大数据计算及应用(八)

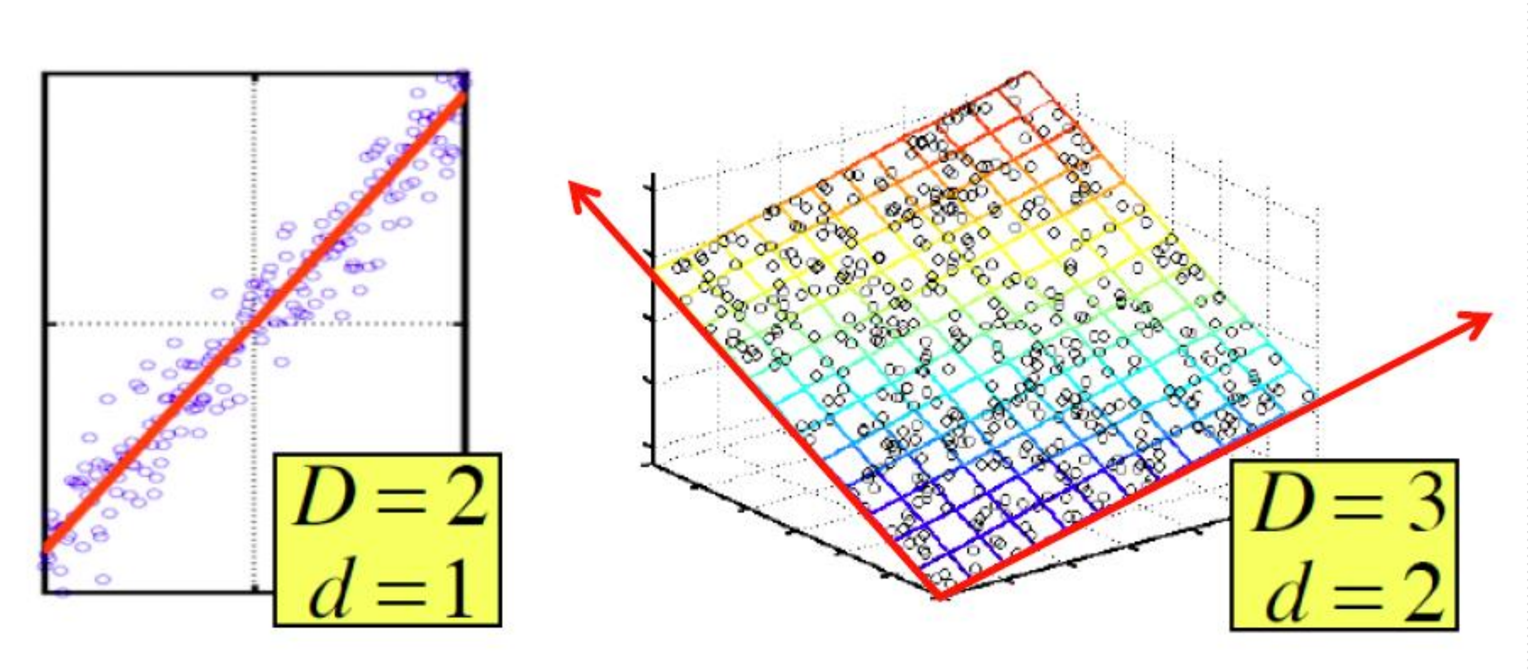
Dimensionality Reduction: SVD

Slides adapted from http://www.mmds.org

Agenda

High dim. data	Graph data	Infinite data	Machine learning	Apps
Locality sensitive hashing	PageRank, SimRank	Filtering data streams	SVM	Recommen der systems
Clustering	Community Detection	Web advertising	Decision Trees	Association Rules
Dimensiona lity reduction	Spam Detection	Queries on streams	Perceptron, kNN	Duplicate document detection

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**

day / customer	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]

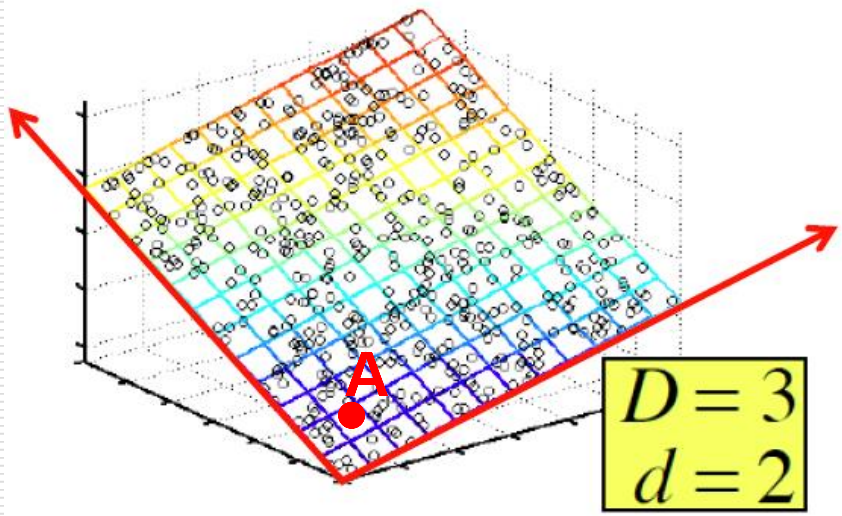
Rank of a Matrix

- **Q:** What is **rank** of a matrix **A**?
- **A:** Number of **linearly independent** rows (columns) of **A**
- **For example:**
  - Matrix **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: [-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:

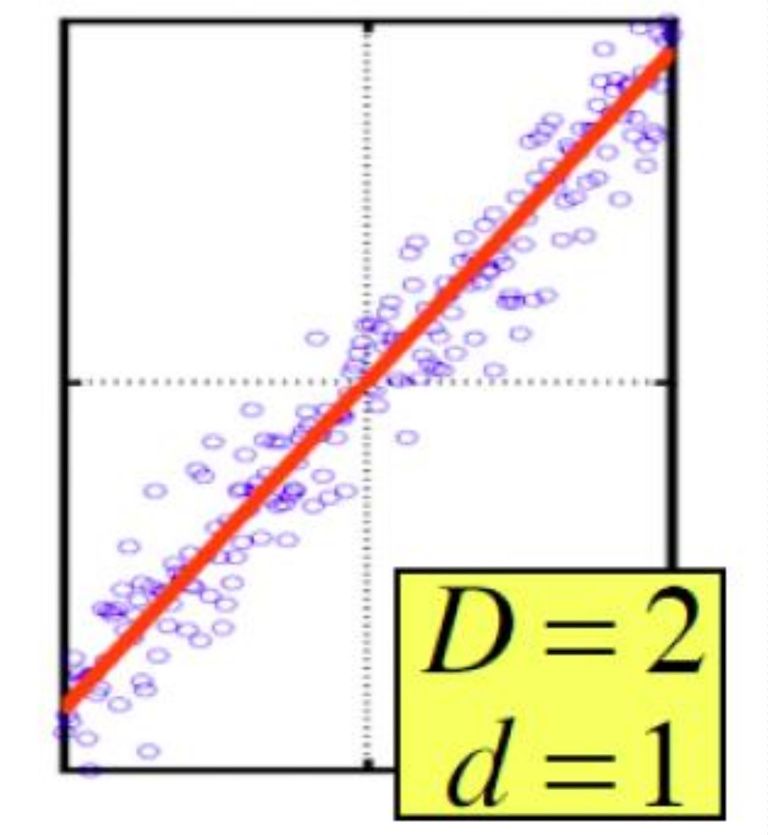
$$\text{1 row per point:} \quad \begin{array}{c} \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix} \end{array} \begin{array}{l} A \\ B \\ C \end{array}$$



