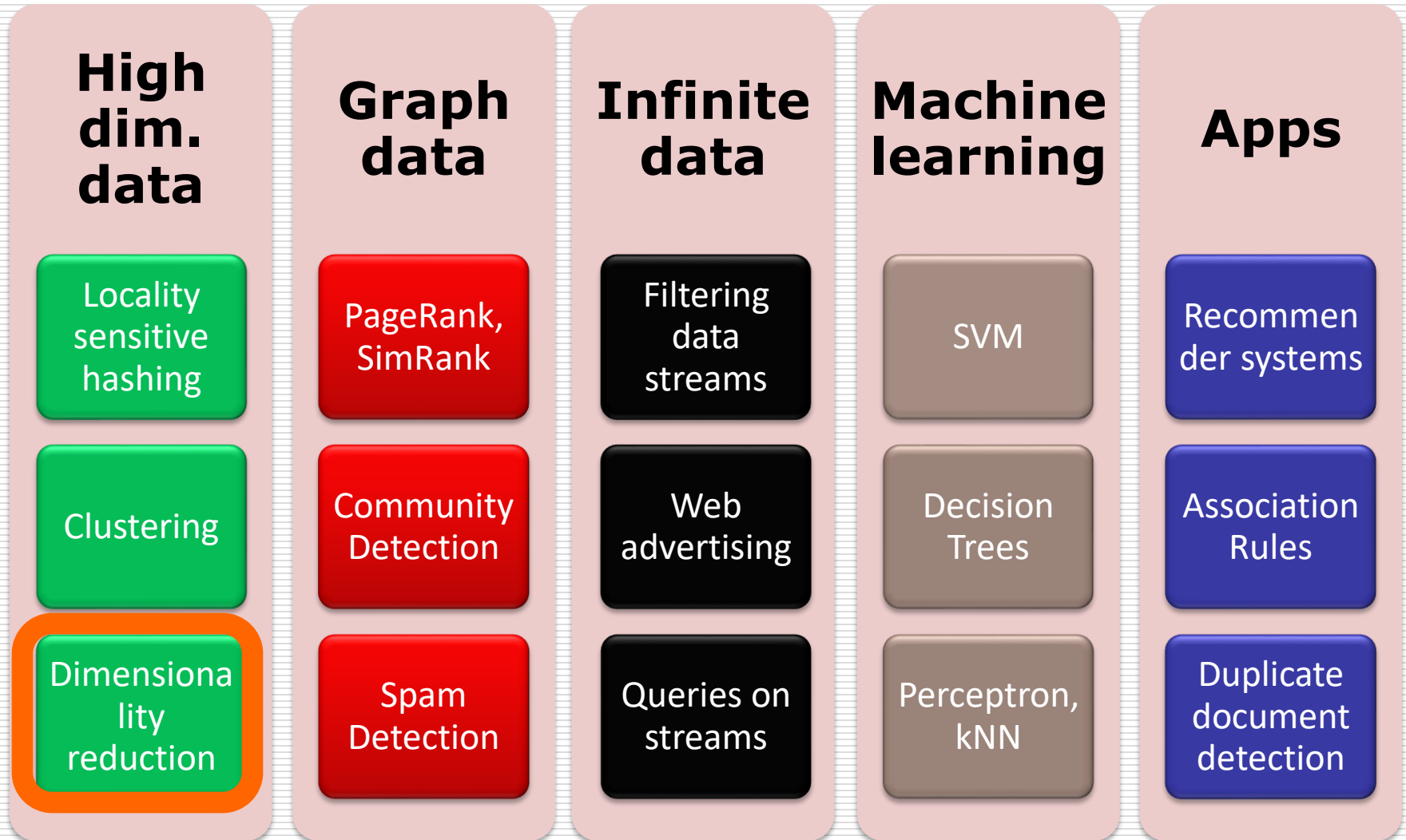


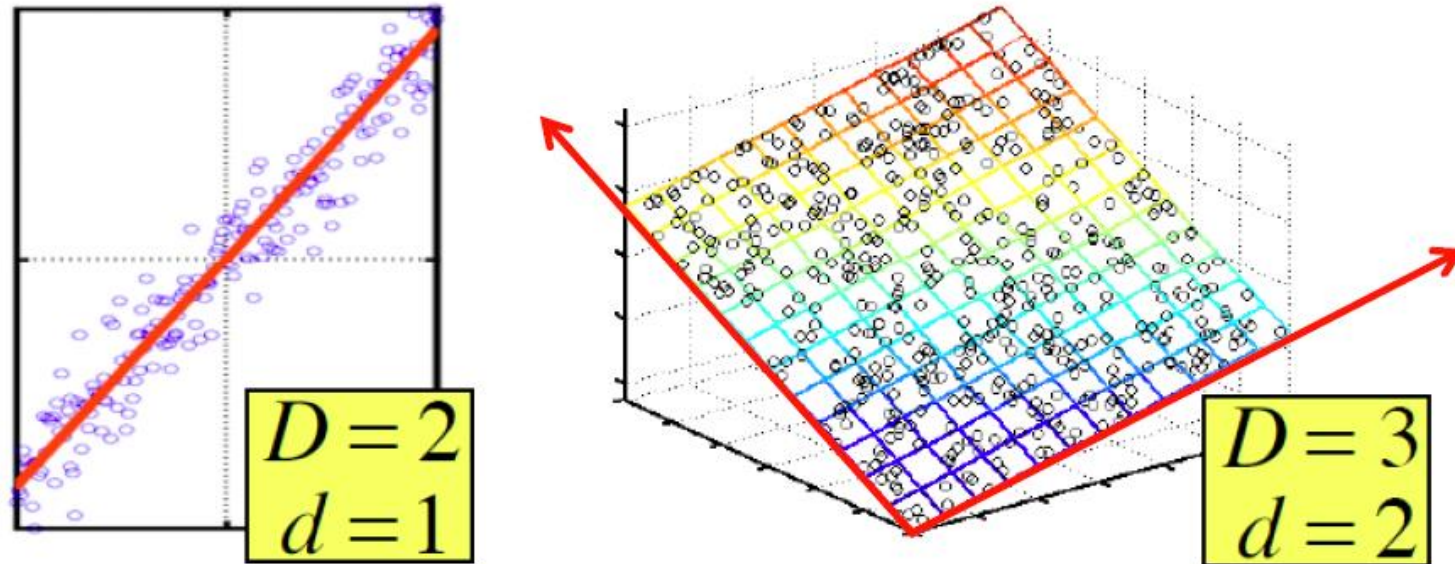
大数据计算及应用(八)

Dimensionality Reduction: SVD

Agenda



Dimensionality Reduction



- **Assumption:** Data lies on or near a low d -dimensional subspace
- **Axes of this subspace are effective representation of the data**

Dimensionality Reduction

□ Compress / reduce dimensionality:

- 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

□ **Q:** What is **rank** of a matrix **A**?

□ **A:** Number of **linearly independent** rows (columns) of **A**

□ **For example:**

■ Matrix $\mathbf{A} = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix}$ 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 \mathbf{A} as two “basis” vectors: $[-2 \ -3 \ 1] \ [3 \ 5 \ 0]$

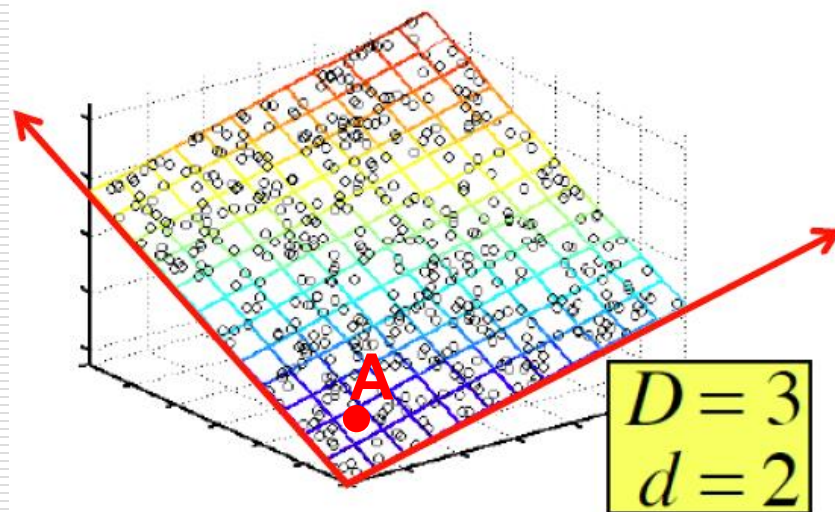
■ And new coordinates of : $[1 \ 1] \ [1 \ 0] \ [0 \ 1]$

Rank is “Dimensionality”

□ Cloud of points 3D space:

- Think of point positions as a matrix:

1 row per point:
$$\begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix} \begin{matrix} \text{A} \\ \text{B} \\ \text{C} \end{matrix}$$

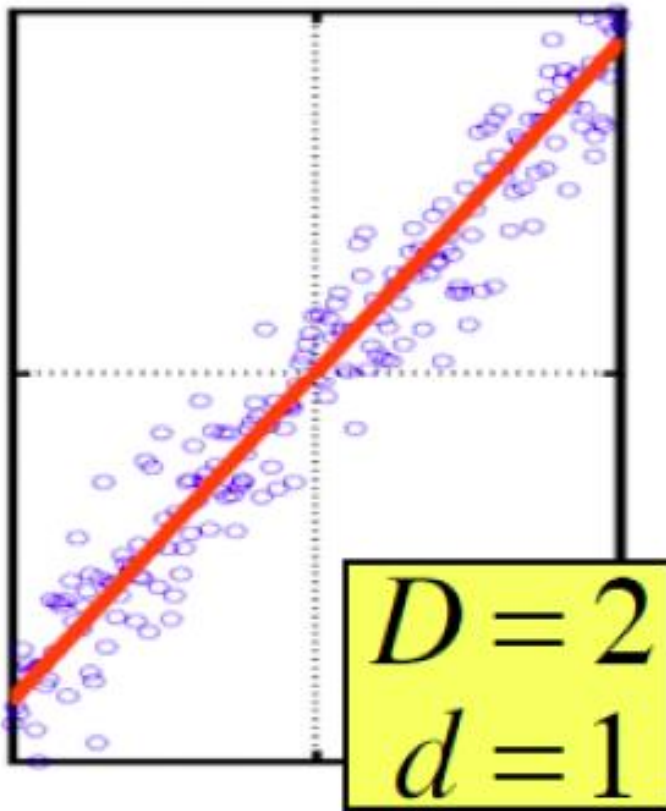


□ We can rewrite coordinates more efficiently!

