Dimensionality Reduction 1

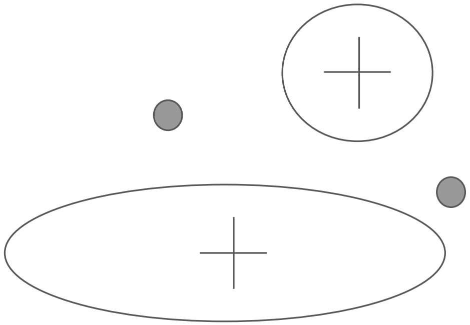
EE412: Foundation of Big Data Analytics

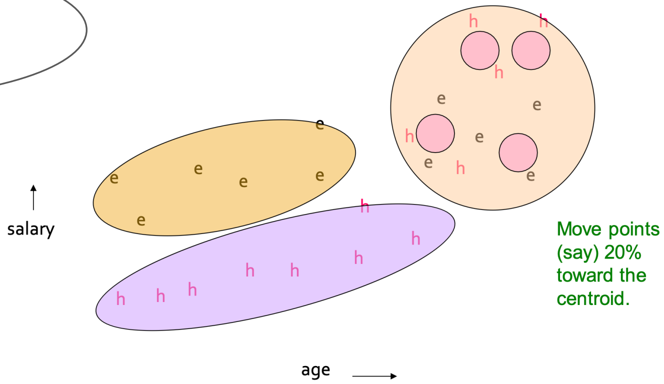
Announcements

- Homeworks
  - HW1 (due: 10/05)
  - HW2 (will be posted at 10/10; due: 11/02)
    - Clustering
    - Dimensionality reduction
    - Recommender systems
  - Midterm (10/19)

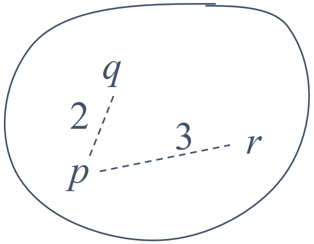
Recap

- BFR Algorithm
  - Cluster representation
  - Three classes of sets
- CURE Algorithm
  - Representative points
- GRGPF Algorithm
  - Rowsum
  - Estimation of a rowsum





$$\textit{rowsum of } p = \sum_{c \in C} d(p, c)^2$$

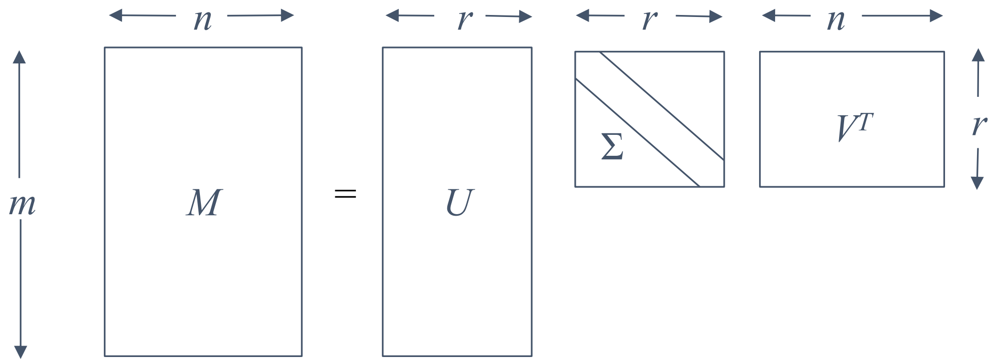


Outline

1. **<u>Dimensionality Reduction</u>**
2. Principal Component Analysis
3. Singular Value Decomposition (SVD)
4. Dimensionality Reduction with SVD

Dimensionality Reduction

- Many sources of data can be viewed as a large matrix
  - E.g., an $m \times n$ matrix for m points with n features
- **Dimensionality reduction**
  - Approximate a matrix by the product of smaller matrices
  - **Example of SVD:** Expect $(m + n + 1)r \ll mn$

