LATENT DIMENSIONS

AP

DIMENSIONALITY REDUCTION

Problem 8: Data reduction

Instance:

- a collection (dataset) $\mathbf{D}$ of datapoints from $\mathbf{X}$, e.g., $\mathbb{R}^m$
- [a distinct independent variable x_i]

Solution: a projection of $\mathbf{D}$ onto $\mathbb{R}^n$, $n < m$

Measure: error in the estimation of x_i

Example: genre identification in consumer behaviour analysis

EXAMPLE

Olivetti faces are 64x64 binary matrices.

Through SVD we discovered that most singular values are in fact 0 or very small

by considering only the top 20 or so singular values we can obtain a very similar image with less data:

- A contains 64x64=4096 values
- $U\Sigma V^T$ contains 20x20 + 20 + 20x20 = 820 values (5:1 compression ratio)

WHY REDUCE DIMENSIONS?

- easier to store, quicker to process
- interpretation and visualisation
- remove redundant or noisy features
- escape the curse of dimensionality and go back to intuitive distance features
- discover hidden correlations/topics

WHEN REDUCE DIMENSIONS?

When any of the goals becomes important

Empirically: when we believe that data essentially represents the mixing of a smaller set of *feature* which are the real *axes* of the data.

There really are only two *features*: sci-fi and romance

	Matrix	Alien	Star Wars	Casablanca	Titanic
Joe	1	1	1	0	0
Jim	3	3	3	0	0
John	4	4	4	0	0
Jack	5	5	5	0	0
Jill	0	0	0	4	4
Jenny	0	0	0	5	5
Jane	0	0	0	2	2