- **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!**



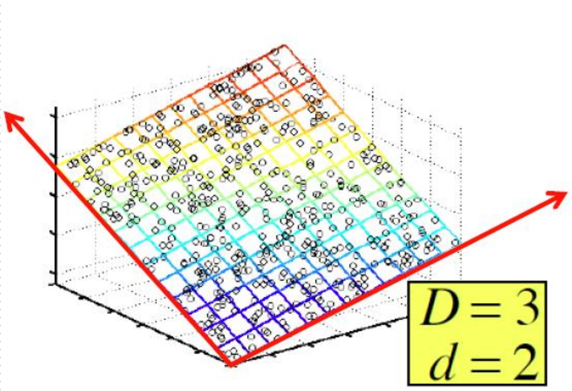
Rather than representing every point with 2 coordinates we represent each point with 1 coordinate (corresponding to the position of the point on the red line).

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**



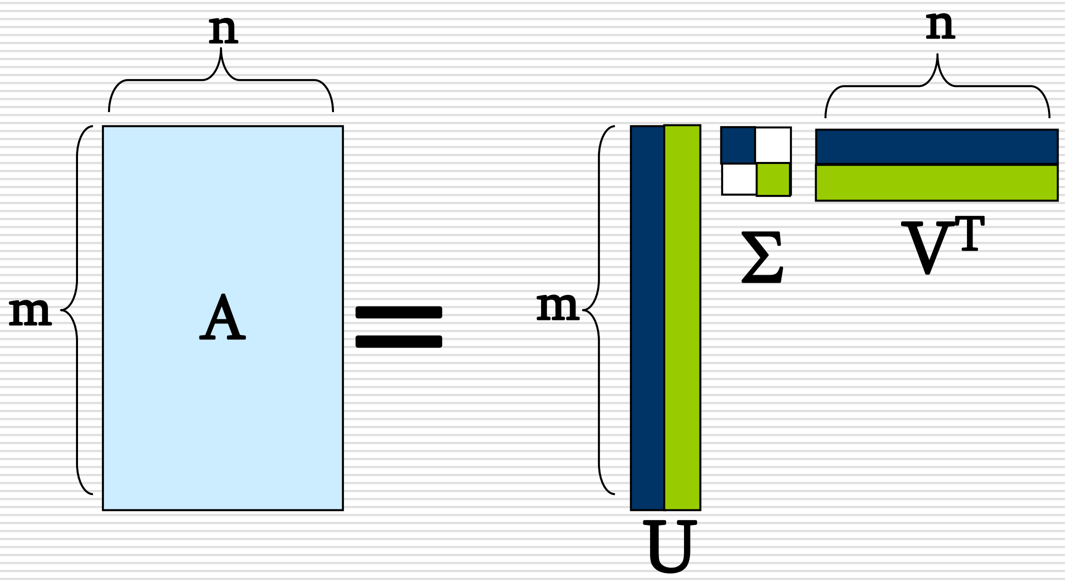
Singular Value Decomposition (SVD)

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

- **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)

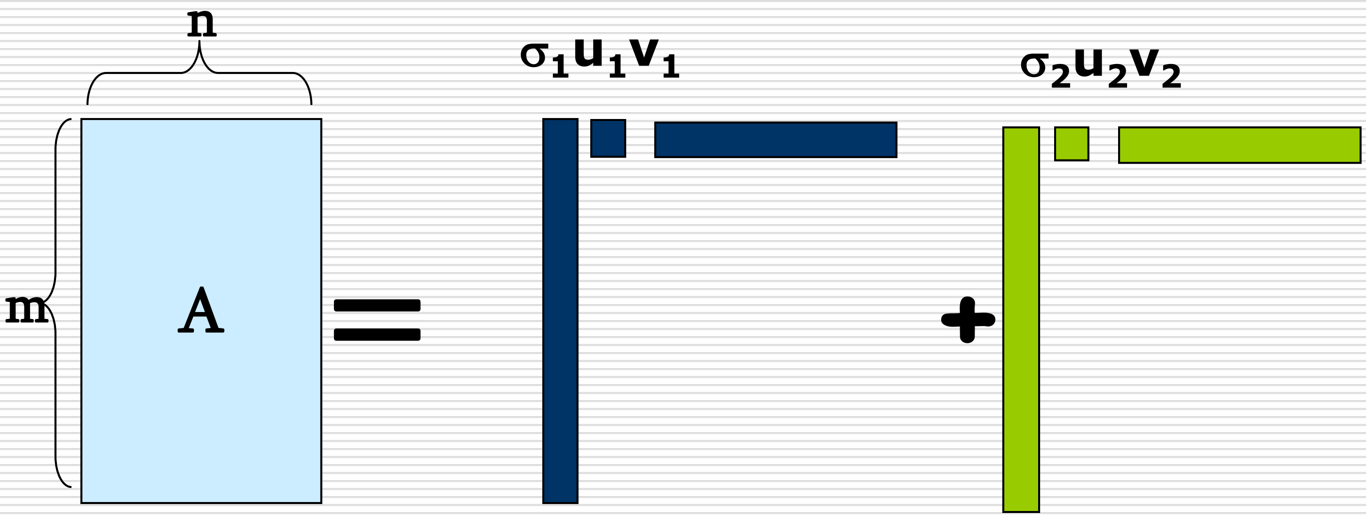
SVD

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



SVD

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



σ_i … scalar

u_i … vector

v_i … vector

SVD - Properties

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

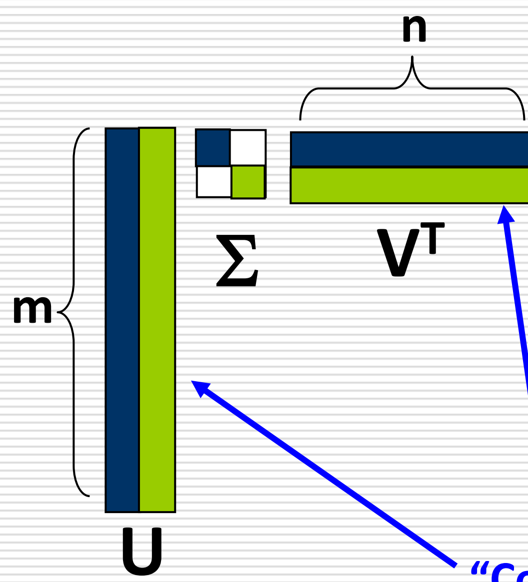
- $\boldsymbol{U}, \Sigma, \boldsymbol{V}$: unique
- $\boldsymbol{U}, \boldsymbol{V}$: column orthonormal
  - $\boldsymbol{U}^T \boldsymbol{U} = \boldsymbol{I}; \boldsymbol{V}^T \boldsymbol{V} = \boldsymbol{I}$ ($\boldsymbol{I}$: identity matrix)
  - (Columns are orthogonal unit vectors)
- Σ: diagonal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