- Old basis vectors: $[1 \ 0 \ 0] \ [0 \ 1 \ 0] \ [0 \ 0 \ 1]$
- New basis vectors: $[-2 \ -3 \ 1] \ [3 \ 5 \ 0]$
- Then **A** has new coordinates: $[1 \ 1]$. **B**: $[1 \ 0]$, **C**: $[0 \ 1]$
- Notice: We reduced the number of coordinates!

Dimensionality Reduction

- Goal of dimensionality reduction is to discover the axis of data!



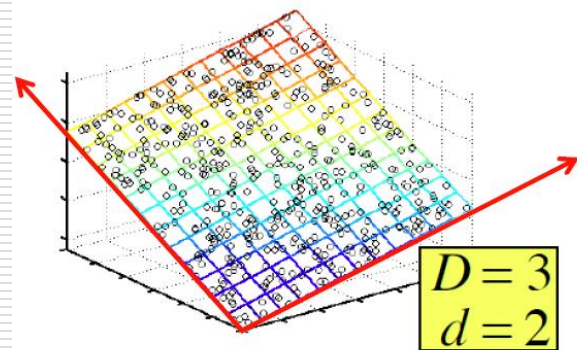
Rather than representing every point with 2 coordinates we represent each point with 1 coordinate (corresponding to the position of the point on the red line).

By doing this we incur a bit of error as the points do not exactly lie on the line

Why Reduce Dimensions?

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**



Singular Value Decomposition (SVD)

$$\mathbf{A}_{[m \times n]} = \mathbf{U}_{[m \times r]} \mathbf{\Sigma}_{[r \times r]} (\mathbf{V}_{[n \times r]})^T$$

□ **A: Input data matrix**

■ $m \times n$ matrix (e.g., m documents, n terms)

□ **U: Left singular vectors**

■ $m \times r$ matrix (m documents, r concepts)

□ **Σ : Singular values**

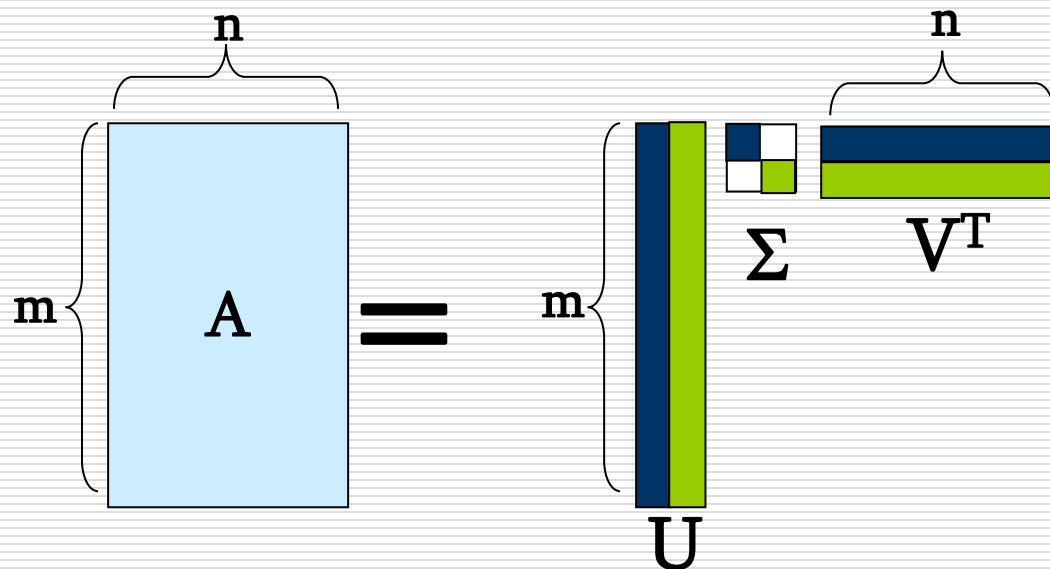
■ $r \times r$ diagonal matrix (strength of each 'concept')
(r : rank of the matrix **A**)

□ **V: Right singular vectors**

■ $n \times r$ matrix (n terms, r concepts)

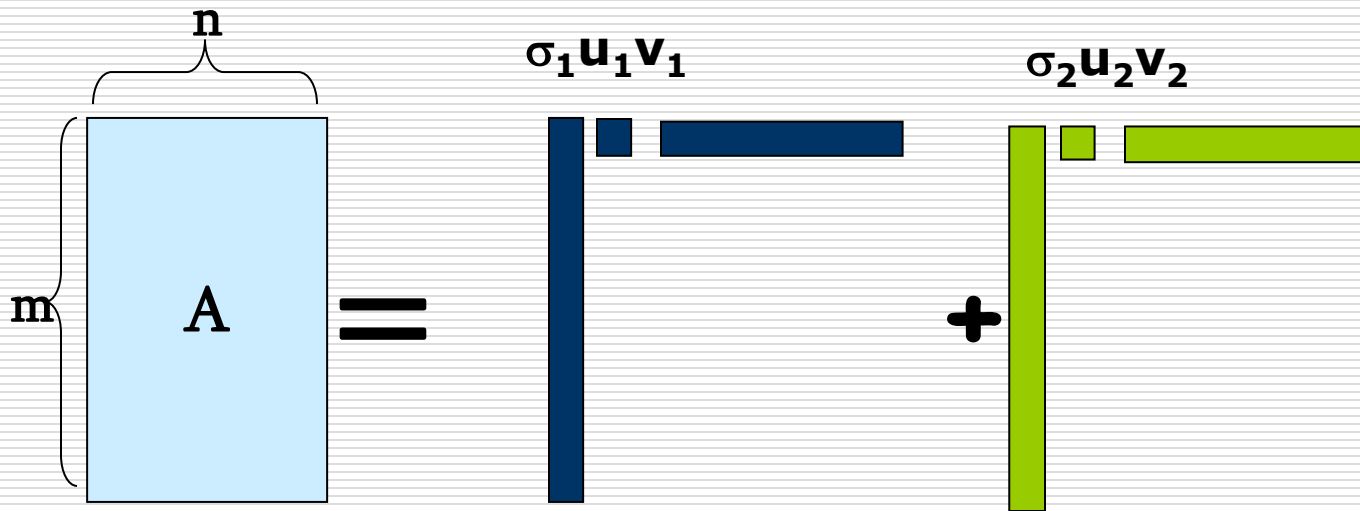
SVD

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



SVD

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



σ_i ... scalar
 \mathbf{u}_i ... vector
 \mathbf{v}_i ... vector

SVD - Properties

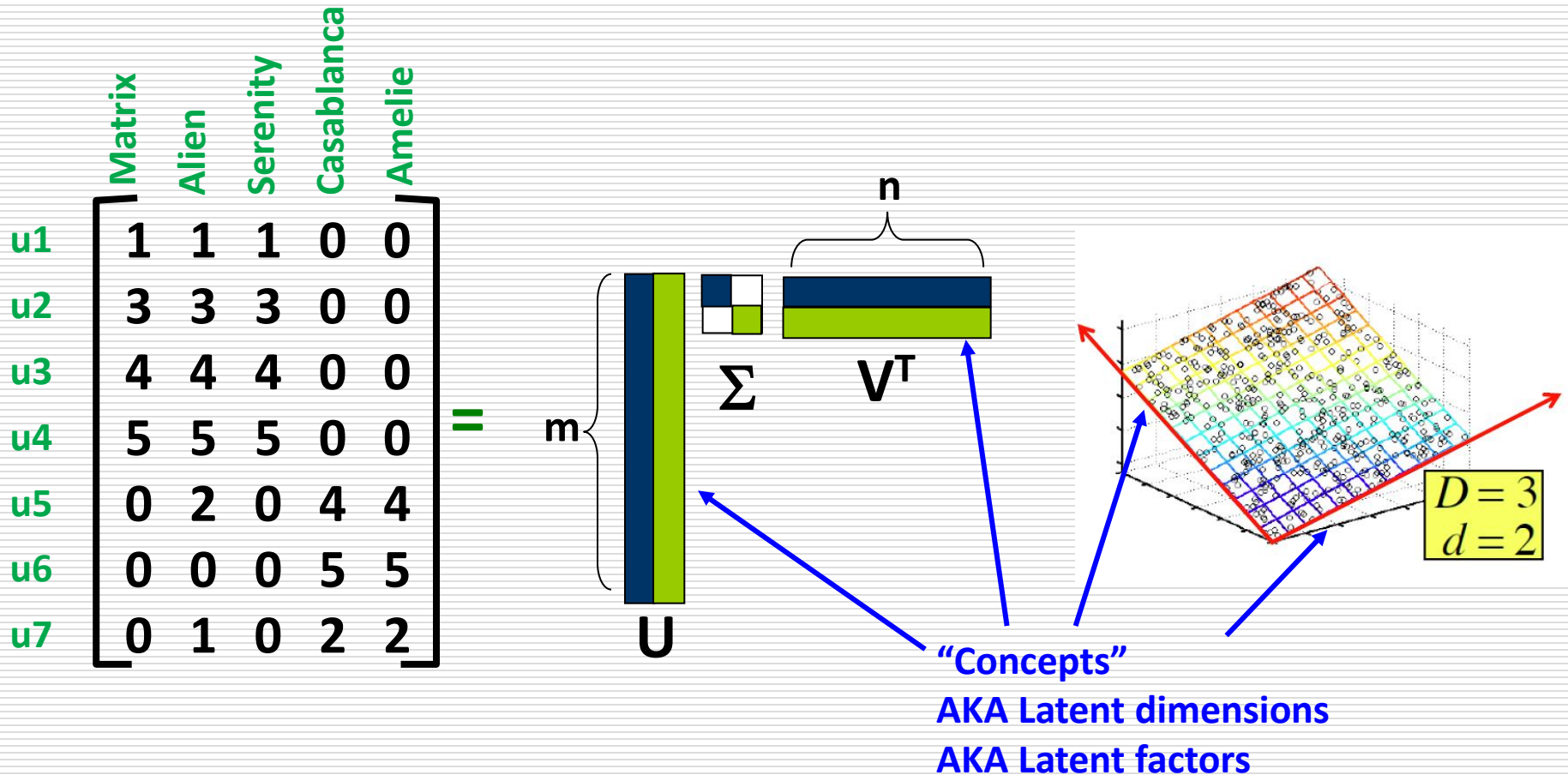
It is **always** possible to decompose a real matrix **A** into **$A = U \Sigma V^T$** , where