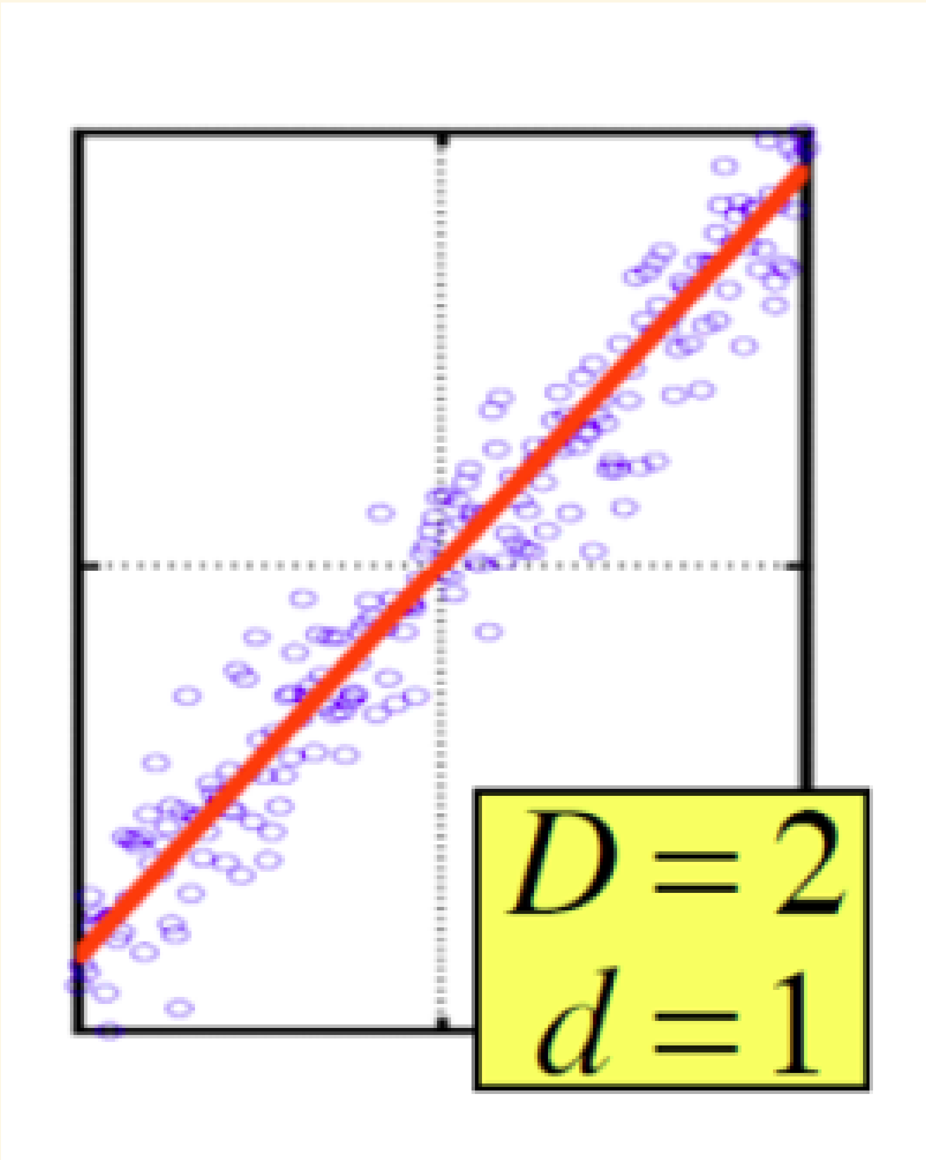
Figure 11.6: Ratings of movies by users

HOW TO REDUCE DIMENSIONS?

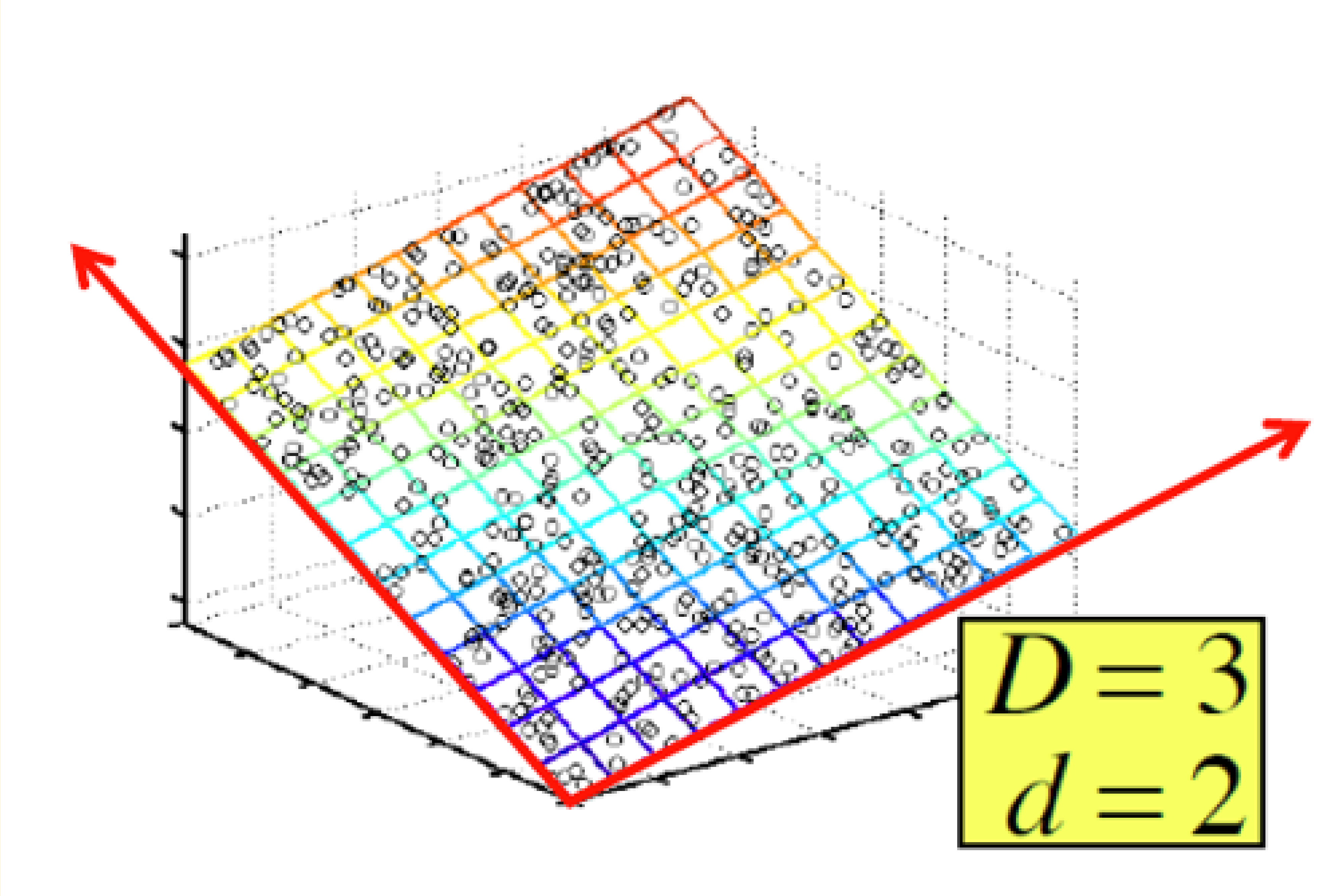
High-level view:

- data lies on or near a low-dimensional space:
- axes of that space are effective representations of data

EXAMPLES: $D \longrightarrow d$



The data axes (in red) are almost never the original measurement axes



BANKGROUND: MATRIX RANK

Rank is an important feature/descriptor of a data matrix

Rank is the maximum number or colums (or rows) that are linearly independent.

Such independent cols/rows are *candidates for the new, reduced reference system* (the red axes)

Rectangular matrices with SVD: $r \leq \min\{n, m\}$

We can map data points is a completely-new, dataset-dependent representation!

The dataset-depent representation will be compact

Let’s create a new, abstract *feature space*

EXAMPLE

Handmade dimensionality reduction, from the MMDS textbook

$$A_{3\times 3} = \begin{pmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{pmatrix}$$

$r = 2$ as the third row can be expressed as the first minus the second:

$$\begin{array}{rrrr} 1 & 2 & 1 & - \\ -2 & -3 & 1 & = \\ \hline \\ 3 & 5 & 0 & \end{array}$$

We create a new 2-d space where the axes are the first two rows:

$[1, 0, 0], [0, 1, 0], [0, 0, 1] \longrightarrow [1, 2, 1], [-2, -3, 1]$

The new rows are $[1, 0]$ $[0, 1]$ and $[1, 1]$

$$A'_{3\times 2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{pmatrix}$$

The new points work as selectors of the new axes: it's easy to go back from this space to the original, no loss of information/precision

DIMENSIONALITY REDUCTION

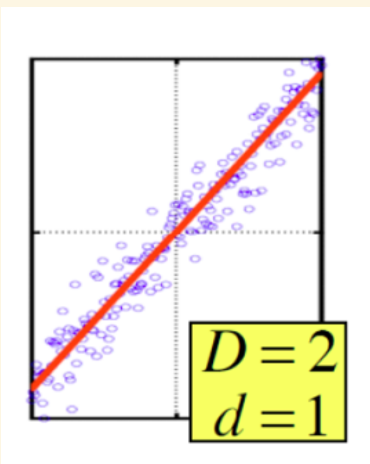
In real dataset mapping to a lower-dimensional space may introduce errors

in the $2 \to 1$ example below, instead of points we just take the measure (distance from the origin) of their projection on the red axis

The red axis is chosen as to minimise the error introduced by the $2 \to 1$ *projection.*

Data mining studies how to find such axes, called concepts

They capture some alignment which is inherent to the data.



SVD

DECOMPOSITION

$$A_{(m\times n)} = U_{(m\times m)} D_{(m\times n)} V^T_{(n\times n)}$$

- U is a orthogonal m. of *left-singular* (col.) vectors
- D is a diagonal matrix of *singular values*
- V is a orthogonal m. of *right-singular* (col.) vectors

Suppose only r ($r < \min\{m, n\}$) singular values are non-zero

We can rewrite the decomposition as follows:

$$A_{(m\times n)} \approx U_{(m\times r)} D_{(r\times r)} V^T_{(r\times n)}$$

Suppose $r = 2$, visualise

$$A_{(m \times n)} \approx U_{(m \times 2)} D_{(2 \times 2)} V^T_{(2 \times n)}$$