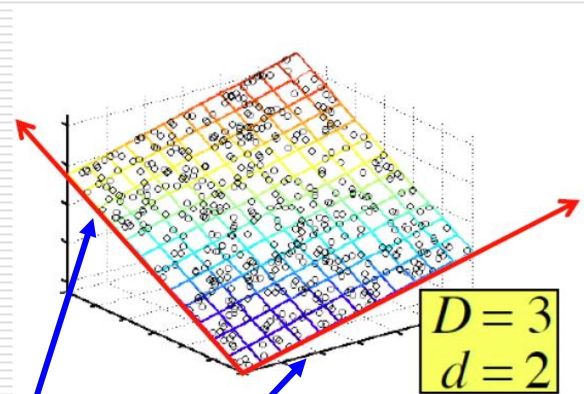
Nice proof of uniqueness: http://www.mpi-inf.mpg.de/~bast/ir-seminar-ws04/lecture2.pdf

SVD – Example: Users-to-Movies

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

$$
\begin{array}{c}
\\
u1\\ u2\\ u3\\ u4\\ u5\\ u6\\ u7
\end{array}
\begin{array}{c}
\begin{array}{ccccc}
\text{Matrix} & \text{Alien} & \text{Serenity} & \text{Casablanca} & \text{Amelie}
\end{array}\\
\begin{bmatrix}
1 & 1 & 1 & 0 & 0\\
3 & 3 & 3 & 0 & 0\\
4 & 4 & 4 & 0 & 0\\
5 & 5 & 5 & 0 & 0\\
0 & 2 & 0 & 4 & 4\\
0 & 0 & 0 & 5 & 5\\
0 & 1 & 0 & 2 & 2
\end{bmatrix}
\end{array}
=
$$





"Concepts"
AKA Latent dimensions
AKA Latent factors

SVD – Example: Users-to-Movies

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

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

SVD – Example: Users-to-Movies

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

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

SVD – Example: Users-to-Movies

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

***U* is "user-to-concept" similarity matrix**

SciFi-concept

Romance-concept

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

SVD – Example: Users-to-Movies

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

SciFi-concept

"strength" of the SciFi-concept

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

SVD – Example: Users-to-Movies

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

***V* is "movie-to-concept" similarity matrix**

SciFi-concept

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

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

SciFi-concept

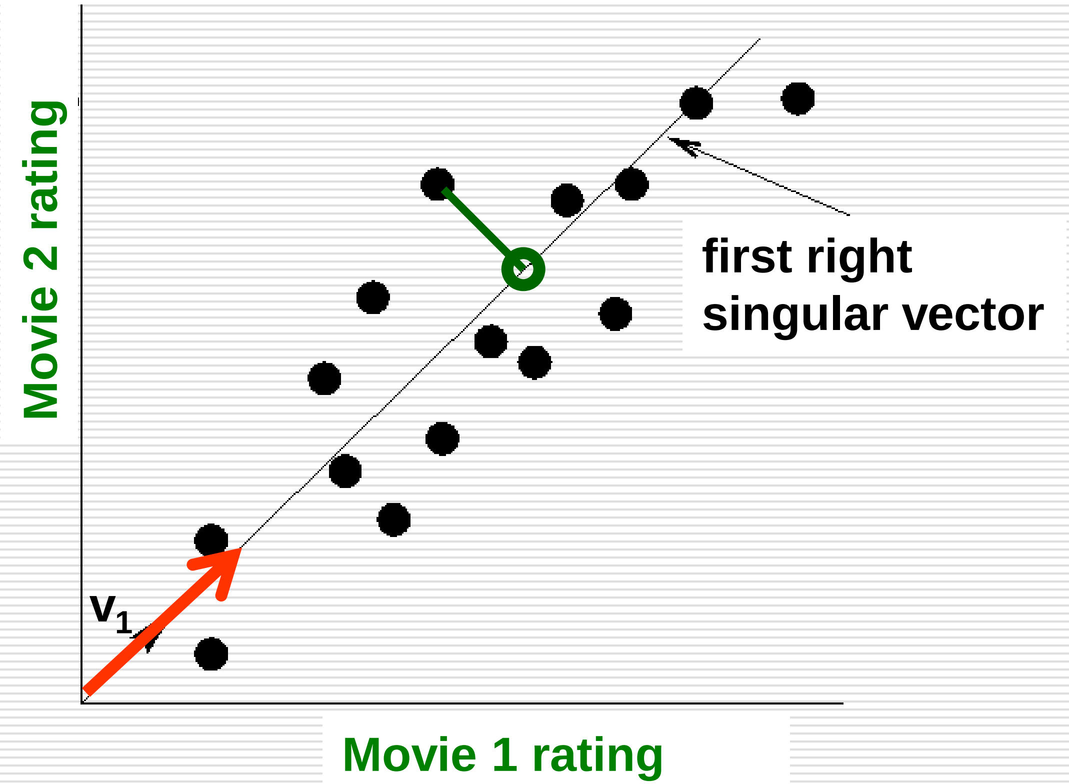
SVD - Interpretation #1

‘**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



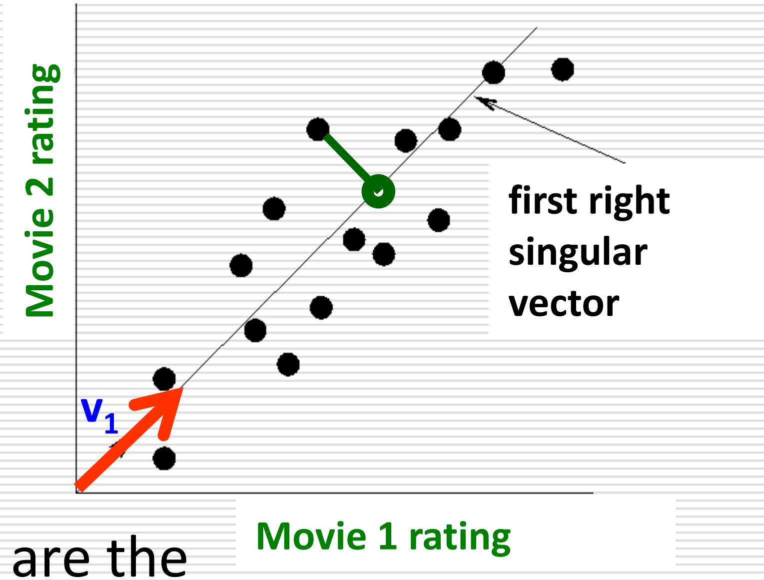
- Instead of using two coordinates (x, y) to describe point locations, let's use only one coordinate (z)
- Point's position is its location along vector v_1
- **How to choose v_1? Minimize reconstruction error**

