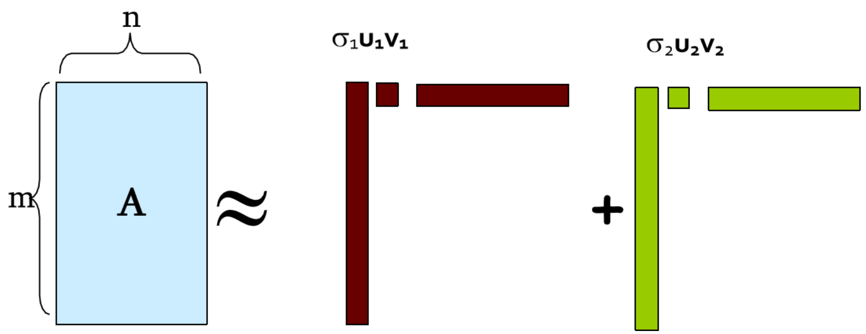
SVD

It is **always** possible to decompose a real matrix $\boldsymbol{A}$ into

$$\mathbf{A} \approx \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_i \sigma_i \mathbf{u}_i \circ \mathbf{v}_i$$

where





- $\boldsymbol{U}, \sum, \boldsymbol{V}$: unique
- A: **Input data matrix**
  - $m \times n$ matrix;
- U: **Left singular vectors**
  - $m \times r$ matrix; column orthonormal $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$
- $\sum$: **Singular values**
  - $r \times r$ diagonal matrix, r : rank of the matrix **A**
  - Entries are non-negative, and sorted in decreasing order($\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$)
- V: **Right singular vectors**
  - $n \times r$ matrix; column orthonormal $\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}$

How to Compute SVD

- First we need a method for finding the **principal eigenvalue** (the largest one) and the corresponding **eigenvector** of a symmetric matrix
  - $\boldsymbol{M}$ is *symmetric* if $m_{ij} = m_{ji}$ for all i and j.
- **Method:**
  - Start with any random eigenvector $\boldsymbol{x_0}$
  - Construct $\boldsymbol{x_{k+1}} = \frac{\boldsymbol{Mx_k}}{\|\boldsymbol{Mx_k}\|}$ for $k = 0,1,\ldots$
    - $\|\ldots\|$ denotes the Frobenius norm
  - Stop when consecutive $\boldsymbol{x_k}$ show little change

Example: Iterative Eigenvector

$$M = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \qquad \mathbf{x}_0 = \begin{matrix} 1 \\ 1 \end{matrix}$$

$$\frac{M\mathbf{x}_0}{||M\mathbf{x}_0||} = \begin{bmatrix} 3 \\ 5 \end{bmatrix} / \sqrt{34} = \begin{bmatrix} 0.51 \\ 0.86 \end{bmatrix} = \mathbf{x}_1$$

$$\frac{M\mathbf{x}_1}{||M\mathbf{x}_1||} = \begin{bmatrix} 2.23 \\ 3.60 \end{bmatrix} / \sqrt{17.93} = \begin{bmatrix} 0.53 \\ 0.85 \end{bmatrix} = \mathbf{x}_2$$

Finding the Principal Eigenvalue

- Once you have the principal eigenvector $\boldsymbol{x}$, you find its eigenvalue λ by $\lambda = \boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}$.
  - We know $\boldsymbol{x}\lambda = \boldsymbol{M}\boldsymbol{x}$ if λ is the eigenvalue; multiply both sides by $\boldsymbol{x}^T$ on the left.
  - Since $\boldsymbol{x}^T\boldsymbol{x} = \mathbf{1}$, we have $\lambda = \boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}$.