- **U, Σ , V: unique**
- **U, V: column orthonormal**
 - **$U^T U = I; V^T V = I$** (**I**: identity matrix)
 - (Columns are orthogonal unit vectors)
- **Σ : diagonal**
 - Entries (**singular values**) are **positive**,
and sorted in decreasing order (**$\sigma_1 \geq \sigma_2 \geq \dots \geq 0$**)

Nice proof of uniqueness: <http://www.mpi-inf.mpg.de/~bast/ir-seminar-ws04/lecture2.pdf>

SVD – Example: Users-to-Movies

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



SVD – Example: Users-to-Movies

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

$$\begin{array}{c} \text{u1} \\ \text{u2} \\ \text{u3} \\ \text{u4} \\ \text{u5} \\ \text{u6} \\ \text{u7} \end{array} \begin{array}{c} \text{Matrix} \\ \text{Alien} \\ \text{Serenity} \\ \text{Casablanca} \\ \text{Amelie} \end{array} \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix} = \begin{bmatrix} 0.13 & 0.02 & -0.01 \\ 0.41 & 0.07 & -0.03 \\ 0.55 & 0.09 & -0.04 \\ 0.68 & 0.11 & -0.05 \\ 0.15 & -0.59 & 0.65 \\ 0.07 & -0.73 & -0.67 \\ 0.07 & -0.29 & 0.32 \end{bmatrix} \times \begin{bmatrix} 12.4 & 0 & 0 \\ 0 & 9.5 & 0 \\ 0 & 0 & 1.3 \end{bmatrix} \times \begin{bmatrix} 0.56 & 0.59 & 0.56 & 0.09 & 0.09 \\ 0.12 & -0.02 & 0.12 & -0.69 & -0.69 \\ 0.40 & -0.80 & 0.40 & 0.09 & 0.09 \end{bmatrix}$$

SVD – Example: Users-to-Movies

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

$$\begin{array}{c}
 \begin{array}{c} \text{u1} \\ \text{u2} \\ \text{u3} \\ \text{u4} \\ \text{u5} \\ \text{u6} \\ \text{u7} \end{array}
 \begin{array}{c} \text{Matrix} \\ \text{Alien} \\ \text{Serenity} \\ \text{Casablanca} \\ \text{Amelie} \end{array}
 \begin{bmatrix}
 1 & 1 & 1 & 0 & 0 \\
 3 & 3 & 3 & 0 & 0 \\
 4 & 4 & 4 & 0 & 0 \\
 5 & 5 & 5 & 0 & 0 \\
 0 & 2 & 0 & 4 & 4 \\
 0 & 0 & 0 & 5 & 5 \\
 0 & 1 & 0 & 2 & 2
 \end{bmatrix}
 =
 \begin{array}{c}
 \text{SciFi-concept} \quad \text{Romance-concept} \\
 \downarrow \quad \swarrow \\
 \begin{bmatrix}
 0.13 & 0.02 & -0.01 \\
 0.41 & 0.07 & -0.03 \\
 0.55 & 0.09 & -0.04 \\
 0.68 & 0.11 & -0.05 \\
 0.15 & -0.59 & 0.65 \\
 0.07 & -0.73 & -0.67 \\
 0.07 & -0.29 & 0.32
 \end{bmatrix}
 \end{array}
 \times
 \begin{bmatrix}
 12.4 & 0 & 0 \\
 0 & 9.5 & 0 \\
 0 & 0 & 1.3
 \end{bmatrix}
 \times
 \begin{bmatrix}
 0.56 & 0.59 & 0.56 & 0.09 & 0.09 \\
 0.12 & -0.02 & 0.12 & -0.69 & -0.69 \\
 0.40 & -0.80 & 0.40 & 0.09 & 0.09
 \end{bmatrix}
 \end{array}$$

SVD – Example: Users-to-Movies

□ $A = U \Sigma V^T$ - example: U is “user-to-concept” similarity matrix

	Matrix	Alien	Serenity	Casablanca	Amelie		SciFi-concept	Romance-concept				
u1	1	1	1	0	0	=	0.13	0.02	-0.01	\times	$\begin{bmatrix} 12.4 & 0 & 0 \\ 0 & 9.5 & 0 \\ 0 & 0 & 1.3 \end{bmatrix}$	\times
u2	3	3	3	0	0		0.41	0.07	-0.03			
u3	4	4	4	0	0		0.55	0.09	-0.04			
u4	5	5	5	0	0		0.68	0.11	-0.05			
u5	0	2	0	4	4		0.15	-0.59	0.65			
u6	0	0	0	5	5		0.07	-0.73	-0.67			
u7	0	1	0	2	2		0.07	-0.29	0.32			
											$\begin{bmatrix} 0.56 & 0.59 & 0.56 & 0.09 & 0.09 \\ 0.12 & -0.02 & 0.12 & -0.69 & -0.69 \\ 0.40 & -0.80 & 0.40 & 0.09 & 0.09 \end{bmatrix}$	

SVD – Example: Users-to-Movies

□ $A = U \Sigma V^T$ - example:

Matrix Alien Serenity Casablanca Amelie

u1

u2

u3

u4

u5

u6

u7

SciFi-concept

“strength” of the SciFi-concept

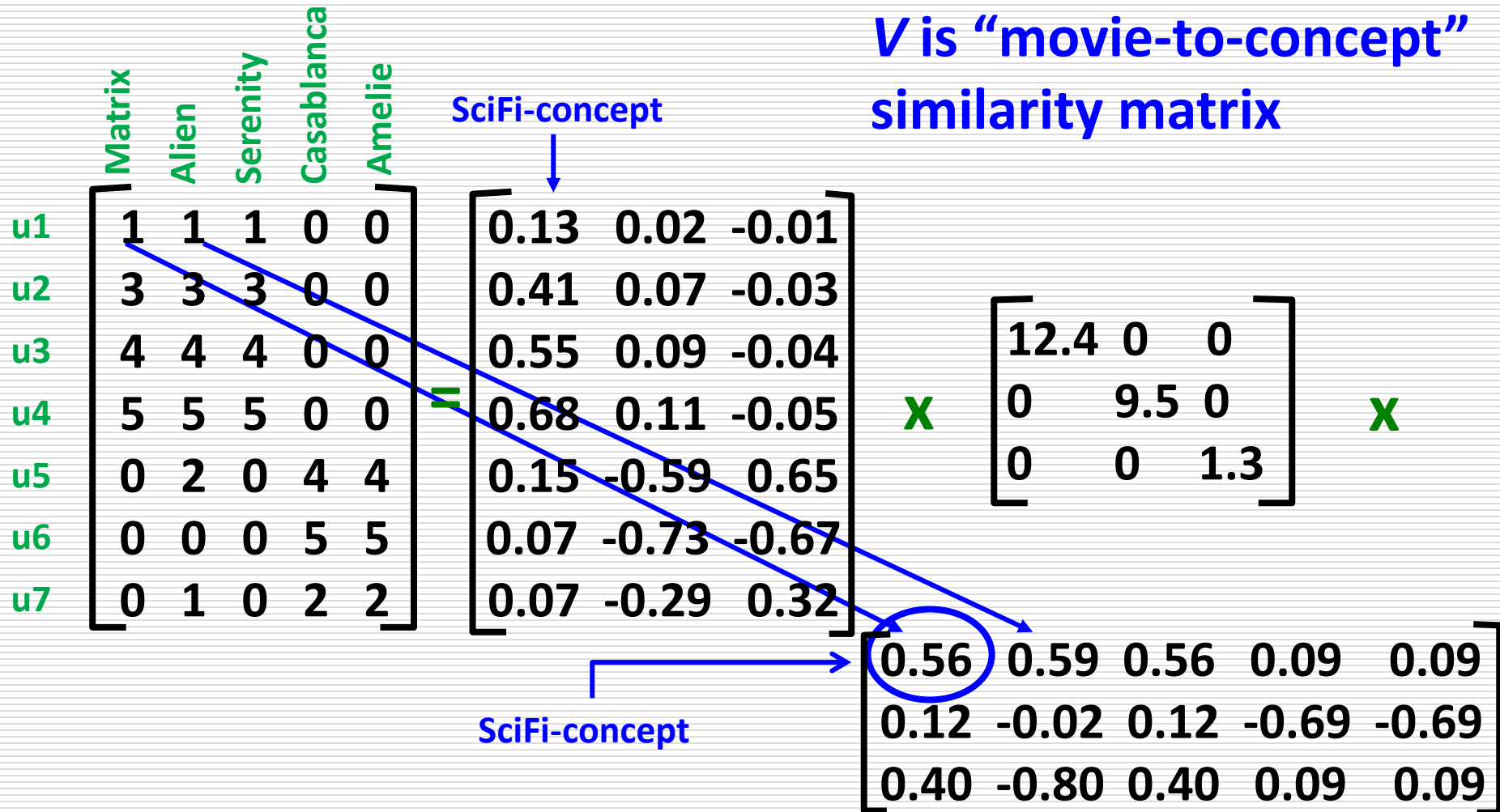
X

X

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

SVD – Example: Users-to-Movies

□ $A = U \Sigma V^T$ - example:



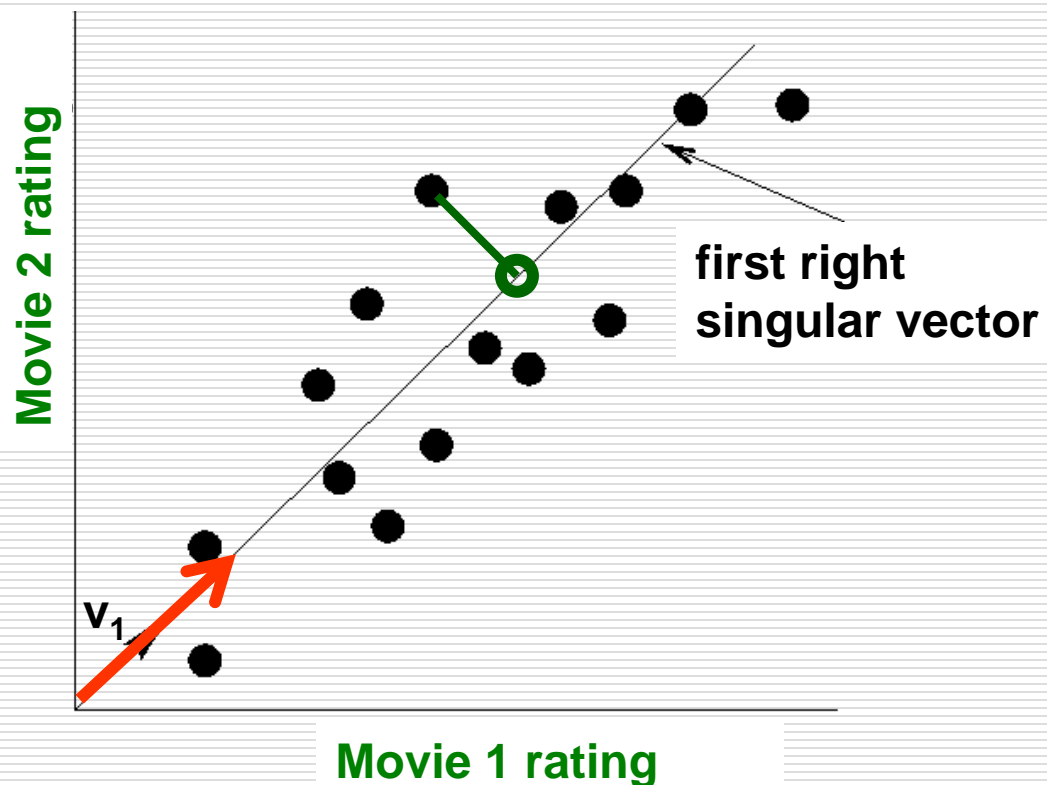
SVD - Interpretation #1

‘**movies**’, ‘**users**’ and ‘**concepts**’:

- U : user-to-concept similarity matrix
- V : movie-to-concept similarity matrix
- Σ : its diagonal elements:
‘strength’ of each concept

Dimensionality Reduction with SVD

SVD – Dimensionality Reduction



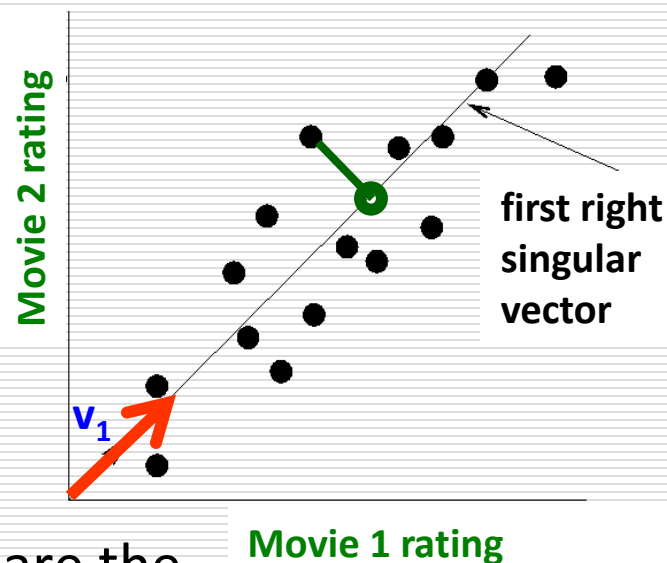
- ❑ Instead of using two coordinates (x, y) to describe point locations, let's use only one coordinate (z)
- ❑ Point's position is its location along vector v_1
- ❑ **How to choose v_1 ? Minimize reconstruction error**

SVD – Dimensionality Reduction

- **Goal:** Minimize the sum of reconstruction errors:

$$\sum_{i=1}^N \sum_{j=1}^D \|x_{ij} - z_{ij}\|^2$$

- where x_{ij} are the “old” and z_{ij} are the “new” coordinates



- SVD gives ‘best’ axis to project on:
 - ‘best’ = minimizing the reconstruction errors
- In other words, **minimum reconstruction error**

SVD - Interpretation #2

□ $A = U \Sigma V^T$ - example:

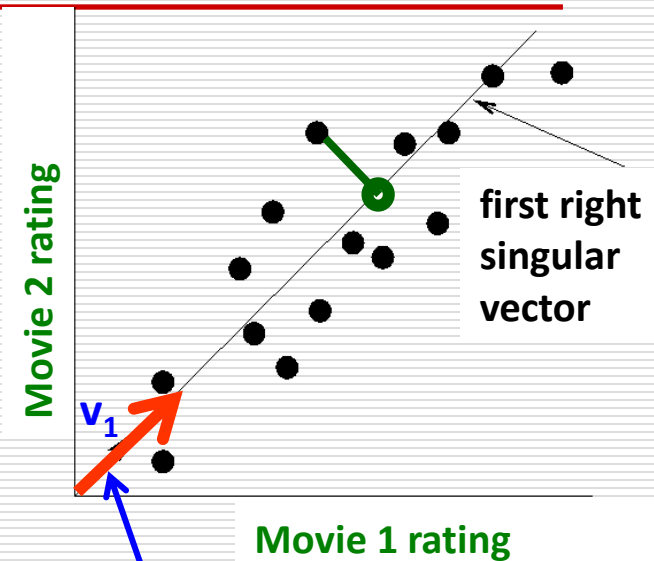
■ V : “movie-to-concept” matrix

■ U : “user-to-concept” matrix

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

$$\begin{bmatrix} 12.4 & 0 & 0 \\ 0 & 9.5 & 0 \\ 0 & 0 & 1.3 \end{bmatrix} \times$$

$$\begin{bmatrix} 0.56 & 0.59 & 0.56 & 0.09 & 0.09 \\ 0.12 & -0.02 & 0.12 & -0.69 & -0.69 \\ 0.40 & -0.80 & 0.40 & 0.09 & 0.09 \end{bmatrix}$$



SVD - Interpretation #2

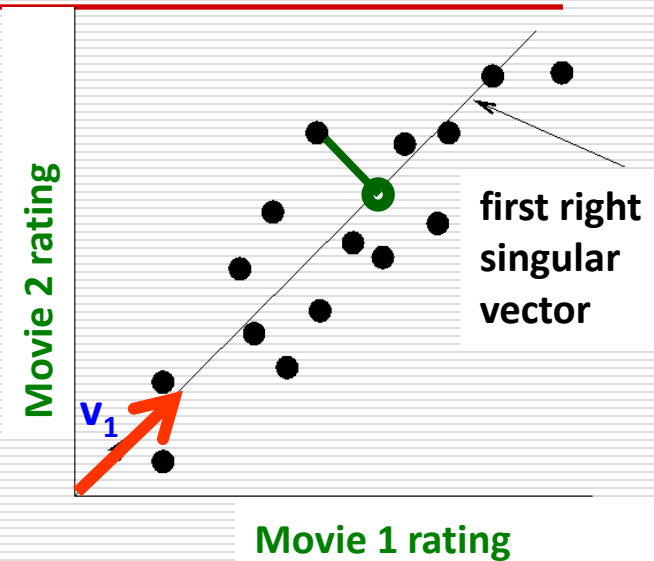
□ $A = U \Sigma V^T$ - example:

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

$$\begin{bmatrix} 12.4 & 0 & 0 \\ 0 & 9.5 & 0 \\ 0 & 0 & 1.3 \end{bmatrix} \times$$

$$\begin{bmatrix} 0.56 & 0.59 & 0.56 & 0.09 & 0.09 \\ 0.12 & -0.02 & 0.12 & -0.69 & -0.69 \\ 0.40 & -0.80 & 0.40 & 0.09 & 0.09 \end{bmatrix}$$

variance ('spread')
on the v_1 axis



SVD - Interpretation #2

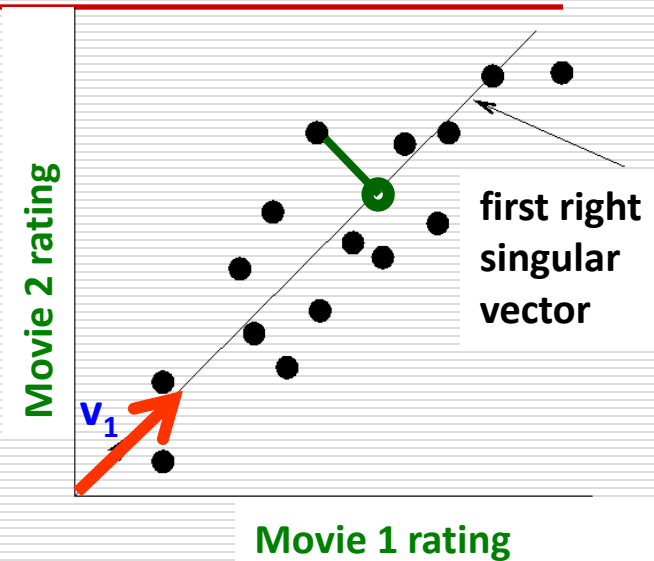
$A = U \Sigma V^T$ - example:

□ **$U \Sigma$:** Gives the coordinates of the points in the projection axis

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

Projection of users on the “Sci-Fi” axis $(U \Sigma)^T$:

1.61	0.19	-0.01
5.08	0.66	-0.03
6.82	0.85	-0.05
8.43	1.04	-0.06
1.86	-5.60	0.84
0.86	-6.93	-0.87
0.86	-2.75	0.41



SVD - Interpretation #2

More details

□ Q: How exactly is dim. reduction done?