SVD – Dimensionality Reduction

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

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

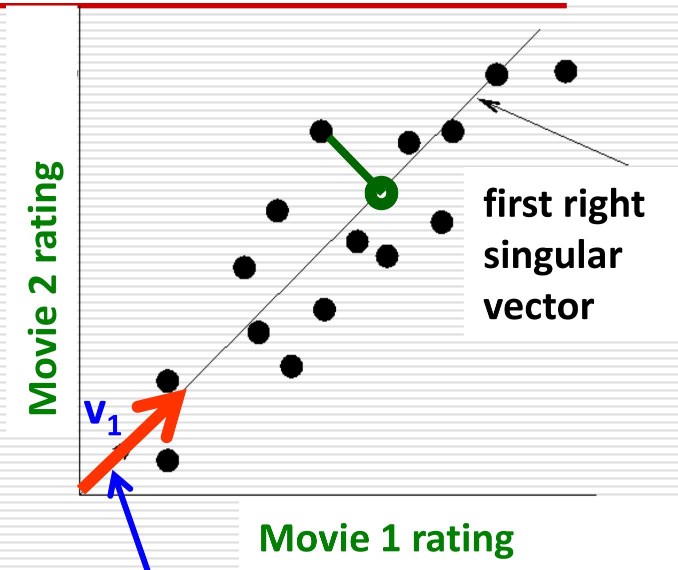
  - where $\boldsymbol{x_{ij}}$ are the "old" and $\boldsymbol{z_{ij}}$ are the "new" coordinates



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

SVD - Interpretation #2

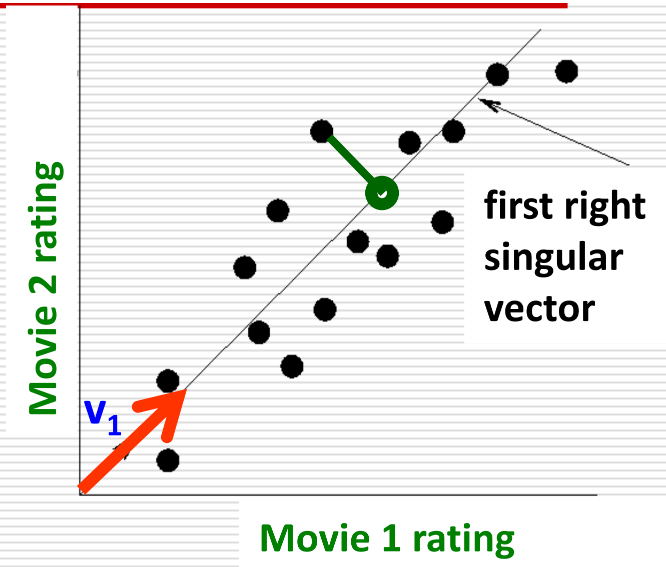
- **$A = U \Sigma V^T$ - example:**
  - **V**: "movie-to-concept" matrix
  - **U**: "user-to-concept" matrix



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

SVD - Interpretation #2

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



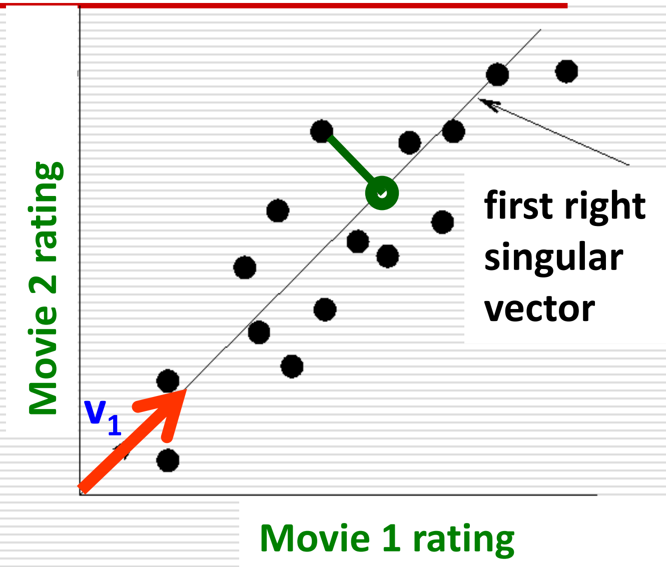
variance ('spread') on the v_1 axis

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

SVD - Interpretation #2

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

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



$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix}$$

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

$$\begin{bmatrix} \boxed{\begin{matrix} 1.61 \\ 5.08 \\ 6.82 \\ 8.43 \\ 1.86 \\ 0.86 \\ 0.86 \end{matrix}} & \begin{matrix} 0.19 \\ 0.66 \\ 0.85 \\ 1.04 \\ -5.60 \\ -6.93 \\ -2.75 \end{matrix} & \begin{matrix} -0.01 \\ -0.03 \\ -0.05 \\ -0.06 \\ 0.84 \\ -0.87 \\ 0.41 \end{matrix} \end{bmatrix}$$

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**

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

SVD - Interpretation #2

More details

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

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

SVD - Interpretation #2

More details

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

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

SVD - Interpretation #2

More details

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

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

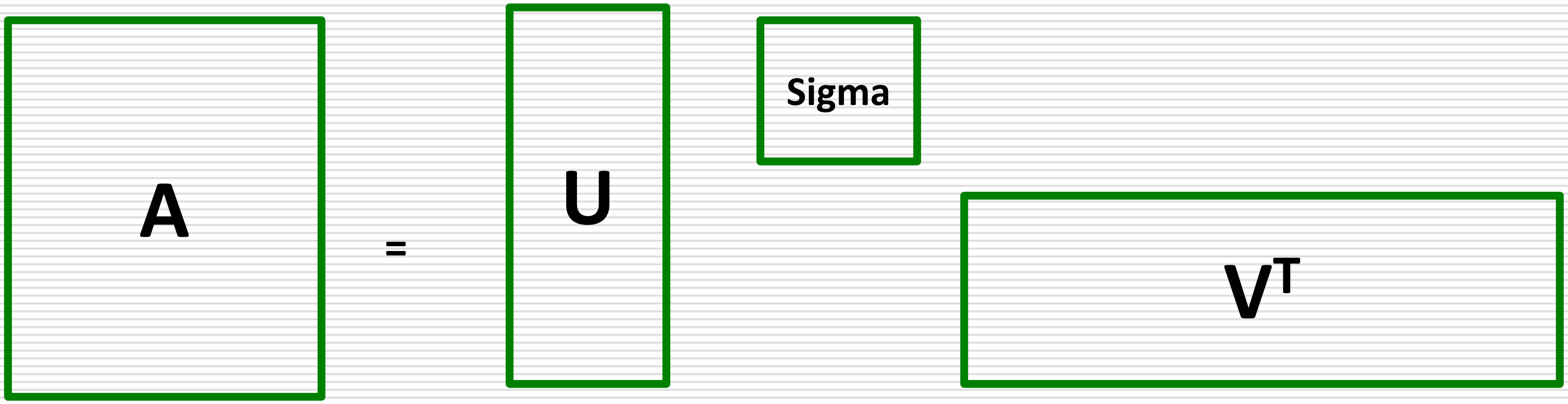
Frobenius norm:

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

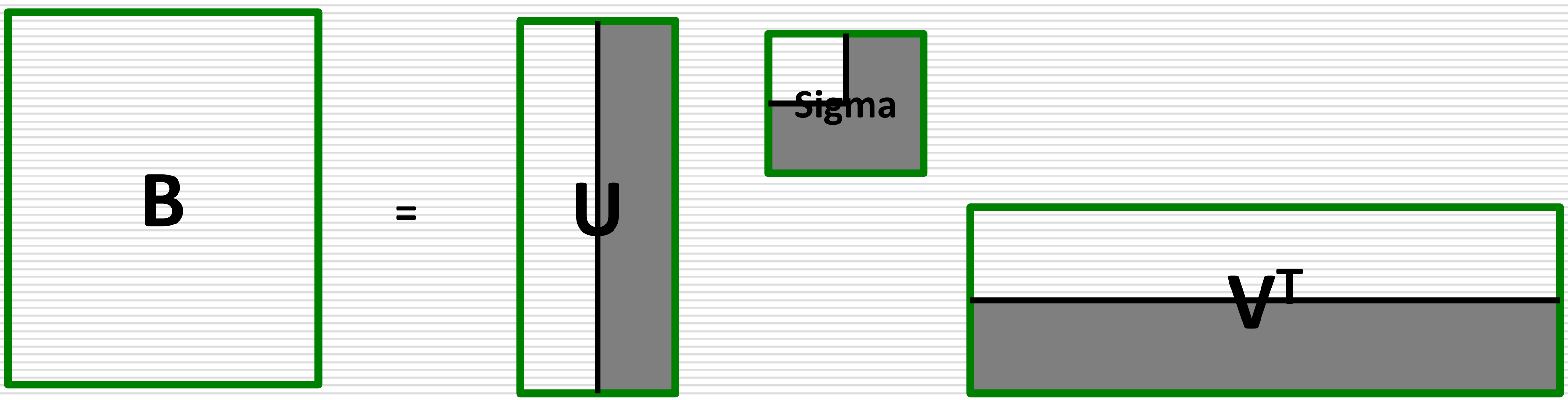
$$\|A-B\|_F = \sqrt{\Sigma_{ij} (A_{ij}-B_{ij})^2}$$

is “small”

SVD – Best Low Rank Approx.



B is best approximation of A

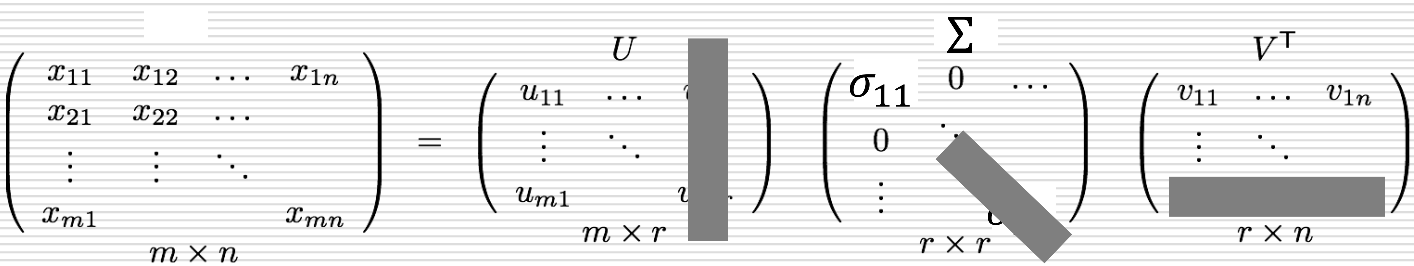


SVD – Best Low Rank Approx.

- **Theorem:**
  Let $\mathbf{A = U\,\Sigma\,V^T}$ and $\mathbf{B = U\,S\,V^T}$ where
  S = diagonal *r*x*r* matrix with $s_i = \sigma_i$ ($i=1\ldots k$) else $s_i = 0$
  then **B** is a **best** rank(**B**)=k approx. to ***A***

What do we mean by "best":

- ***B* is a solution to min$_B$ ||*A*-*B*||$_F$ where rank(*B*)=*k***



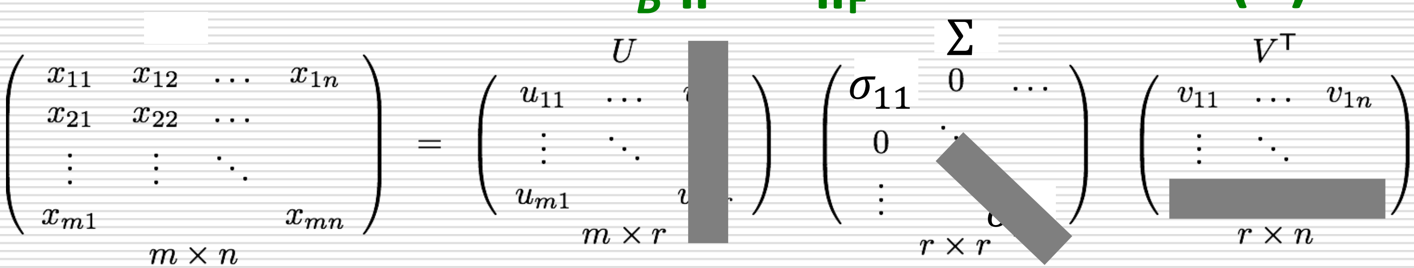
$$\|A - B\|_F = \sqrt{\sum_{ij} \left(A_{ij} - B_{ij}\right)^2}$$

SVD – Best Low Rank Approx.

Details!

- **Theorem:** Let $\mathbf{A} = \mathbf{U}\,\Sigma\,\mathbf{V}^T$ $(\sigma_1 \geq \sigma_2 \geq \ldots, \text{rank}(\boldsymbol{A}) = \boldsymbol{r})$ **then** $\mathbf{B} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^T$
  - **S = diagonal *rxr* matrix** where $s_i = \sigma_i$ $(i=1\ldots k)$ else $s_i = 0$

  is a best rank-$\boldsymbol{k}$ approximation to $\boldsymbol{A}$:
  - $\boldsymbol{B}$ **is a solution to** $\min_B \|A-B\|_F$ **where** $\text{rank}(B)=k$



- **We will need:**
  - $\mathbf{U}\,\Sigma\,\mathbf{V}^T - \mathbf{U}\,\mathbf{S}\,\mathbf{V}^T = \mathbf{U}\,(\Sigma - \mathbf{S})\,\mathbf{V}^T$

SVD – Best Low Rank Approx.

Details!

- **A = U Σ V^T , B = U S V^T** ($\sigma_1 \geq \sigma_2 \geq \ldots \geq 0$, rank(***A***)=***r***)
  - **S** = diagonal *n*x*n* matrix where $s_i = \sigma_i$ (*i=1…k*) else $s_i = 0$

then ***B*** is solution to $\min_B \|A-B\|_F$, rank(***B***)=***k***

- **Why?**

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

We used: U Σ V^T - U S V^T = U (Σ - S) V^T

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

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

SVD - Interpretation #2

Equivalent:

'spectral decomposition' of the matrix:

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

SVD - Interpretation #2

Equivalent:

'spectral decomposition' of the matrix

$$
\begin{array}{c c}
 & \xleftarrow{\quad m \quad} \\
n \updownarrow & \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix}
\end{array}
= \overbrace{\sigma_1 \underbrace{u_1}_{n \times 1} \underbrace{v^T_1}_{1 \times m} + \sigma_2 u_2 v^T_2 + \ldots}^{k \text{ terms}}
$$

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

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 σs to keep?