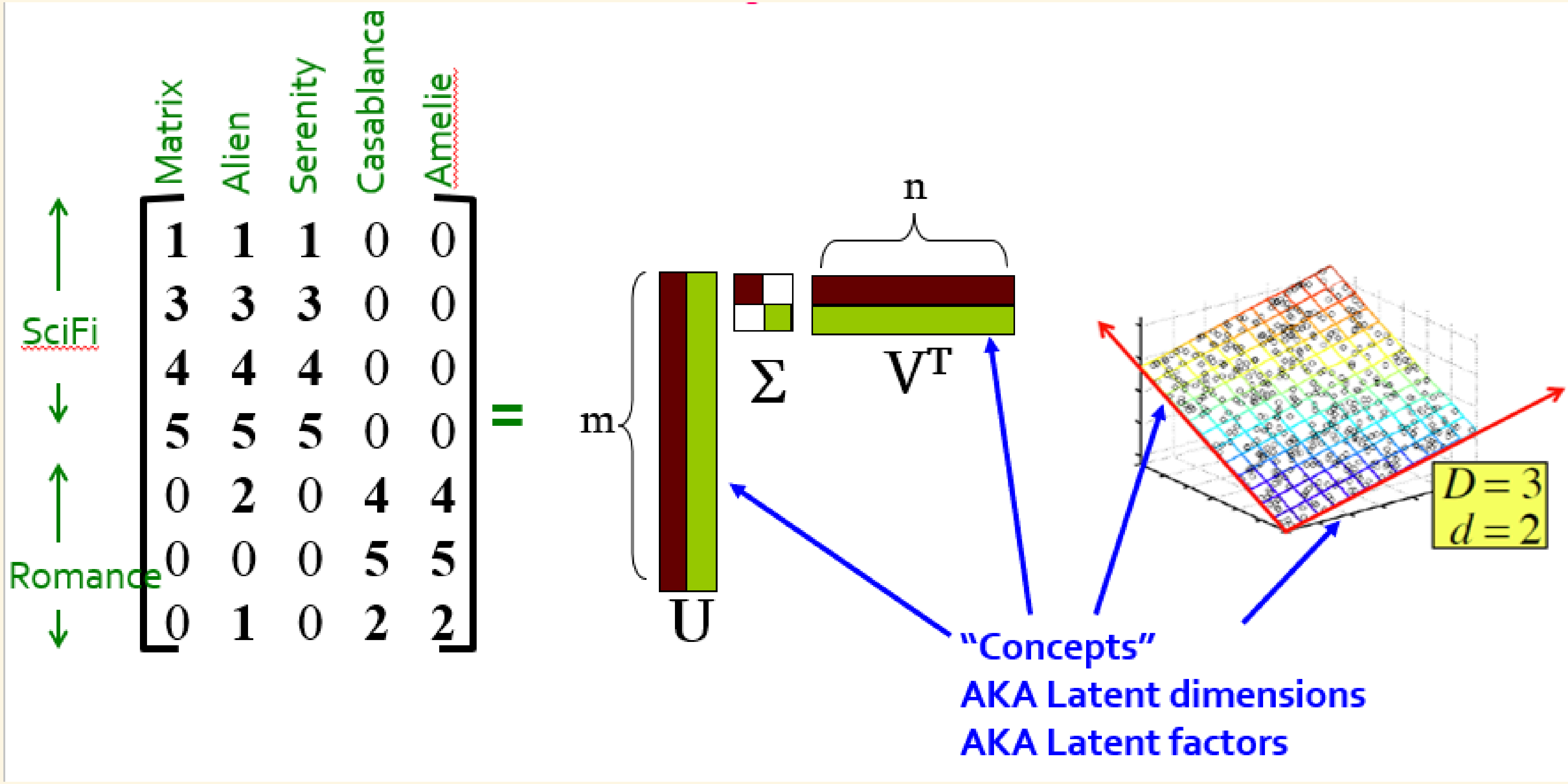
$$A_{(m \times n)} \approx \sum_i \sigma_i U_i \circ V_i^T$$

where U_i is the i-th column of U and $\circ$ is matrix multiplication

Now A is represented as the sum of independent *factors* that were not explicit in the original data

EXAMPLE: USERS TO FILMS BECOMES USERS TO CONCEPTS TO FILMS

(sligthy different data)