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

SVD - Interpretation #2

More details

☐ **Q:** How exactly is dim. reduction done?

☐ **A:** Set smallest singular values to zero

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

SVD - Interpretation #2

More details

❑ **Q:** How exactly is dim. reduction done?

❑ **A:** Set smallest singular values to zero

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

SVD - Interpretation #2

More details

☐ **Q:** How exactly is dim. reduction done?

☐ **A:** Set smallest singular values to zero

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

The diagram illustrates the SVD decomposition of a matrix. The first matrix is a 7x5 matrix. It is approximated by the product of three matrices: a 7x3 matrix, a 3x3 diagonal matrix, and a 3x5 matrix. The 7x3 matrix and the 3x5 matrix have their third columns crossed out with red lines, indicating that the smallest singular values are being set to zero. The 3x3 diagonal matrix has its third diagonal element (1.3) crossed out with a red line, indicating that the smallest singular value is being set to zero. The green 'X' marks indicate the dimensionality reduction.

SVD - Interpretation #2

More details

☐ **Q:** How exactly is dim. reduction done?

☐ **A:** Set smallest singular values to zero

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

SVD - Interpretation #2

More details

❑ Q: How exactly is dim. reduction done?

❑ A: Set smallest singular values to zero

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

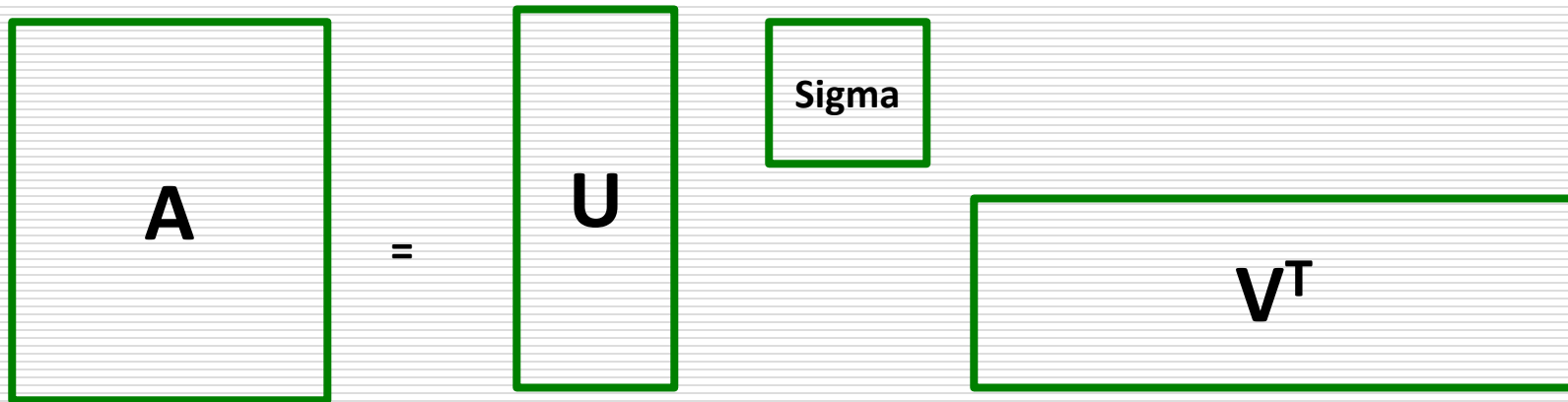
Frobenius norm:

$$\|M\|_F = \sqrt{\sum_{ij} M_{ij}^2}$$

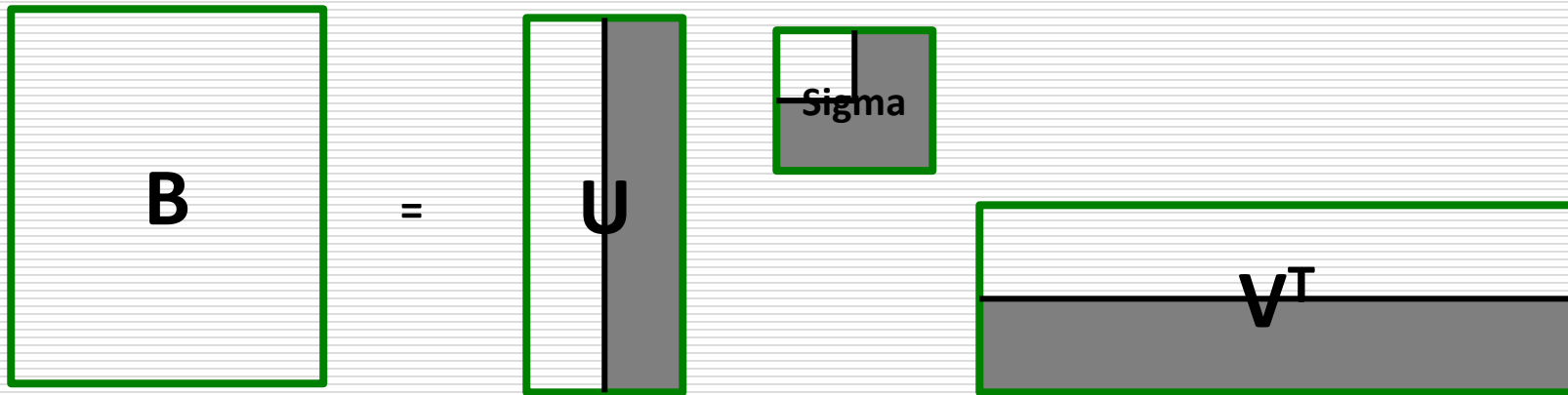
$$\|A-B\|_F = \sqrt{\sum_{ij} (A_{ij}-B_{ij})^2}$$

is “small”

SVD – Best Low Rank Approx.



B is best approximation of A



SVD – Best Low Rank Approx.

□ Theorem:

Let $A = U \Sigma V^T$ and $B = U S V^T$ where
 $S = \text{diagonal } r \times r \text{ matrix}$ with $s_i = \sigma_i$ ($i=1 \dots k$) else $s_i=0$
 then B is a **best** $\text{rank}(B)=k$ approx. to A

What do we mean by “best”:

■ B is a solution to $\min_B \|A-B\|_F$ where $\text{rank}(B)=k$

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & \\ \vdots & \vdots & \ddots & \\ x_{m1} & & & x_{mn} \end{pmatrix}_{m \times n} = \begin{pmatrix} u_{11} & \dots & & \\ \vdots & \ddots & & \\ u_{m1} & & & \end{pmatrix}_{m \times r} \begin{pmatrix} \sigma_{11} & 0 & \dots \\ 0 & \ddots & \\ \vdots & & \end{pmatrix}_{r \times r} \begin{pmatrix} v_{11} & \dots & v_{1n} \\ \vdots & \ddots & \\ & & \end{pmatrix}_{r \times n}$$

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

SVD – Best Low Rank Approx.

Details!

□ **Theorem:** Let $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$ ($\sigma_1 \geq \sigma_2 \geq \dots$, $\text{rank}(\mathbf{A}) = r$)

then $\mathbf{B} = \mathbf{U} \mathbf{S} \mathbf{V}^T$

■ \mathbf{S} = diagonal $r \times r$ matrix where $s_i = \sigma_i$ ($i=1 \dots k$) else $s_i = 0$
is a best rank- k approximation to \mathbf{A} :

■ \mathbf{B} is a solution to $\min_{\mathbf{B}} \|\mathbf{A} - \mathbf{B}\|_F$ where $\text{rank}(\mathbf{B}) = k$

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & \\ \vdots & \vdots & \ddots & \\ x_{m1} & & & x_{mn} \end{pmatrix}_{m \times n} = \begin{pmatrix} u_{11} & \dots & & \\ \vdots & \ddots & & \\ u_{m1} & & & \end{pmatrix}_{m \times r} \begin{pmatrix} \sigma_{11} & 0 & \dots \\ 0 & \ddots & \\ \vdots & & \end{pmatrix}_{r \times r} \begin{pmatrix} v_{11} & \dots & v_{1n} \\ \vdots & \ddots & \\ & & \end{pmatrix}_{r \times n}$$

□ **We will need:**

■ $\mathbf{U} \mathbf{\Sigma} \mathbf{V}^T - \mathbf{U} \mathbf{S} \mathbf{V}^T = \mathbf{U} (\mathbf{\Sigma} - \mathbf{S}) \mathbf{V}^T$

SVD – Best Low Rank Approx.

Details!

- $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$, $\mathbf{B} = \mathbf{U} \mathbf{S} \mathbf{V}^T$ ($\sigma_1 \geq \sigma_2 \geq \dots \geq 0$, $\text{rank}(\mathbf{A})=r$)
- \mathbf{S} = diagonal $n \times n$ matrix where $s_i = \sigma_i$ ($i=1 \dots k$) else $s_i=0$

then \mathbf{B} is solution to $\min_{\mathbf{B}} \|\mathbf{A} - \mathbf{B}\|_F$, $\text{rank}(\mathbf{B})=k$

■ Why?

$$\min_{\mathbf{B}, \text{rank}(\mathbf{B})=k} \|\mathbf{A} - \mathbf{B}\|_F = \min \|\mathbf{\Sigma} - \mathbf{S}\|_F = \min_{s_i} \sum_{i=1}^r (\sigma_i - s_i)^2$$

We used: $\mathbf{U} \mathbf{\Sigma} \mathbf{V}^T - \mathbf{U} \mathbf{S} \mathbf{V}^T = \mathbf{U} (\mathbf{\Sigma} - \mathbf{S}) \mathbf{V}^T$

- We want to choose s_i to minimize $\sum_i (\sigma_i - s_i)^2$
- Solution is to set $s_i = \sigma_i$ ($i=1 \dots k$) and other $s_i=0$

$$= \min_{s_i} \sum_{i=1}^k (\sigma_i - s_i)^2 + \sum_{i=k+1}^r \sigma_i^2 = \sum_{i=k+1}^r \sigma_i^2$$

SVD - Interpretation #2

Equivalent:

'spectral decomposition' of the matrix:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix} = \begin{bmatrix} | & | \\ u_1 & u_2 \\ | & | \end{bmatrix} \times \begin{bmatrix} \sigma_1 & \text{ } \\ \text{ } & \sigma_2 \end{bmatrix} \times \begin{bmatrix} \text{ } & v_1 \\ \text{ } & v_2 \end{bmatrix}$$

The diagram illustrates the Spectral Decomposition of a matrix. The matrix is decomposed into three components: a matrix of left singular vectors u_1 and u_2 , a diagonal matrix of singular values σ_1 and σ_2 , and a matrix of right singular vectors v_1 and v_2 . The singular values are represented by circles with diagonal lines through them, indicating they are scalars. The vectors u_1 , u_2 , v_1 , and v_2 are represented by vertical and horizontal lines, indicating they are column and row vectors respectively.