- **Example:** If we take $\boldsymbol{x}^T = [0.53, 0.85]$, then

$$\lambda = [0.53\ 0.85] \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 0.53 \\ 0.85 \end{bmatrix} = 4.25$$

Finding More Eigenpairs

- Eliminate the portion of the matrix that can be generated by the first eigenpair, $\boldsymbol{\lambda}$ and $\boldsymbol{x}$:

$$\boldsymbol{M}^* := \boldsymbol{M} - \lambda \boldsymbol{x}\boldsymbol{x}^T$$

- Recursively find the principal eigenpair for $\boldsymbol{M}^*$, eliminate the effect of that pair, and so on.
- Example:

$$\boldsymbol{M}^* = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} - 4.25 \begin{bmatrix} 0.53 \\ 0.85 \end{bmatrix} [0.53\ 0.85] = \begin{bmatrix} -0.19 & 0.09 \\ 0.09 & 0.07 \end{bmatrix}$$

How to Compute SVD

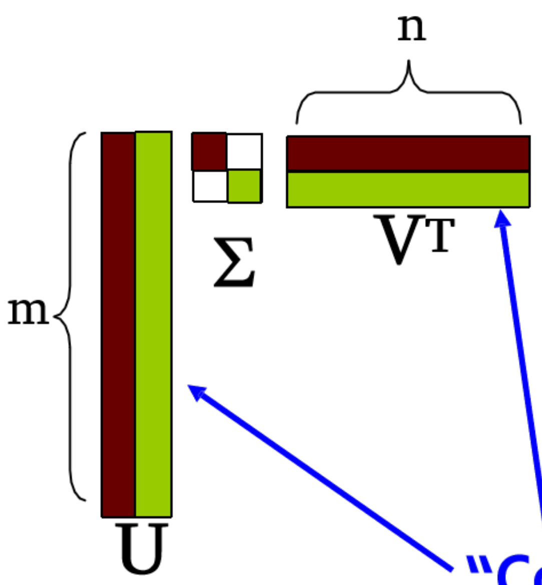
- Start by $\boldsymbol{A} = \boldsymbol{U}\sum\boldsymbol{V}^{\boldsymbol{T}}$
- $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A} = \boldsymbol{V}\sum\boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{U}\sum\boldsymbol{V}^{\boldsymbol{T}} = \boldsymbol{V}\sum^{\boldsymbol{2}}\boldsymbol{V}^{\boldsymbol{T}}$
  - $\boldsymbol{U}$ is orthonormal, so $\boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{U}$ is an identity matrix.
  - Also note that $\sum^{2}$ is a diagonal matrix whose i-th element is the square of the i -th element of $\sum$.
- $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}\boldsymbol{V} = \boldsymbol{V}\sum^{\boldsymbol{2}}\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{V} = \boldsymbol{V}\sum^{\boldsymbol{2}}$
  - $\boldsymbol{V}$ is also orthonormal.
  - **Note** that therefore the i-th column of $\boldsymbol{V}$ is an eigenvector of $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$, and its eigenvalue is the i-th element of $\sum^{\boldsymbol{2}}$.

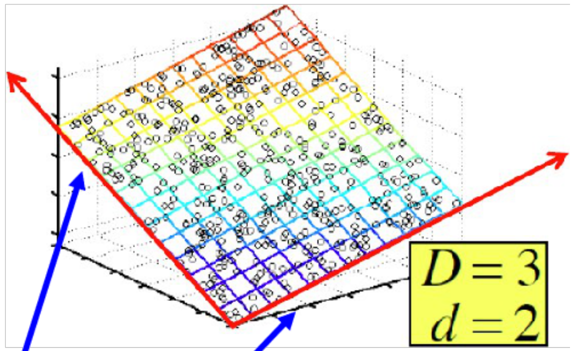
- Symmetric argument, $\boldsymbol{A}\boldsymbol{A}^{\boldsymbol{T}}$ gives us $\boldsymbol{U}$.

Interpreting SVD

- **What does SVD do?**

$$
\begin{array}{c|ccccc}
 & \text{Matrix} & \text{Alien} & \text{Serenity} & \text{Casablanca} & \text{Amelie} \\
\hline
\text{SciFi} & & & & & \\
\text{Romance} & & & & &
\end{array}
\qquad
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
3 & 3 & 3 & 0 & 0 \\
4 & 4 & 4 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 2 & 0 & 4 & 4 \\
0 & 0 & 0 & 5 & 5 \\
0 & 1 & 0 & 2 & 2
\end{bmatrix}
=
$$





Ratings matrix where each column corresponds to a movie and each row to a user. First 4 users prefer SciFi, while others prefer Romance.