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

CONCEPTS

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

U is a user-to-concept similarity matrix

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

σs reveal the strength of each concept

SciFi-concept

“strength” of the SciFi-concept

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

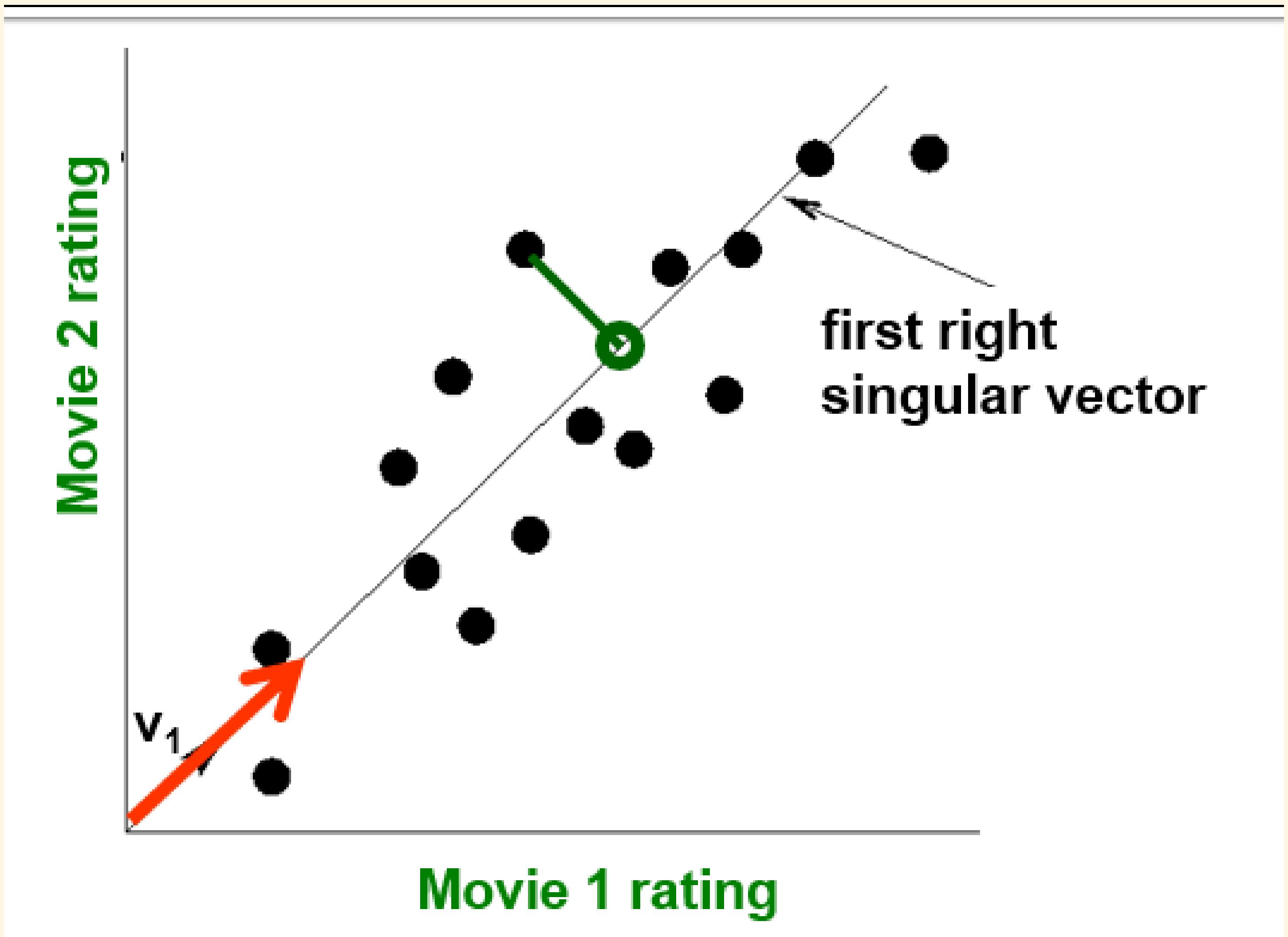
V^T is a concept-to-film similarity matrix

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

SVD INTERPRETATION -1

FROM SVD BACK TO GEOMETRY

The *singular vectors* that make up V (and U) are the new axes for projection



The *singular vectors* that make up V (and U) are the new axes for projection

They will minimise the *reconstruction error* (z is the the value obtained by the reduced SVD)

$$\epsilon = \sum_{i=1}^{m} \sum_{i=1}^{n} \|a_{ij} - z_{ij}\|$$

SVD INTERPRETATION -2

STEP-BY-STEP

Again, we use the singular vectors as the new axes