SVD - Interpretation #2

Equivalent:

‘spectral decomposition’ of the matrix

$$\begin{array}{c} \updownarrow n \\ \left[\begin{array}{ccccc} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{array} \right] \end{array} = \overbrace{\sigma_1 \underbrace{u_1}_{n \times 1} \underbrace{v_1^T}_{1 \times m}}^{k \text{ terms}} + \sigma_2 u_2 v_2^T + \dots$$

Assume: $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots \geq 0$

Why is setting small σ_i to 0 the right thing to do?
Vectors u_i and v_i are unit length, so σ_i scales them.
So, zeroing small σ_i introduces less error.

SVD - Interpretation #2

Q: How many σ s to keep?

A: Rule-of-a thumb:

keep 80-90% of 'energy' = $\sum_i \sigma_i^2$

$$\begin{array}{c} \begin{array}{c} \leftarrow m \rightarrow \\ \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix} \end{array} \\ \begin{array}{c} \uparrow \\ n \\ \downarrow \end{array} \end{array} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \dots$$

Assume: $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots$

Relation to Eigen-decomposition

□ SVD gives us:

- $A = U \Sigma V^T$

□ Eigen-decomposition:

- $M = X L X^T$

- M is symmetric

- U, V, X are orthonormal ($U^T U = I$),

- L, Σ are diagonal

□ Now let's calculate:

- $AA^T =$

- $A^T A = V \Sigma^T U^T (U \Sigma V^T) = V \Sigma \Sigma^T V^T$

Relation to Eigen-decomposition

□ SVD gives us:

■ $A = U \Sigma V^T$

□ Eigen-decomposition:

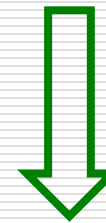
■ $M = X L X^T$

□ M is symmetric

□ U, V, X are orthonormal ($U^T U = I$),

□ L, Σ are diagonal

Shows how to compute
SVD using eigenvalue
decomposition!



□ Now let's calculate:

■ $AA^T = U \Sigma V^T (U \Sigma V^T)^T = U \Sigma V^T (V \Sigma^T U^T) = U \Sigma \Sigma^T U^T$

■ $A^T A = V \Sigma^T U^T (U \Sigma V^T) = V \Sigma \Sigma^T V^T$

$\begin{matrix} \uparrow & \uparrow & \uparrow \\ X & L & X^T \end{matrix}$

where $L = \Sigma^2$

An Example of SVD Computation

$$A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$A = U \Sigma V$$

$$A A^T = X L X^T = U \Sigma^2 U^T$$

$$A^T A = Y L Y^T = V \Sigma^2 V^T$$

$$A A^T = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Set $A A^T = M$, let λ be the eigenvalue of M , then we have

$$M X = \lambda X \Rightarrow (\lambda I - M) X = 0 \Rightarrow |\lambda I - M| = 0$$

I: identity matrix

$$\begin{vmatrix} \lambda - 2 & 1 & -1 \\ 1 & \lambda - 1 & 0 \\ -1 & 0 & \lambda - 1 \end{vmatrix} = \lambda(\lambda - 1)(\lambda - 3) = 0$$

$$\lambda_1 = 3, \lambda_2 = 1, \lambda_3 = 0$$

An Example of SVD Computation

$$A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \quad A A^T = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad \mathbf{A A^T = X L X^T = U \Sigma^2 U^T}$$

$$\lambda_1=3, \lambda_2=1, \lambda_3=0$$

When $\lambda_1=3$, we have $\mathbf{M X} = \lambda \mathbf{X}$

$$\left\{ \begin{array}{l} \begin{bmatrix} 2 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 3 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ x_1^2 + x_2^2 + x_3^2 = 1 \end{array} \right\} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{6}} \\ \frac{-1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix} = U_1$$

$$\lambda_1=3, U_1 = \begin{bmatrix} \frac{2}{\sqrt{6}} \\ \frac{-1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix} \quad \lambda_2=1, U_2 = \begin{bmatrix} \frac{0}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \quad \lambda_3=0, U_3 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{-1}{\sqrt{3}} \end{bmatrix}$$

An Example of SVD Computation

$$A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$A = U \Sigma V$$

$$A A^T = X L X^T = U \Sigma^2 U^T$$

$$A^T A = Y L Y^T = V \Sigma^2 V^T$$

$$A^T A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

$$\lambda_1=3, \lambda_2=1$$

$$\lambda_1=3, V_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{\sqrt{2}}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\lambda_2=1, V_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{\sqrt{2}}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

An Example of SVD Computation

$$A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$A = U \Sigma V$$

$$A A^T = X L X^T = U \Sigma^2 U^T$$

$$A^T A = Y L Y^T = V \Sigma^2 V^T$$

$$A = U \Sigma V$$

$$\begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{6}} & 0 \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 1 & 1 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

SVD - Complexity

- To compute SVD:

- $O(nm^2)$ or $O(n^2m)$ (whichever is less)

- But:

- Less work, if we just want singular values
- or if we want first k singular vectors
- or if the matrix is sparse

- Implemented in linear algebra packages like

- LINPACK, Matlab, SPlus, Mathematica ...

SVD - Conclusions so far

□ SVD: $A = U \Sigma V^T$: **unique**

- U : user-to-concept similarities
- V : movie-to-concept similarities
- Σ : strength of each concept

□ Dimensionality reduction:

- keep the few largest singular values (80-90% of 'energy')
- SVD: picks up linear correlations

Example of SVD & Conclusion

Case study: How to query?

- ❑ Q: Find users that like 'Matrix'
- ❑ A: Map query into a 'concept space' – how?

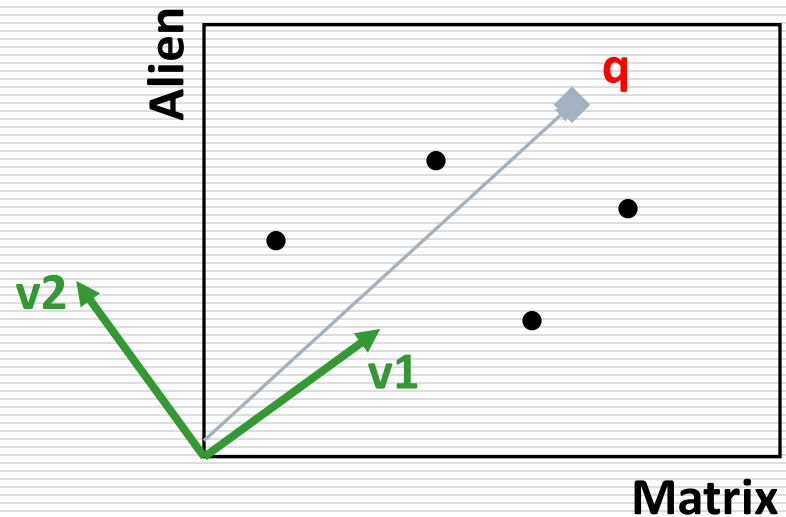
$$\begin{array}{c} \text{u1} \\ \text{u2} \\ \text{u3} \\ \text{u4} \\ \text{u5} \\ \text{u6} \\ \text{u7} \end{array} \begin{array}{c} \text{Matrix} \\ \text{Alien} \\ \text{Serenity} \\ \text{Casablanca} \\ \text{Amelie} \end{array} \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix} = \begin{bmatrix} 0.13 & 0.02 & -0.01 \\ 0.41 & 0.07 & -0.03 \\ 0.55 & 0.09 & -0.04 \\ 0.68 & 0.11 & -0.05 \\ 0.15 & -0.59 & 0.65 \\ 0.07 & -0.73 & -0.67 \\ 0.07 & -0.29 & 0.32 \end{bmatrix} \times \begin{bmatrix} 12.4 & 0 & 0 \\ 0 & 9.5 & 0 \\ 0 & 0 & 1.3 \end{bmatrix} \times \begin{bmatrix} 0.56 & 0.59 & 0.56 & 0.09 & 0.09 \\ 0.12 & -0.02 & 0.12 & -0.69 & -0.69 \\ 0.40 & -0.80 & 0.40 & 0.09 & 0.09 \end{bmatrix}$$

Case study: How to query?

- ❑ Q: Find users that like 'Matrix'
- ❑ A: Map query into a 'concept space' – how?

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

Project into concept space:
Inner product with each
'concept' vector v_i

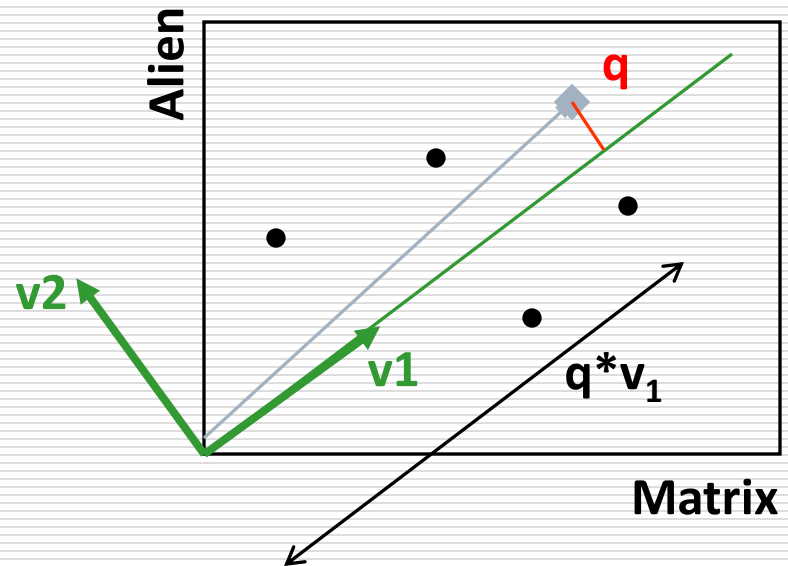


Case study: How to query?