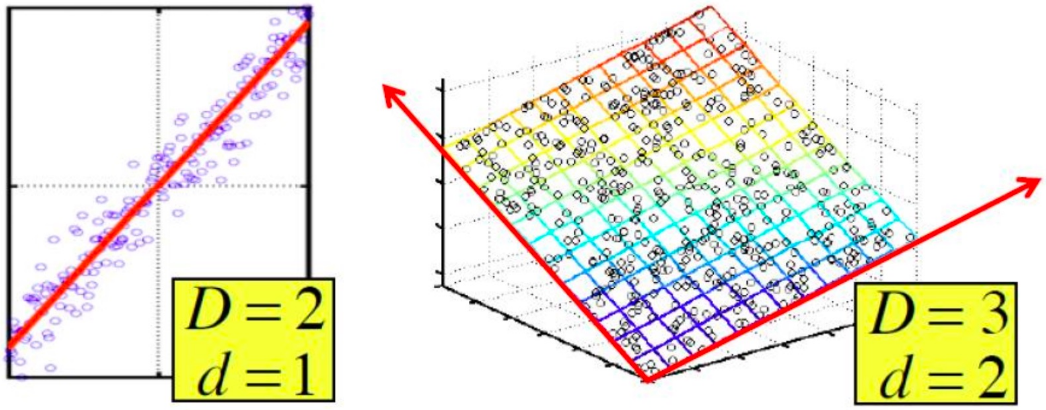


Latent Factors

- **Q:** Do we lose information after dimensionality reduction?

잠재하는

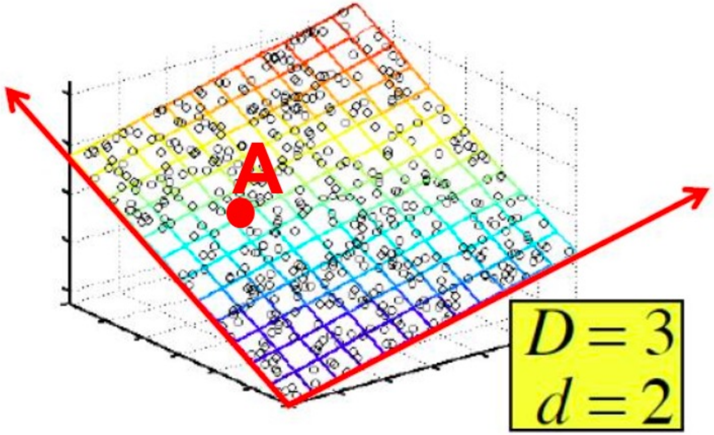
- **A:** Maybe not. There can be hidden, or **latent factors** that
  - Explain why the values are as they appear in the data matrix



Source: Stanford CS246 (2022)

Rank is "Dimensionality"

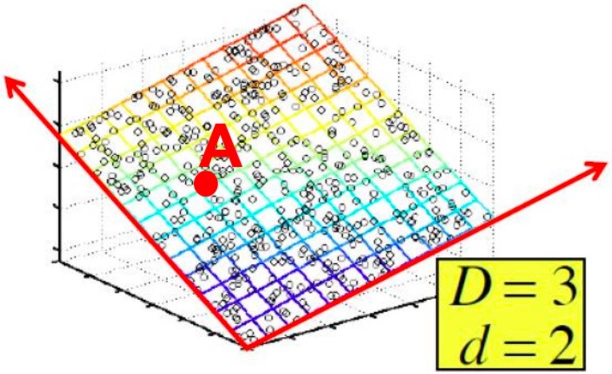
- **Q:** What is the **rank** of a matrix M?
- **A:** Number of linearly independent rows (or columns) of M
  - **Independence:** There is no nonzero linear sum of these rows that is $\mathbf{0}$
- Consider 3 points in a 3D space:
  - $\begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix}$ (1 row per point)
- Rank is two, since ⇔ dependent on A & B
  - $A - B - C = \mathbf{0}$ (in terms of rows)
  - $5c_1 - 3c_2 + c_3 = \mathbf{0}$ (in terms of columns)



Source: Stanford CS246 (2022)

Rank is "Dimensionality"

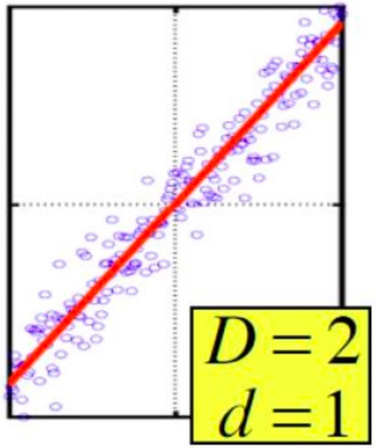
- We can **rewrite the coordinates** efficiently
  - $\begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix}$
- Old basis vectors: $[1\ 0\ 0], [0\ 1\ 0], [0\ 0\ 1]$
  - $A = 1u_1 + 2u_2 + 1u_3$
- **New basis vectors:** $[1\ 2\ 1], [-2\ \ -3\ 1]$
  - $A = 1u_1 + 0u_2$
  - M has new coordinates: A: $[1\ 0], B$: $[0\ 1], C$: $[1\ \ -1]$
  - We reduced the number of dimensions!