"Concepts"
AKA Latent dimensions
AKA Latent factors

SVD-Example: Users-to-Movies

- $A = U\Sigma V^T$-example: Users to Movies

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

SVD-Example: Users-to-Movies

U is "user-to-concept" factor matrix

SciFi-concept

Romance-concept

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

SVD-Example: Users-to-Movies

SciFi-concept

“strength” of the SciFi-concept

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

SVD-Example: Users-to-Movies

V is "movie-to-concept" factor matrix

SciFi-concept

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

SciFi-concept

SVD-Interpretation #1

- **Movies, users and concepts:**
  - $\boldsymbol{U}$: user-to-concept matrix

  - $\boldsymbol{V}$: movie-to-concept matrix

  - $\sum$: its diagonal elements: 'strength' of each concept

SVD-Interpretation #2

- More details
  - Q: Further dimension reduction?

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

SVD-Interpretation #2

- **More details**
  - Q: Further dimension reduction?
  - **A: Set smallest singular values to zero.**

This is Rank 2 approximation to A. We could also do Rank 1 approx.
The larger the rank the more accurate the approximation.

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

SVD-Interpretation #2

- **More details**
  - **Q:** Further dimension reduction?
  - **A: Set smallest singular values to zero.**

This is Rank 2 approximation to A. We could also do Rank 1 approx.
The larger the rank the more accurate the approximation.

Matrix A

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

Reconstructed matrix B

Reconstruction Error is quantified by the Frobenius norm:

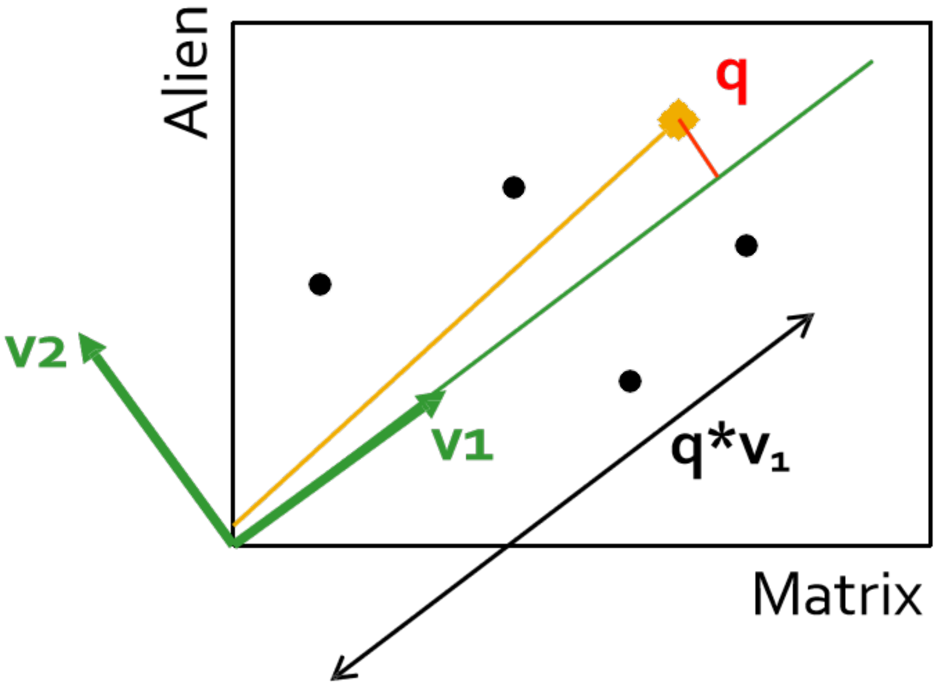
$$\|A - B\|_F = \sqrt{\sum_{ij} (A_{ij} - B_{ij})^2} = \sqrt{tr((\Sigma_A - \Sigma_B)^2)}$$

Case Study: How to query?