A: Rule-of-a thumb:

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

$$
\begin{array}{cc}
 & \xleftarrow{\quad m \quad} \\
n \updownarrow & \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix}
\end{array}
= \sigma_1 \quad u_1 \quad v^T_1 \quad + \quad \sigma_2 \quad u_2 \quad v^T_2 \quad + \ldots
$$

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

Relation to Eigen-decomposition

- **SVD gives us:**
  - $A = U \Sigma V^T$
- **Eigen-decomposition:**
  - $M = X \Lambda X^T$
    - M is symmetric
    - U, V, X are orthonormal (**$U^TU=I$**),
    - **Λ, Σ** are diagonal
- **Now let's calculate:**
  - **$AA^T=$**
  - **$A^TA = V \Sigma^T U^T (U\Sigma V^T) = V \Sigma\Sigma^T V^T$**

Relation to Eigen-decomposition

- **SVD gives us:**
  - $A = U\,\Sigma\,V^T$
- **Eigen-decomposition:**
  - $M = X\,\mathrm{L}\,X^T$
    - M is symmetric
    - U, V, X are orthonormal (**$U^TU=I$**),
    - **L**, Σ are diagonal
- **Now let's calculate:**
  - $\mathbf{AA^T = U\Sigma\, V^T(U\Sigma\, V^T)^T = U\Sigma\, V^T(V\Sigma^T U^T) = \underbrace{U}_{X}\underbrace{\Sigma\Sigma^T}_{L}\,\underbrace{U^T}_{X^T}}$
  - $\mathbf{A^TA = V\,\Sigma^T\,U^T\,(U\Sigma\,V^T) = \underbrace{V}_{X}\,\underbrace{\Sigma\Sigma^T}_{L}\,\underbrace{V^T}_{X^T}}$

Shows how to compute SVD using eigenvalue decomposition!

where L = Σ^2

An Example of SVD Computation

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

A = U Σ V
$A\,A^T = X\,L\,X^T = U\,\Sigma^2\,U^T$
$A^T A = Y\,L\,Y^T = V\,\Sigma^2\,V^T$

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

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

$\mathbf{M\,X = \lambda\,X} \Rightarrow (\boldsymbol{\lambda}\,\mathbf{I} - \mathbf{M})\,\mathbf{X} = 0 \Rightarrow |\,\boldsymbol{\lambda}\,\mathbf{I} - \mathbf{M}| = 0$

I: identity matrix

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

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

An Example of SVD Computation

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

$\mathbf{A\,A^T = X\,L\,X^T = U\,\Sigma^2\,U^T}$

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

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

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

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

An Example of SVD Computation

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

A = U Σ V

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

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

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

λ_1=3, λ_2=1

λ_1=3, V_1= $\begin{bmatrix} \frac{\mathbf{1}}{\sqrt{\mathbf{2}}} \\ \frac{\mathbf{-1}}{\sqrt{\mathbf{2}}} \end{bmatrix}$

λ_2=1, V_2= $\begin{bmatrix} \frac{\mathbf{1}}{\sqrt{\mathbf{2}}} \\ \frac{\mathbf{1}}{\sqrt{\mathbf{2}}} \end{bmatrix}$

An Example of SVD Computation

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

A = U Σ V

A A^T = X L X^T = U Σ^2 U^T

A^T A = Y L Y^T = V Σ^2 V^T

A = U Σ V

$$\begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{6}} & 0 & [illegible] \\ \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & [illegible] \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & [illegible] \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ [illegible] & [illegible] \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

SVD - Complexity

- **To compute SVD:**
  - **O(nm^2)** or **O(n^{2}m)** (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 Σ $\mathbf{V}^T$: 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?

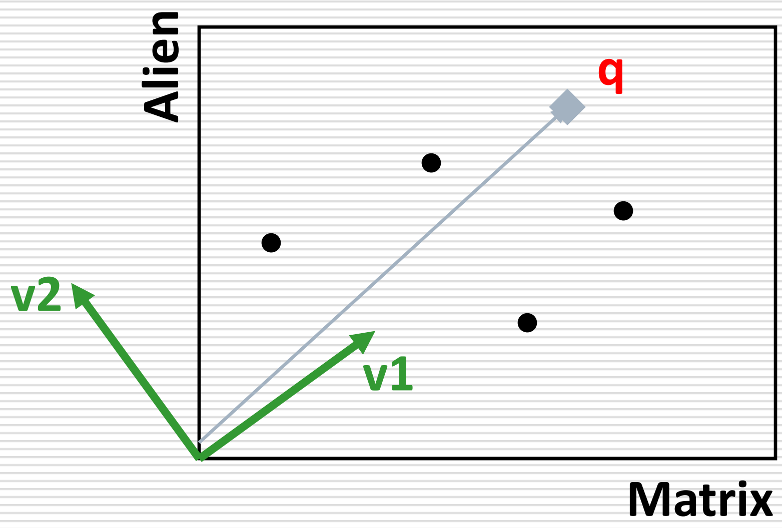
$$\begin{array}{cc} & \begin{array}{ccccc} \text{Matrix} & \text{Alien} & \text{Serenity} & \text{Casablanca} & \text{Amelie} \end{array} \\ \begin{array}{c} u1 \\ u2 \\ u3 \\ u4 \\ u5 \\ u6 \\ u7 \end{array} & \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 &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

Case study: How to query?

- **Q: Find users that like ‘Matrix’**
- **A: Map query into a ‘concept space’ – how?**

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

Project into concept space:
Inner product with each
‘concept’ vector v_i

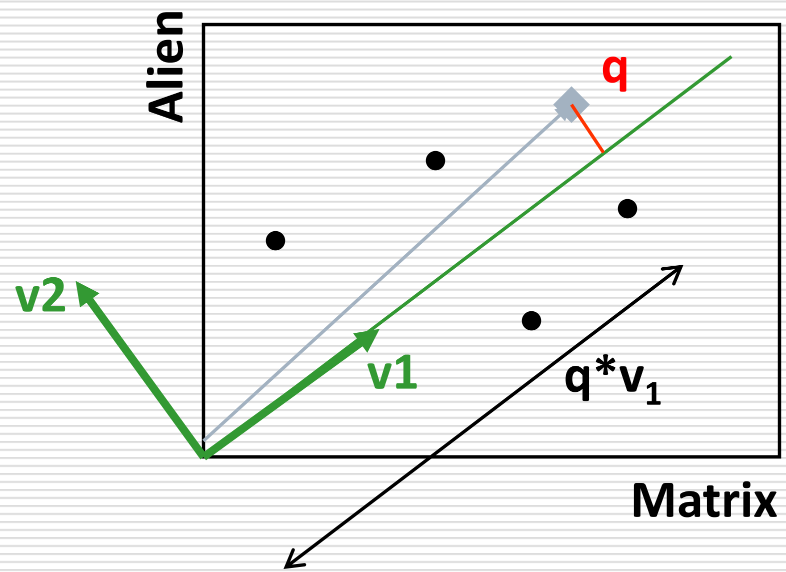


Case study: How to query?

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

$$q = \begin{array}{c} \begin{array}{ccccc} \text{Matrix} & \text{Alien} & \text{Serenity} & \text{Casablanca} & \text{Amelie} \end{array} \\ \left[\begin{array}{ccccc} 5 & 0 & 0 & 0 & 0 \end{array}\right] \end{array}$$

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



Case study: How to query?

Compactly, we have:

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

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 similarities (V)

SciFi-concept

Case study: How to query?