Source: Stanford CS246 (2022)

Dimensionality Reduction

- **Goal:** Discover **the axes** of data
  - Represent each point with 1 coordinate
  - Follow the read line (a.k.a. the manifold)
- By doing this, we incur a **bit of error**
  - Since the points do not exactly lie on the line
  - There will be no error if the rank is one



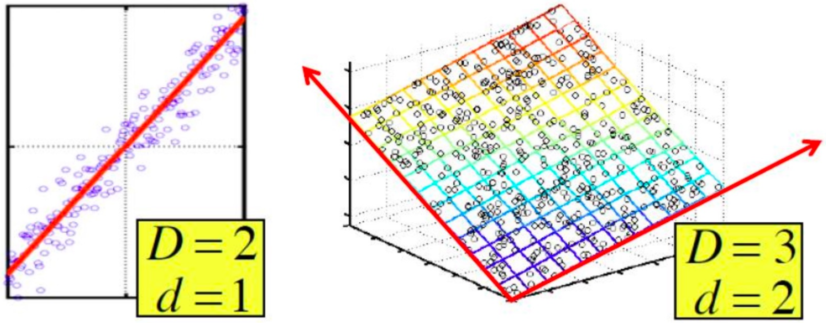
Source: Stanford CS246 (2022)

Outline

1. Dimensionality Reduction
2. **<u>Principal Component Analysis</u>**
3. Singular Value Decomposition (SVD)
4. Dimensionality Reduction with SVD

Principal Component Analysis

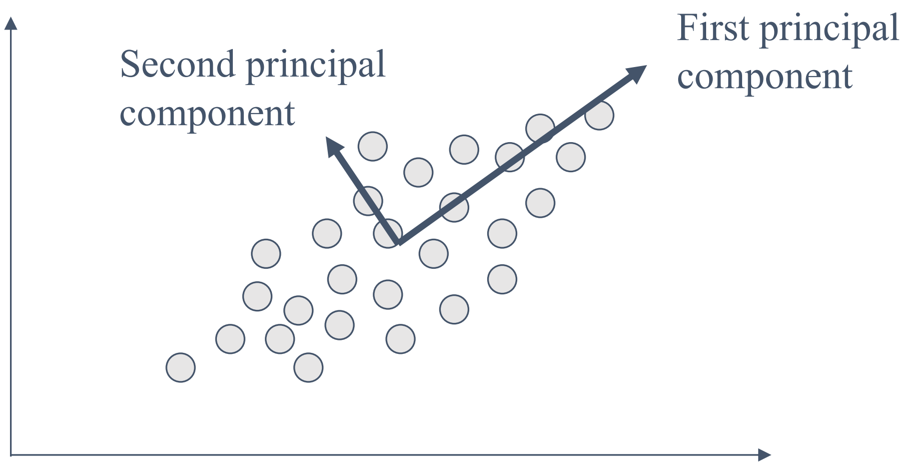
- The axes of these dimensions can be chosen by:
  - **First dimension:** Direction in which the points exhibit the greatest variance
  - **Second dimension:** Direction in which points show the 2nd greatest variance
    - Orthogonal to the first
  - And so on…, until you have enough dimensions that variance is really low



Source: Stanford CS246 (2022)