- ❑ Q: Find users that like 'Matrix'
- ❑ A: Map query into a 'concept space' – how?

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

Project into concept space:
Inner product with each
'concept' vector \mathbf{v}_i



Case study: How to query?

Compactly, we have:

$$\mathbf{q}_{\text{concept}} = \mathbf{q} \mathbf{V}$$

E.g.:

$$\mathbf{q} = \begin{bmatrix} \text{Matrix} \\ \text{Alien} \\ \text{Serenity} \\ \text{Casablanca} \\ \text{Amelie} \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 & 0 & 0 \end{bmatrix} \mathbf{X} \begin{bmatrix} 0.56 & 0.12 \\ 0.59 & -0.02 \\ 0.56 & 0.12 \\ 0.09 & -0.69 \\ 0.09 & -0.69 \end{bmatrix} = \begin{bmatrix} 2.8 & 0.6 \end{bmatrix}$$

movie-to-concept similarities (V)

SciFi-concept

Case study: How to query?

□ How would the user d that rated ('Serenity') be handled?

$$\mathbf{d}_{\text{concept}} = \mathbf{d} \mathbf{V}$$

E.g.:

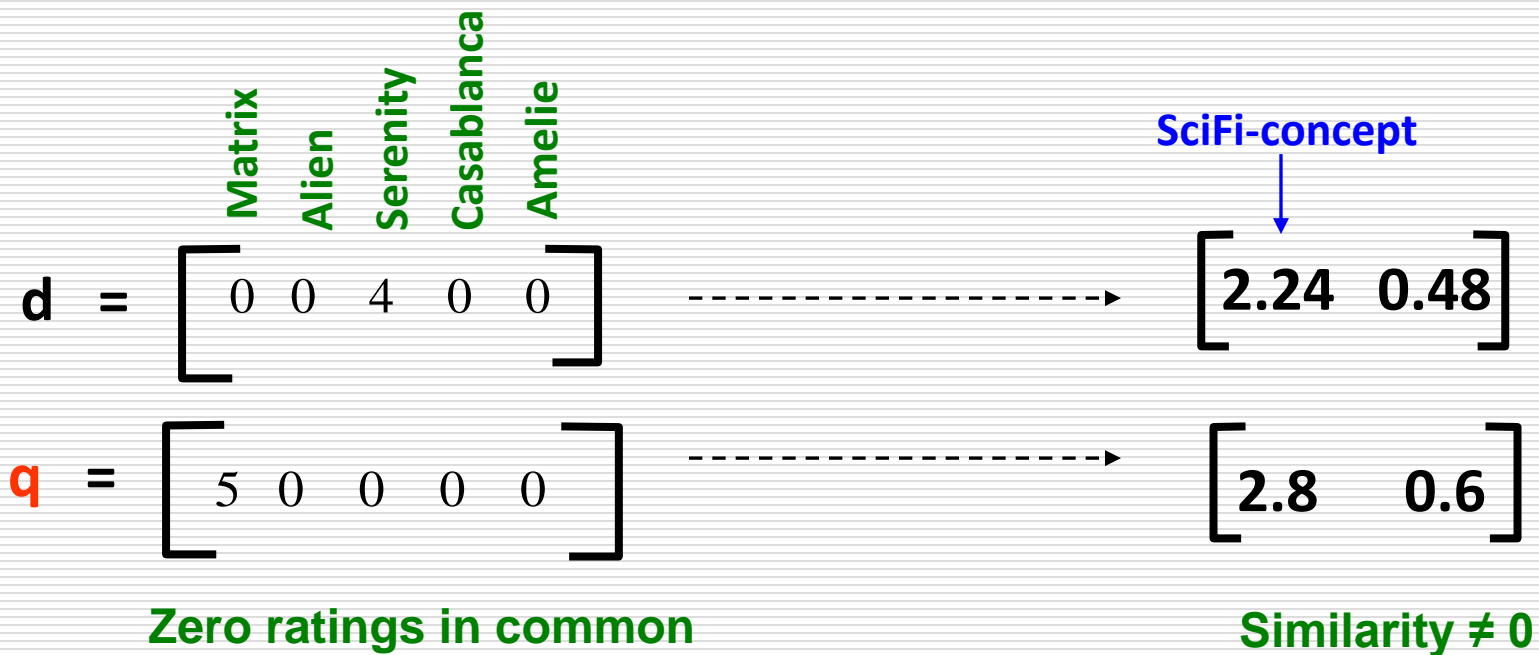
$$\mathbf{q} = \begin{matrix} & \text{Matrix} & \text{Alien} & \text{Serenity} & \text{Casablanca} & \text{Amelie} \\ \begin{bmatrix} 0 & 0 & 4 & 0 & 0 \end{bmatrix} & \times & \begin{bmatrix} 0.56 & 0.12 \\ 0.59 & -0.02 \\ 0.56 & 0.12 \\ 0.09 & -0.69 \\ 0.09 & -0.69 \end{bmatrix} & = & \begin{bmatrix} 2.24 & 0.48 \end{bmatrix} \end{matrix}$$

movie-to-concept
similarities (V)

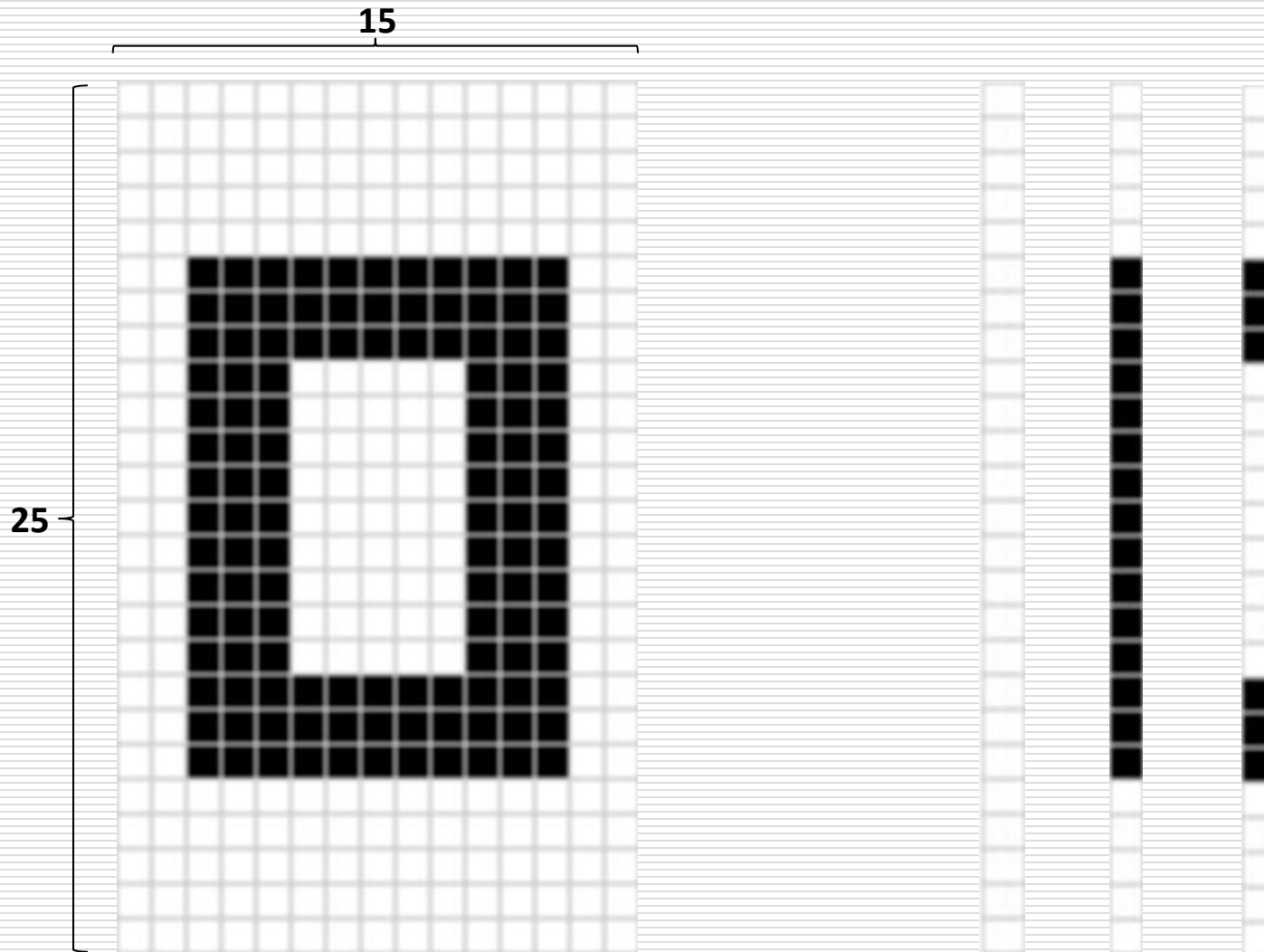
SciFi-concept
↓

Case study: How to query?

