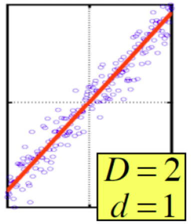
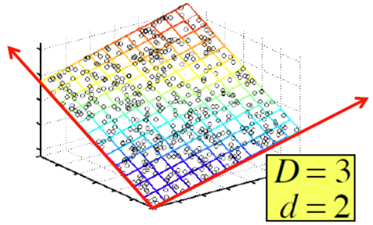
大数据分析

Large-scale computing

刘盛华

Dimensionality Reduction





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

J. Leskovec, A. Rajaraman, J. Ullman: Mining of Massive Datasets, http://www.mmds.org

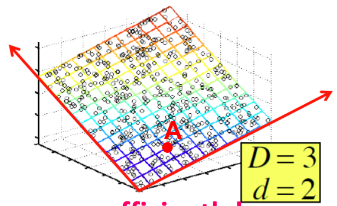
Rank of a Matrix

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

Rank is "Dimensionality"

- **Cloud of points 3D space:**
  - Think of point positions as a matrix: $\begin{array}{c} \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix} \end{array} \begin{array}{l} \mathbf{A} \\ \mathbf{B} \\ \mathbf{C} \end{array}$

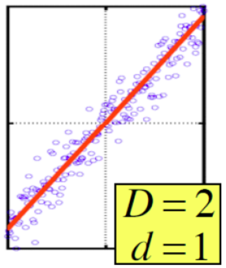
    1 row per point:



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

Dimensionality Reduction

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



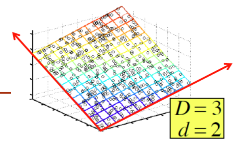
Rather than representing every point with 2 coordinates we represent each point with 1 coordinate (corresponding to the position of the point on the red line).

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



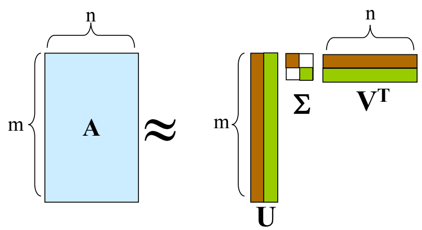
SVD - Definition

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

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

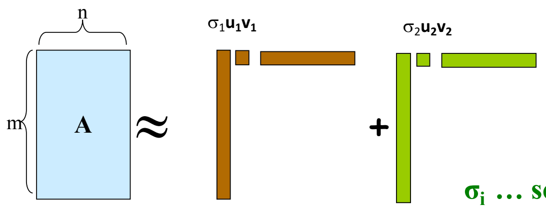
SVD

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



SVD

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



σ_i … scalar
u_i … vector
v_i … vector

SVD - Properties

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

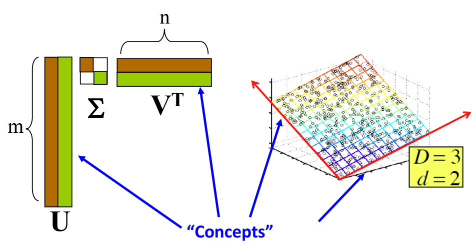
- U, Σ, V: unique
- U, V: column orthonormal
  - $U^T U = I$; $V^T V = I$ (I: identity matrix)
  - (Columns are orthogonal unit vectors)
- Σ: diagonal
  - Entries (singular values) are positive, and sorted in decreasing order ($\sigma_1 \geq \sigma_2 \geq ... \geq 0$)

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

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



"Concepts"
AKA Latent dimensions
AKA Latent factors

SVD – Example: Users-to-Movies

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

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

SVD – Example: Users-to-Movies

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

SciFi-concept

Romance-concept

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

SVD – Example: Users-to-Movies

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

U is "user-to-concept" similarity matrix

SciFi-concept

Romance-concept

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

SVD – Example: Users-to-Movies

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

SciFi-concept

"strength" of the SciFi-concept

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

SVD – Example: Users-to-Movies

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

V is "movie-to-concept" similarity matrix

SciFi-concept

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

SciFi-concept

SVD - Interpretation #1

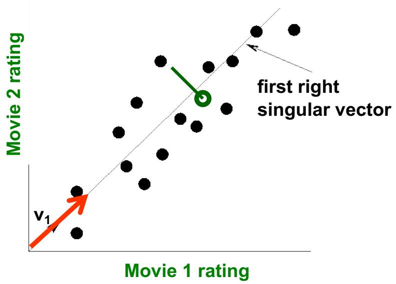
'movies', 'users' and 'concepts':

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

17

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

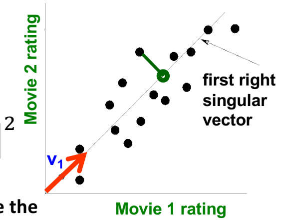
19

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 x_{ij} are the "old" and z_{ij} are the "new" coordinates