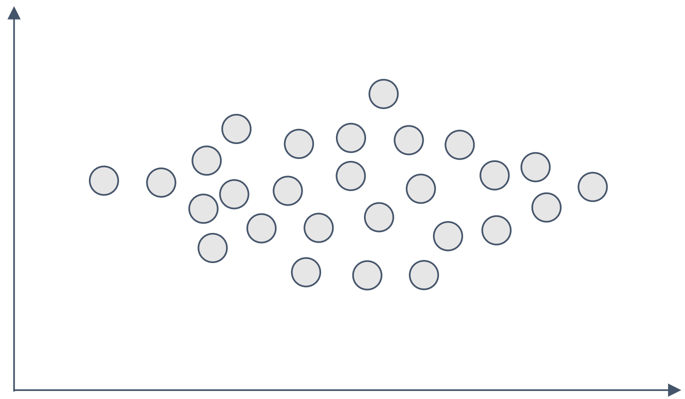
Principal Component Analysis

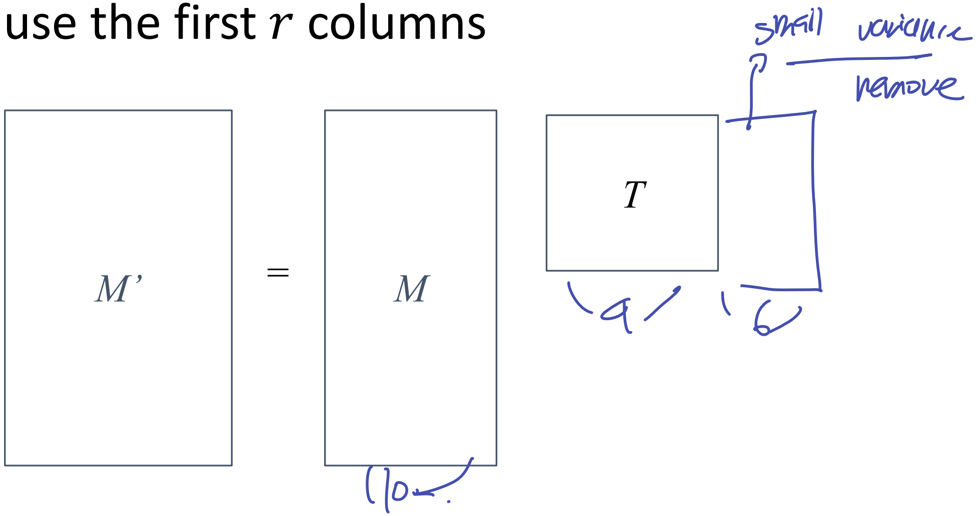
- **Principal component analysis (PCA)** is an algorithm for that
  - The new dimensions are called **principal components**



Output of PCA

- The goal of PCA is to create an $n \times n$ **transformation matrix** T
  - $M' = MT$ puts an $m \times n$ data matrix M on new axes
  - The i-th column of T creates the i-th principal component
    - We want it to be a **unit vector** (i.e., $\|t_i\|_2 = 1$)
  - Dimensionality reduction is done if we use the first r columns





Optimization Problem

- **Q:** How can we efficiently find the principal components?
- In order to maximize variance, the first column t_1 should satisfy

$$t_1 = \operatorname{argmax}_{\|t=1\|} \sum_{i=1}^{m} (M_i' - \bar{\mathbf{m}}')^2$$

- $M_i{}'$: The i-th row of the new data matrix M'
- $\bar{\mathbf{m}}'$: The elementwise mean of all rows in M'

Optimization Problem

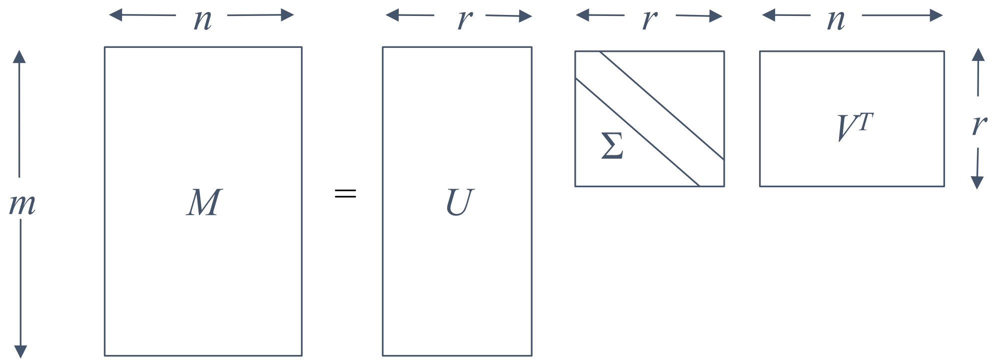
- **Q:** How can we efficiently find the principal components?
- **A:** **Singular value decomposition (SVD)** is a standard way to do that
  - The problem is related to the **eigenvectors** of $M^\top M$
  - PCA has its own way, which is effective if r is very low