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

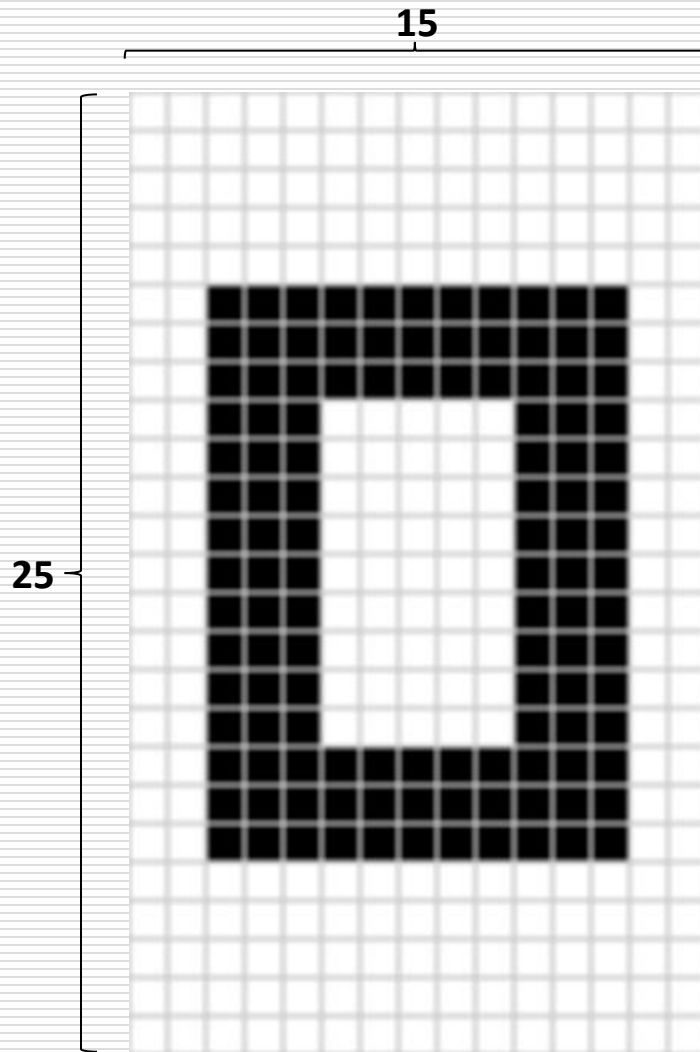


Case study: Image compression



Space cost= $25 \times 15 = 375$

Case study: Image compression



Space cost=25*15=375

$A =$

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	0	0	0	0	0	0	0	0	0	0	0	1	1
1	1	0	0	0	0	0	0	0	0	0	0	0	1	1
1	1	0	0	0	0	0	0	0	0	0	0	0	1	1
1	1	0	0	0	1	1	1	1	1	0	0	0	1	1
1	1	0	0	0	1	1	1	1	1	0	0	0	1	1
1	1	0	0	0	1	1	1	1	1	0	0	0	1	1
1	1	0	0	0	1	1	1	1	1	0	0	0	1	1
1	1	0	0	0	1	1	1	1	1	0	0	0	1	1
1	1	0	0	0	0	0	0	0	0	0	0	0	1	1
1	1	0	0	0	0	0	0	0	0	0	0	0	1	1
1	1	0	0	0	0	0	0	0	0	0	0	0	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

white color=1, black color=0

Case study: Image compression

15

25

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	0	0	0	0	0	0	0	0	0	0	0	1	1
1	1	0	0	0	0	0	0	0	0	0	0	0	1	1
1	1	0	0	0	0	0	0	0	0	0	0	0	1	1
1	1	0	0	0	1	1	1	1	1	0	0	0	1	1
1	1	0	0	0	1	1	1	1	1	0	0	0	1	1
1	1	0	0	0	1	1	1	1	1	0	0	0	1	1
1	1	0	0	0	1	1	1	1	1	0	0	0	1	1
1	1	0	0	0	1	1	1	1	1	0	0	0	1	1
1	1	0	0	0	1	1	1	1	1	0	0	0	1	1
1	1	0	0	0	1	1	1	1	1	0	0	0	1	1
1	1	0	0	0	0	0	0	0	0	0	0	0	1	1
1	1	0	0	0	0	0	0	0	0	0	0	0	1	1
1	1	0	0	0	0	0	0	0	0	0	0	0	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Space cost=25*15=375

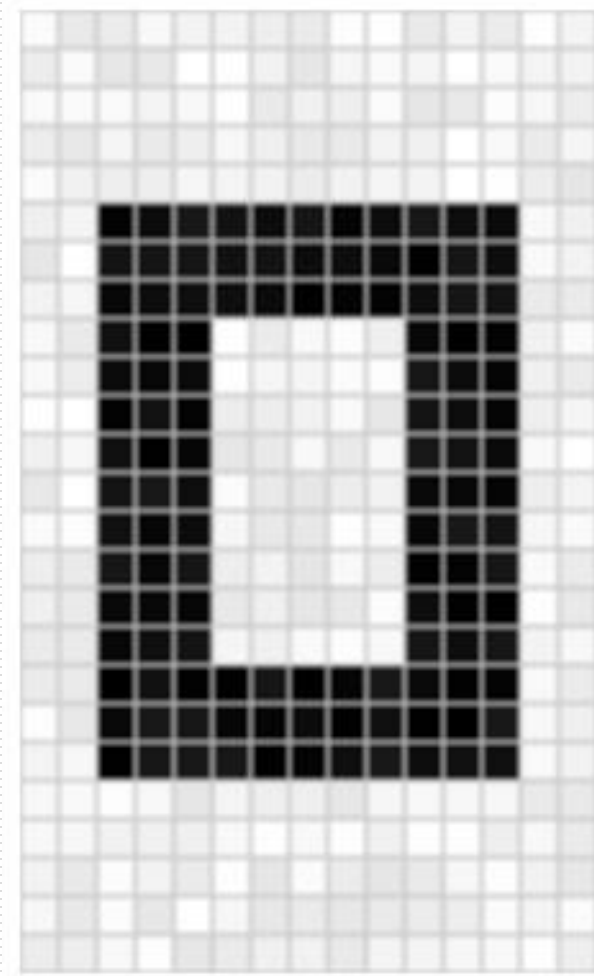
SVD:

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

$$\sigma_1=14.72, \sigma_2=5.22, \sigma_3=3.31$$

$$\text{Space cost}=25*3+15*3+3=123$$

Case study: Noise reduction



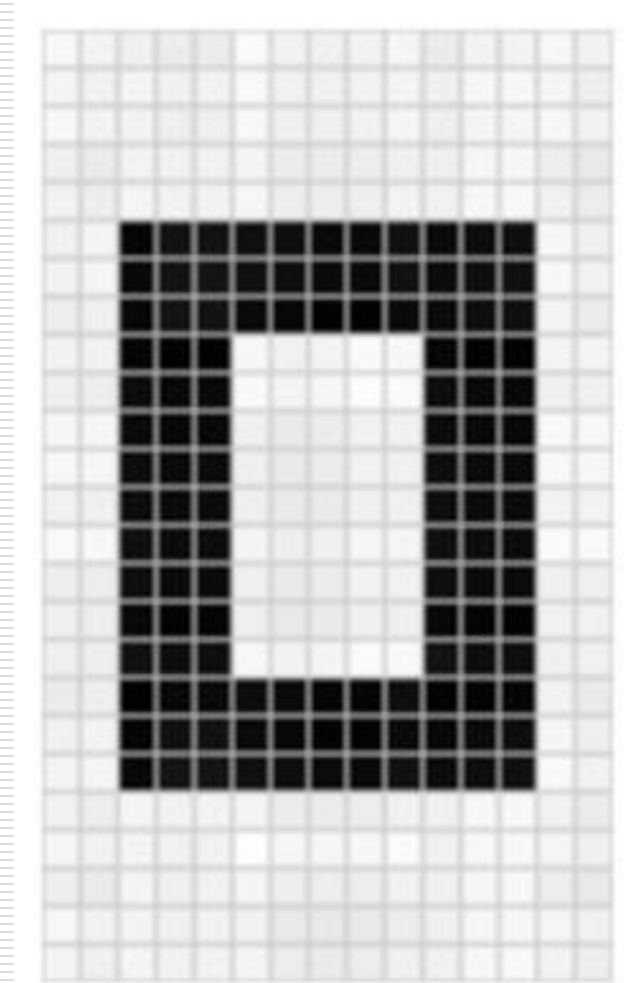
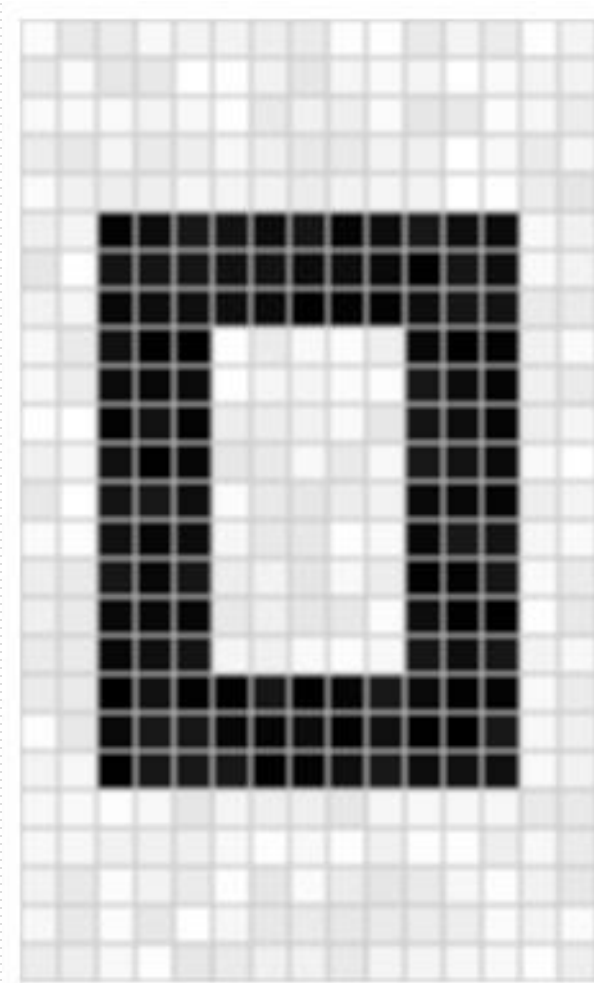
SVD:

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

$\sigma_1=14.15, \sigma_2=4.67, \sigma_3=3.00,$
 $\sigma_4=0.21, \sigma_5=0.19, \dots, \sigma_{15}=0.05$

$$\mathbf{A} \approx \mathbf{u}_1 \sigma_1 \mathbf{v}_1^T + \mathbf{u}_2 \sigma_2 \mathbf{v}_2^T + \mathbf{u}_3 \sigma_3 \mathbf{v}_3^T$$

Case study: Noise reduction



SVD with low rank approximation

Acknowledgement

- Slides are adapted from:
 - Prof. Jeffrey D. Ullman
 - Dr. Anand Rajaraman
 - Dr. Jure Leskovec