- **How would the user *d* that rated ('Serenity') be handled?**
  $\mathbf{d_{concept} = d\ V}$

E.g.:

$$
\begin{array}{cc}
 & \begin{array}{ccccc} \text{Matrix} & \text{Alien} & \text{Serenity} & \text{Casablanca} & \text{Amelie} \end{array} \\
\mathbf{q} = & \begin{bmatrix} 0 & 0 & 4 & 0 & 0 \end{bmatrix}
\end{array}
\ \mathbf{x} \
\begin{bmatrix} 0.56 & 0.12 \\ 0.59 & -0.02 \\ 0.56 & 0.12 \\ 0.09 & -0.69 \\ 0.09 & -0.69 \end{bmatrix}
=
\begin{array}{c} \text{SciFi-concept} \\ \begin{bmatrix} 2.24 & 0.48 \end{bmatrix} \end{array}
$$

movie-to-concept similarities (V)

Case study: How to query?

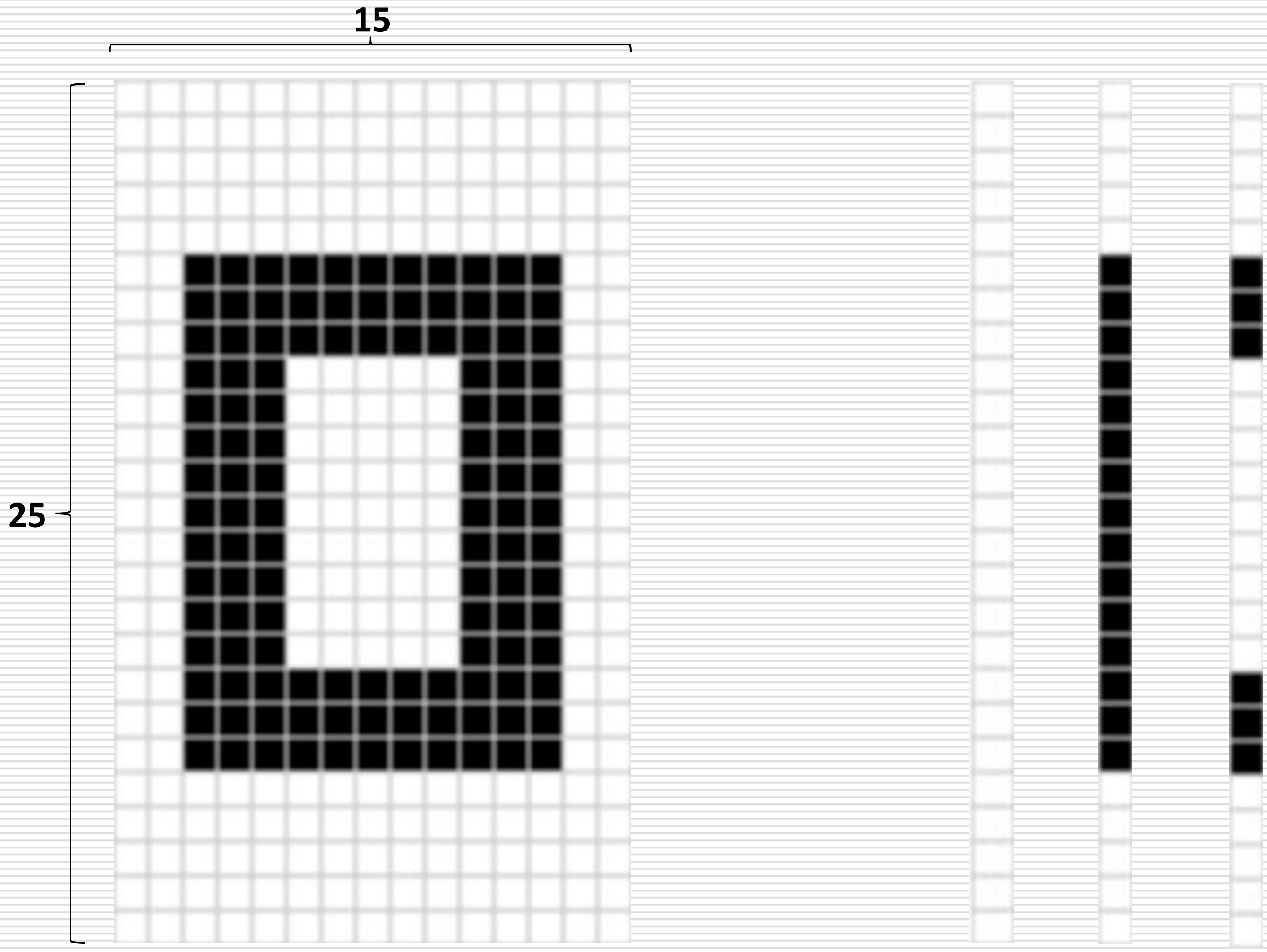
- **Observation:** User ***d*** that rated (‘*Serenity*’) will be **similar** to user **q** that rated (‘*Matrix*’), although ***d*** and **q** have **zero ratings in common**!

$$\begin{array}{cc} & \begin{array}{ccccc} \text{Matrix} & \text{Alien} & \text{Serenity} & \text{Casablanca} & \text{Amelie} \end{array} \\ \mathbf{d} = & \begin{bmatrix} 0 & 0 & 4 & 0 & 0 \end{bmatrix} \\ \mathbf{q} = & \begin{bmatrix} 5 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \quad \begin{array}{c} \\ \dashrightarrow \\ \dashrightarrow \end{array} \quad \begin{array}{c} \begin{array}{cc} \text{SciFi-concept} & \end{array} \\ \begin{bmatrix} \mathbf{2.24} & \mathbf{0.48} \end{bmatrix} \\ \begin{bmatrix} \mathbf{2.8} & \mathbf{0.6} \end{bmatrix} \end{array}$$

Zero ratings in common

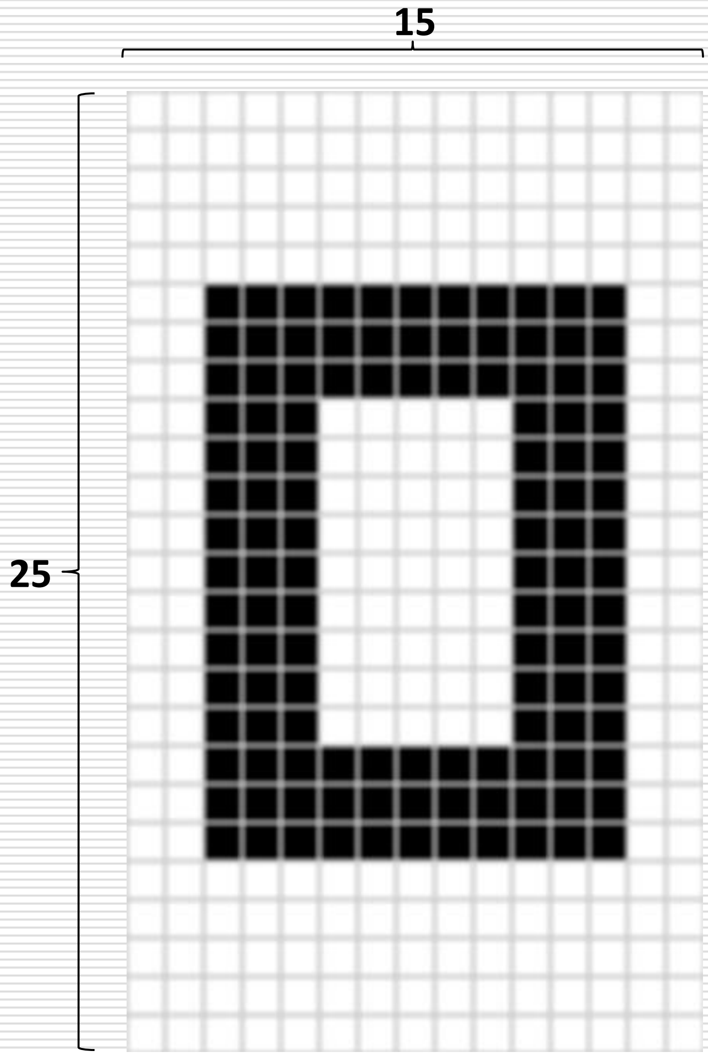
Similarity ≠ 0

Case study: Image compression



Space cost=25*15=375

Case study: Image compression



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

Space cost=25*15=375

white color=1, black color=0

Case study: Image compression

$$
25\left\{\overbrace{\begin{bmatrix}
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1
\end{bmatrix}}^{15}\right.
$$