Outline

1. Dimensionality Reduction
2. Principal Component Analysis
3. **Singular Value Decomposition (SVD)**
4. Dimensionality Reduction with SVD

Singular Value Decomposition

- Decomposition of any matrix into a product of three matrices
- Choose any number r of **intermediate concepts** (latent factors)
    - In a way that minimizes the reconstruction error
    - The error is zero when $r \geq \mathrm{rank}(M)$



Definition of SVD

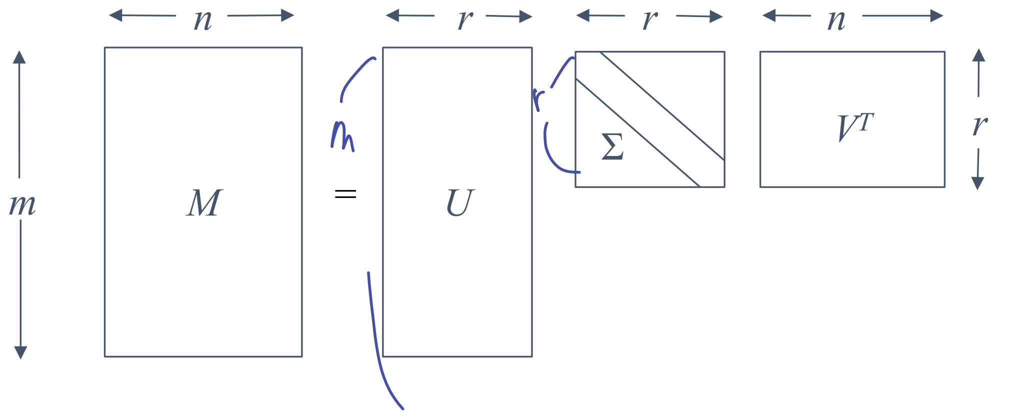
$u = n \times 1$

$v = n \times 1$

$M^T v = s$

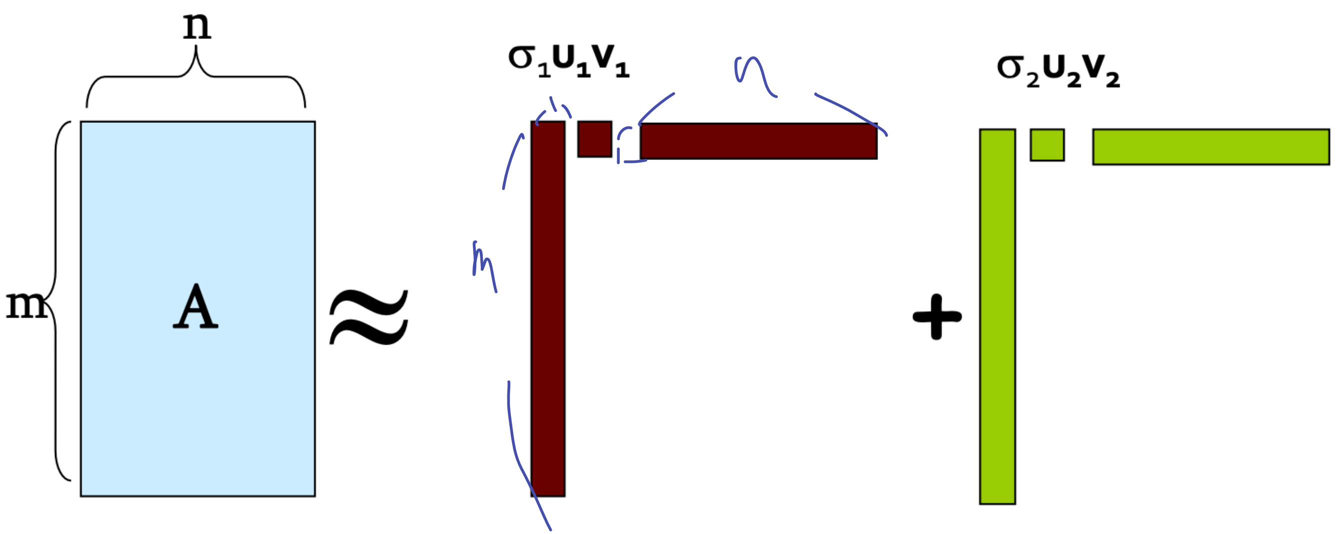
$uv^T = \square$ $n \times n$

- $M \approx U\Sigma V^\top$ (equality if $r \geq \text{rank}(M)$)
  - M is an $m \times n$ **input data matrix**
  - U is an $m \times r$ matrix of **left singular vectors**
  - V is an $n \times r$ matrix of **right singular vectors**
  - Σ is an $r \times r$ diagonal matrix of **singular values** (strength of each "concept")