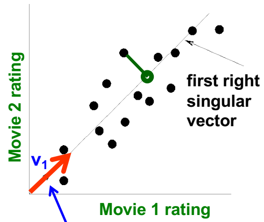


- SVD gives 'best' axis to project on:
  - 'best' = minimizing the reconstruction errors
- In other words, minimum reconstruction error

20

SVD - Interpretation #2

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

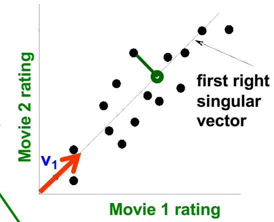


$$\begin{bmatrix} 1&1&1&0&0\\ 3&3&3&0&0\\ 4&4&4&0&0\\ 5&5&5&0&0\\ 0&2&0&4&4\\ 0&0&0&5&5\\ 0&1&0&2&2 \end{bmatrix} = \begin{bmatrix} \mathbf{0.13}&0.02&-0.01\\ \mathbf{0.41}&0.07&-0.03\\ \mathbf{0.55}&0.09&-0.04\\ \mathbf{0.68}&0.11&-0.05\\ 0.15&\mathbf{-0.59}&\mathbf{0.65}\\ 0.07&\mathbf{-0.73}&\mathbf{-0.67}\\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

SVD - Interpretation #2

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

variance (‘spread’) on the v_1 axis

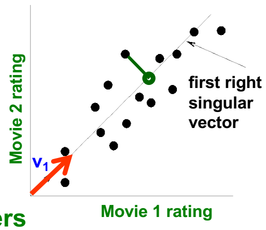


$$\begin{bmatrix} 1&1&1&0&0\\ 3&3&3&0&0\\ 4&4&4&0&0\\ 5&5&5&0&0\\ 0&2&0&4&4\\ 0&0&0&5&5\\ 0&1&0&2&2 \end{bmatrix} = \begin{bmatrix} \mathbf{0.13}&0.02&-0.01\\ \mathbf{0.41}&0.07&-0.03\\ \mathbf{0.55}&0.09&-0.04\\ \mathbf{0.68}&0.11&-0.05\\ 0.15&\mathbf{-0.59}&\mathbf{0.65}\\ 0.07&\mathbf{-0.73}&\mathbf{-0.67}\\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

SVD - Interpretation #2

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

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

SVD - Interpretation #2

More details

- Q: How exactly is dim. reduction done?

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

SVD - Interpretation #2

More details

- **Q: 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**

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

SVD - Interpretation #2

More details

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

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

28

SVD - Interpretation #2

More details

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

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

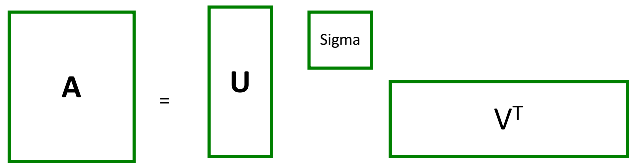
Frobenius norm:

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

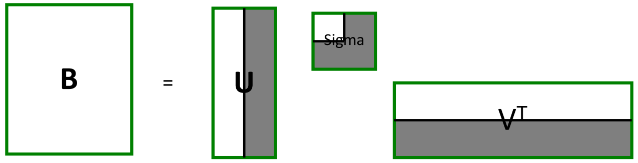
$\|A-B\|_F = \sqrt{\Sigma_{ij} (A_{ij}-B_{ij})^2}$ is "small"

29

SVD – Best Low Rank Approx.



B is best approximation of A



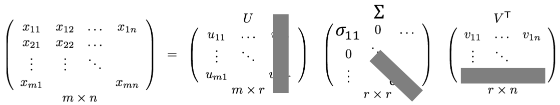
30

SVD – Best Low Rank Approx.

- **Theorem:**
  Let $A = U \Sigma V^T$ and $B = U S V^T$ where
  S = diagonal $r \times r$ 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 $\text{rank}(B)=k$



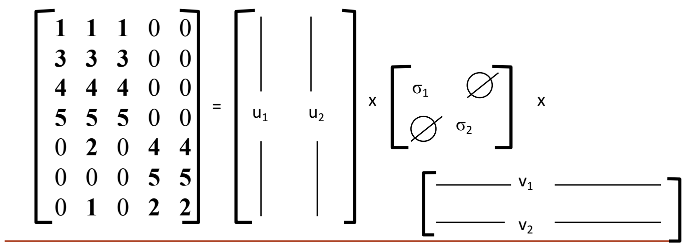
$$\|A-B\|_F = \sqrt{\sum_{ij} (A_{ij}-B_{ij})^2}$$

31

SVD - Interpretation #2

Equivalent:

'spectral decomposition' of the matrix:



32

SVD - Interpretation #2

Equivalent:

'spectral decomposition' of the matrix

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

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 $\mathbf{u}_i$ and $\mathbf{v}_i$ 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' $= \sum_i \sigma_i^2$

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

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

SVD - Complexity

- **To compute SVD:**
  - **O(nm²) or O(n²m) (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

- **SVD: A= U Σ 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**