Space cost=25*15=375

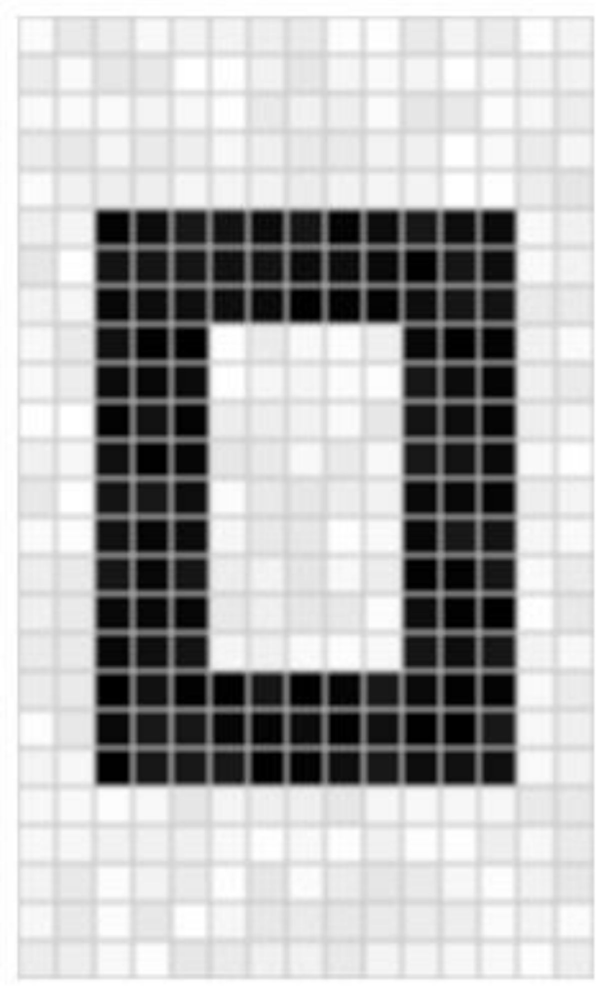
SVD:

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

σ1=14.72, σ2=5.22, σ3=3.31

Space cost=25*3+15*3+3=123

Case study: Noise reduction



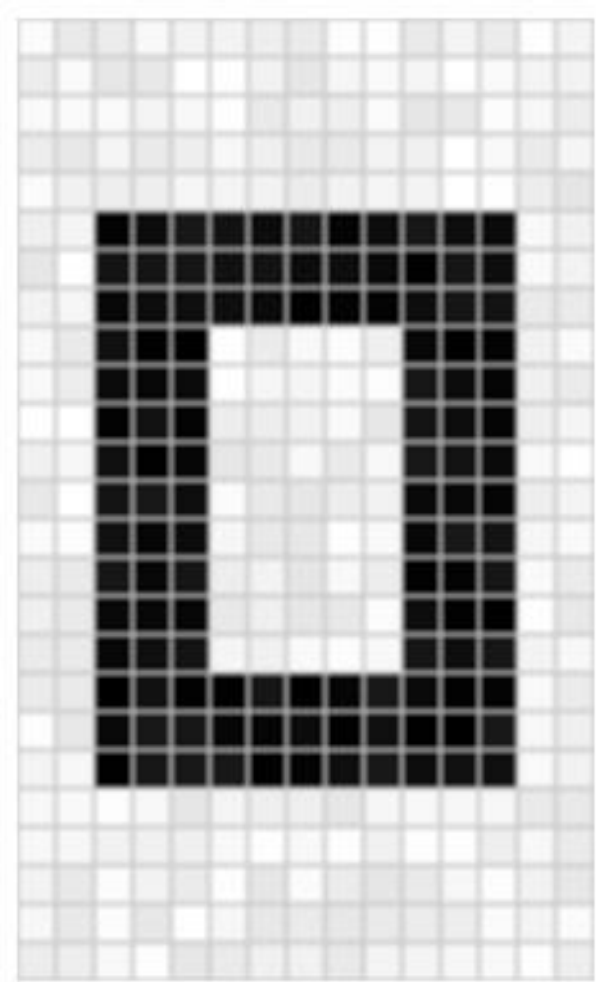
SVD:

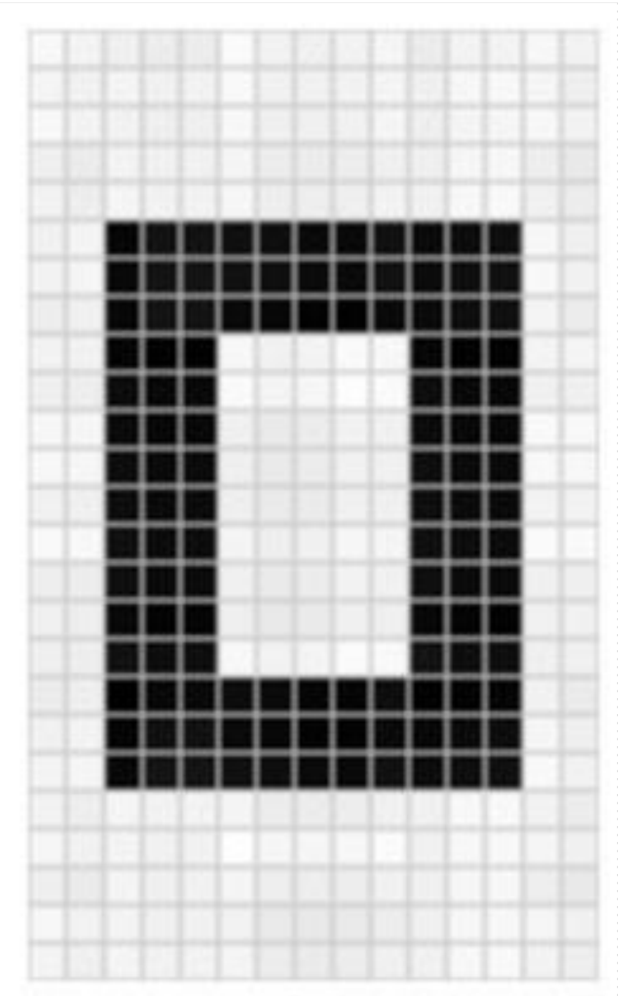
$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_i \sigma_i \mathbf{u}_i \circ \mathbf{v}_i$$

σ1=14.15, σ2=4.67, σ3=3.00, σ4=0.21 , σ5=0.19, …… , σ15=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