Different View of SVD

- $M \approx U\Sigma V^\top = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$ (σ_i: scalar, u_i, v_i: vectors)



Source: Stanford CS246 (2022)

Properties of SVD

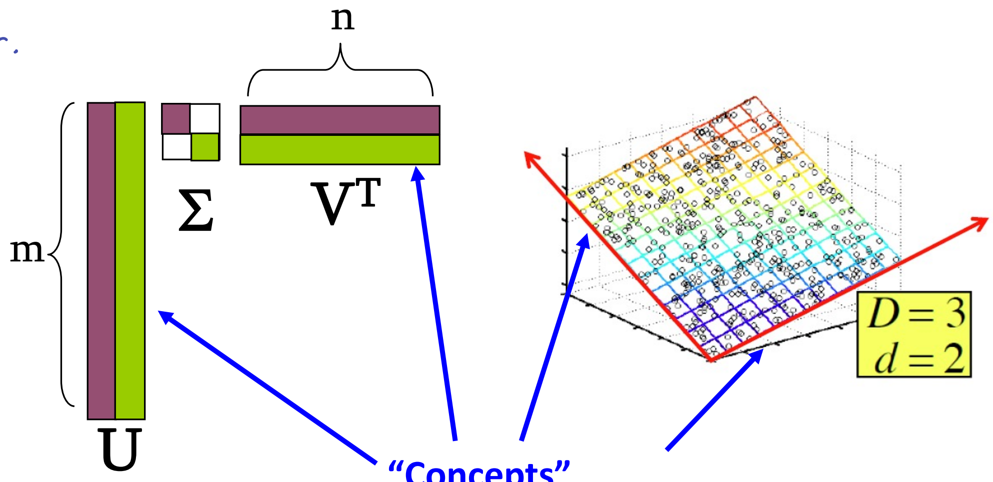
- **Always possible to decompose a real matrix** as $M = U\Sigma V^{\top}$
- U, Σ, V: Unique
- U, V: Column orthonormal
  - Columns are orthogonal unit vectors
  - $U^{\top}U = I; V^{\top}V = I$ (I: Identity matrix)
- Σ: diagonal
  - Entries (singular values) are non-negative, and sorted in decreasing order
  - $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$

Example: SVD

- Consider a matrix of ratings from users; **What does SVD do?**

$$
\begin{array}{c|c}
 & \begin{array}{ccccc} \text{Matrix} & \text{Alien} & \text{Serenity} & \text{Casablanca} & \text{Amelie} \end{array} \\
\begin{array}{c} \text{SciFi} \\ \\ \\ \\ \text{Romnce} \\ \\ \\ \end{array} & \left[\begin{array}{ccccc} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{array}\right]
\end{array}
=
$$

user.



"Concepts"
AKA Latent dimensions
AKA Latent factors

Source: Stanford CS246 (2014)

KAIST

Example: SVD

"strength" of the SciFi-concept

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

Example: SVD

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

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

noise

Example: SVD

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

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

Example: SVD

What concept?

concept

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

movie

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

Example: SVD

Movies, users and concepts:

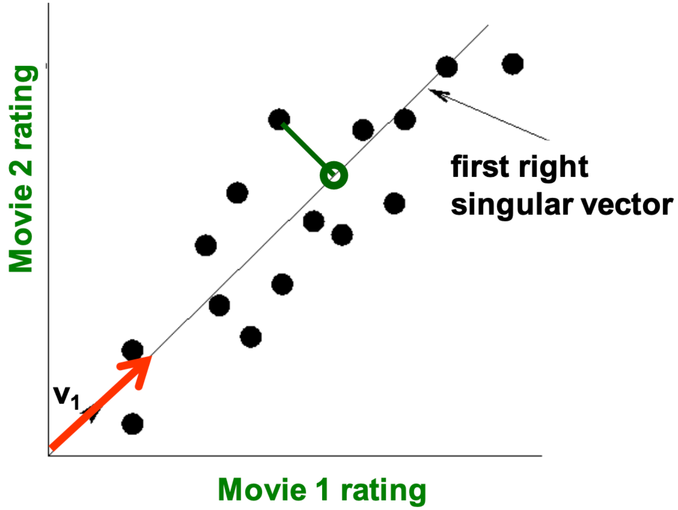
- U: User (row)-to-concept matrix
- V: Movie (column)-to-concept matrix
- Σ: Its diagonal elements: “strength” of each concept