- Q: A user likes `Matrix', what is her/his taste?
- A: Map query into a 'concept space' – how?

$$\mathbf{q} = \begin{array}{c} \begin{array}{ccccc} \text{Matrix} & \text{Alien} & \text{Serenity} & \text{Casablanca} & \text{Amelie} \end{array} \\ \begin{bmatrix} 5 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Project into concept space:
Inner product with each 'concept' vector $\mathbf{v_i}$



Case Study: Are they similar?

- **Observation:** For user $\boldsymbol{d}$ that rated (‘Alien’, ‘Serenity’) and user $\boldsymbol{q}$ that rated (‘Matrix’), are they similar or not? How to measure?

$$\begin{array}{cc} & \begin{array}{ccccc}\text{Matrix} & \text{Alien} & \text{Serenity} & \text{Casablanca} & \text{Amelie}\end{array} \\ \mathbf{d} = & \begin{bmatrix} 0 & 4 & 5 & 0 & 0 \end{bmatrix} \\ \mathbf{q} = & \begin{bmatrix} 5 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Zero ratings in common

Case Study: How to query?

- Compactly, we have:

$$q_{concept} = qV$$

- E.g.:

$$\mathbf{q} = \begin{array}{ccccc} \text{Matrix} & \text{Alien} & \text{Serenity} & \text{Casablanca} & \text{Amelie} \\ \end{array}$$

$$\mathbf{q} = \begin{bmatrix} 5 & 0 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 0.56 & 0.12 \\ 0.59 & -0.02 \\ 0.56 & 0.12 \\ 0.09 & -0.69 \\ 0.09 & -0.69 \end{bmatrix} = \begin{bmatrix} 2.8 & 0.6 \end{bmatrix}$$

movie-to-concept factors (V)

SciFi-concept

Case Study: How to query?

- How would the user d that rated ('Alien', 'Serenity') be handled?

$$d_{concept} = dV$$

- E.g.:

$$\mathbf{d} = \begin{array}{c} \begin{array}{ccccc} \text{Matrix} & \text{Alien} & \text{Serenity} & \text{Casablanca} & \text{Amelie} \end{array} \\ \begin{bmatrix} 0 & 4 & 5 & 0 & 0 \end{bmatrix} \end{array} \times \begin{bmatrix} 0.56 & 0.12 \\ 0.59 & -0.02 \\ 0.56 & 0.12 \\ 0.09 & -0.69 \\ 0.09 & -0.69 \end{bmatrix} = \begin{array}{c} \text{SciFi-concept} \\ \begin{bmatrix} 5.2 & 0.4 \end{bmatrix} \end{array}$$

movie-to-concept factors (V)

Case Study: How to query?

- **Observation:** User $\boldsymbol{d}$ that rated ('Alien', 'Serenity') will be **similar** to user $\boldsymbol{q}$ that rated ('Matrix'), although $\boldsymbol{d}$ and $\boldsymbol{q}$ have **zero ratings in common**!

$$
\begin{array}{cc}
 & \begin{array}{ccccc} \text{Matrix} & \text{Alien} & \text{Serenity} & \text{Casablanca} & \text{Amelie} \end{array} \\
\mathbf{d} = & \begin{bmatrix} 0 & 4 & 5 & 0 & 0 \end{bmatrix} \\
\mathbf{q} = & \begin{bmatrix} 5 & 0 & 0 & 0 & 0 \end{bmatrix}
\end{array}
\quad
\begin{array}{c}
\\
\dashrightarrow \\
\dashrightarrow
\end{array}
\quad
\begin{array}{cc}
\text{SciFi-concept} & \\
\downarrow & \\
\begin{bmatrix} 5.2 \end{bmatrix}. & \\
\end{array}
$$

$$
\mathbf{d} \dashrightarrow \begin{bmatrix} 5.2 & 0.4 \end{bmatrix}, \qquad \mathbf{q} \dashrightarrow \begin{bmatrix} 2.8 & 0.6 \end{bmatrix}
$$

Zero ratings in common

Similarity > 0