Outline

1. Dimensionality Reduction
2. Principal Component Analysis
3. Singular Value Decomposition (SVD)
4. **Dimensionality Reduction with SVD**

SVD for Dimensionality Reduction

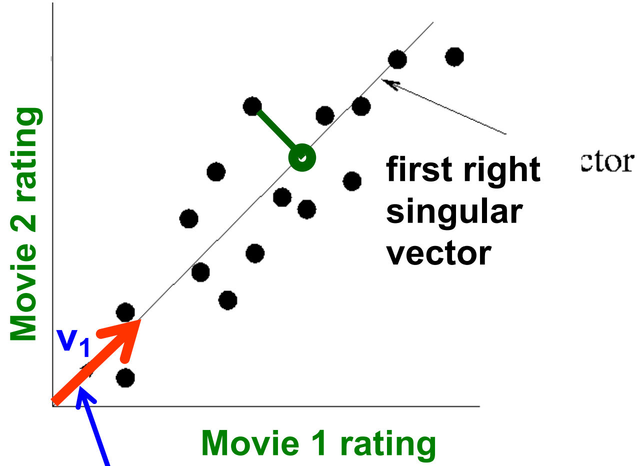
- Let's use fewer coordinates to describe point positions
  - Consider an example of N users for 2 movies of 1 concept
  - Point's position is its location along vector v_1



Source: Stanford CS246 (2022)

Example: SVD

- U is about users, V is about movies
- Each column of V becomes a **new axis**



U

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

V

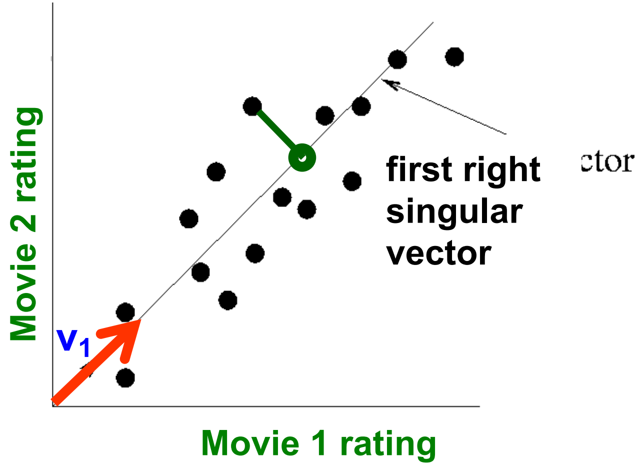
$$\begin{bmatrix} \boxed{\mathbf{0.56} \quad \mathbf{0.59} \quad \mathbf{0.56} \quad 0.09 \quad 0.09} \\ 0.12 \quad -0.02 \quad 0.12 \quad \mathbf{-0.69} \quad \mathbf{-0.69} \\ 0.40 \quad \mathbf{-0.80} \quad 0.40 \quad 0.09 \quad 0.09 \end{bmatrix}$$

$$M = U\Sigma V^{\top}$$

Example: SVD

- First column of V
  - Has the largest variance

variance ('spread') on the v_1 axis

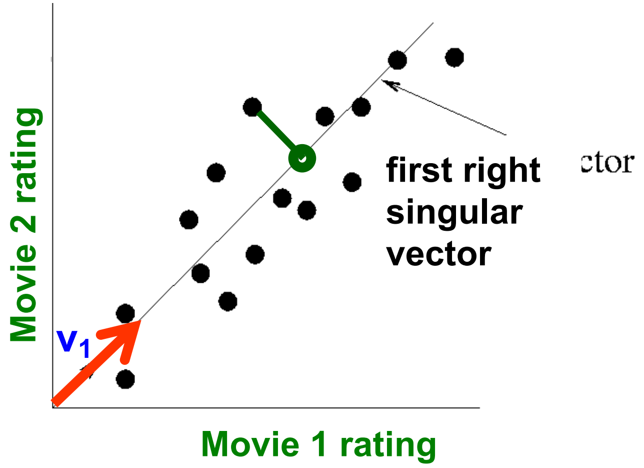


$$
\begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & 0 & 0 \\ \mathbf{3} & \mathbf{3} & \mathbf{3} & 0 & 0 \\ \mathbf{4} & \mathbf{4} & \mathbf{4} & 0 & 0 \\ \mathbf{5} & \mathbf{5} & \mathbf{5} & 0 & 0 \\ 0 & \mathbf{2} & 0 & \mathbf{4} & \mathbf{4} \\ 0 & 0 & 0 & \mathbf{5} & \mathbf{5} \\ 0 &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

$$M = U\Sigma V^{\top}$$

Example: SVD

- $U\Sigma$ 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 Σ□^{T}:

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

$$M = U\Sigma V^{\top}$$

SVD for Dimensionality Reduction

- **Q:** How exactly is dimensionality reduction done by SVD?
- **A: Set smallest singular values to zero**

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

SVD for Dimensionality Reduction

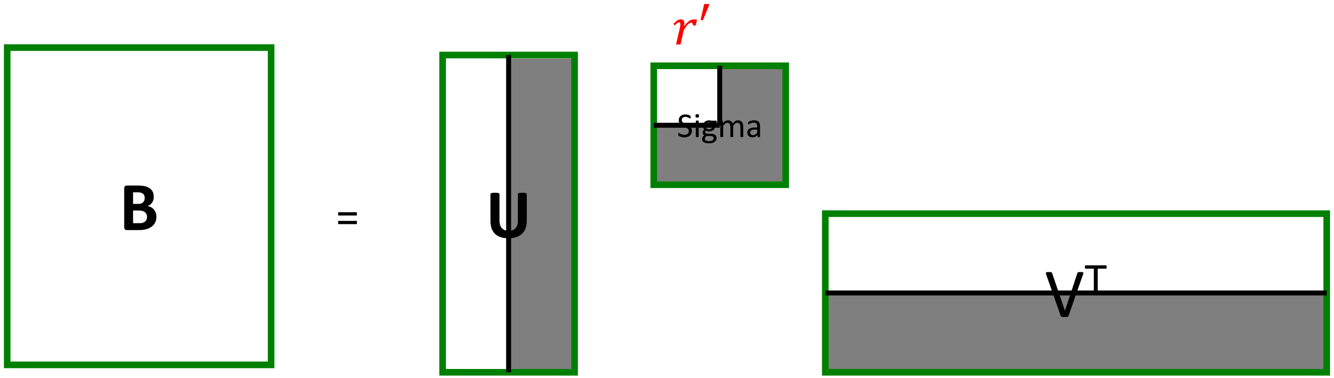
- Reconstruction error is quantified by the **Frobenius norm**:
  - **Definition:** $\|A\|_\mathrm{F} = \sqrt{\sum_{ij} A_{ij}^2}$, **This case:** $\|M - B\|_\mathrm{F} = \sqrt{\sum_{ij} \left(M_{ij} - B_{ij}\right)^2}$

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

Reconstructed data matrix B

Best Low Rank Approximation

- Given $r' < r$, SVD gives the **best axes** to project on
  - Minimizing the reconstruction error $\|M - B\|_F = \sqrt{\sum_{ij}(M_{ij} - B_{ij})^2}$



Source: Stanford CS246 (2022)

Determining the Low Rank

- **Q:** What is a good value for r' (# of latent factors)?
- **A:** Pick r' to have at least 90% of the total energy
  - Let the **energy** of singular values be the sum of their squares
- Back to our example:
  - Singular values are 12.4, 9.5, and 1.3 (total energy = 245.7)
  - If we drop 1.3, whose square is 1.7, the remaining is energy 244.0
  - The remaining energy is over 99% of the total energy

Summary

1. Dimensionality Reduction
    - Latent factors
    - Rank of a matrix
2. Principal Component Analysis
3. Singular Value Decomposition (SVD)
    - Singular values and singular vectors
4. Dimensionality Reduction with SVD
    - Reconstruction error
    - Best low rank approximation