Dimensionality reduction: PCA

Predictive Analytics

Acknowledgment

Materials were adapted from lectures by **Fragkiskos D. Malliaros** (CentraleSupélec, Université Paris-Saclay).

High-Dimensional Data: Why Reduce?

- Real datasets often have **thousands to millions** of features.
  - *Text*: one dimension per vocabulary term (e.g., bag-of-words/TF–IDF).
  - *Social networks*: one dimension per user/connection.
- High dimensionality brings the **curse of dimensionality**:
  - Data are extremely **sparse**; density notions become unreliable $\Rightarrow$ density-based clustering degrades.
  - Algorithmic **complexity grows with** dimensionality d $\Rightarrow$ time/memory blow up.

Can we compress?

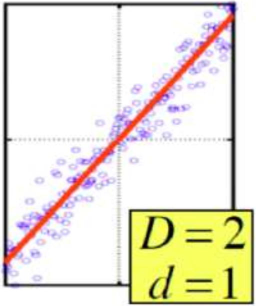
Week days

Customers

Customer	Wed	Thu	Fri	Sat	Sun
George	1	1	1	0	0
Maria	2	2	2	0	0
Ian	5	5	5	0	0
Zoe	0	0	0	2	2
Helen	0	0	0	3	3
Marc	0	0	0	1	1

change axis

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



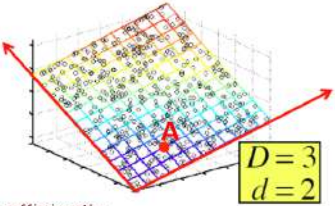
- Rather than representing every point with 2 coordinates we represent each point with1 coordinate (corresponding to the position of the point on the red line)
- By doing this we incur a bit of error as the points do not exactly lie on the line

From 3D to 2D

- Cloud of points 3D space
  - Think of point positions as a matrix:

$$\mathbf{A} = \begin{array}{c} \begin{pmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{pmatrix} \end{array} \begin{array}{l} \text{A} \\ \text{B} \\ \text{C} \end{array}$$

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

Why reduce dimensions

- Aim: uncover the **true dimension** of the data.
- Real datasets are messy, but **useful reduction is still possible**.
- Working assumption: data $=$ **signal** $+$ **noise**; the signal is well-approximated in a **lower-dimensional** subspace/manifold.
- Dimensionality reduction not only **shrinks** data size, it often **reveals structure**—making the informative part more salient.

Why reduce dimensions

- **Find hidden patterns:** discover words or features that often go together (e.g., in texts).
- **Remove useless noise:** throw away features that don't help (not every word is important).
- **Easier to understand:** results are simpler to explain.
- **Better visuals:** easier to draw and see patterns on plots.
- **Less storage:** fewer numbers to save.
- **Faster computing:** algorithms run quicker and use less memory.

Matrix setup for dimensionality reduction

- **Data:** m objects described by n numeric attributes.
- Represent as a matrix $A \in \mathbb{R}^{m \times n}$ (rows = objects, columns = attributes).
- Use **linear algebra** on A to analyze/transform the data.
- **Goal:** build a new matrix $B \in \mathbb{R}^{m \times k}$ with $k \ll n$ such that:
  - it preserves as much information from A as possible;
  - it **reveals structure** (latent patterns, groups, topics) present in A.

From n columns to $k \ll n$ columns

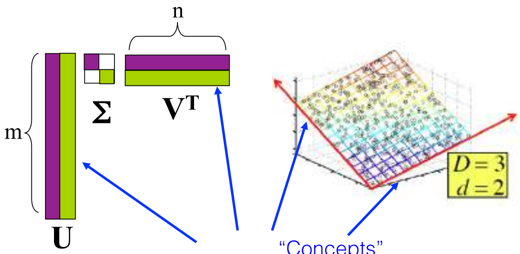
$$\begin{bmatrix} | & | & & | \\ a^{(1)} & a^{(2)} & \dots & a^{(n)} \\ | & | & & | \end{bmatrix}_{m \times n} \longrightarrow \begin{bmatrix} | & & | \\ b^{(1)} & \dots & b^{(k)} \\ | & & | \end{bmatrix}_{m \times k}$$

SVD – Example: Users-to-Movies

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

NETFLIX

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



"Concepts"
AKA Latent dimensions or factors
(e.g., SciFi, Romance concepts)

SVD – Example: Users-to-Movies

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

SciFi-concept

Romance-concept

$$
\begin{array}{c}
\begin{array}{cc}
 & \begin{array}{ccccc}\text{Matrix} & \text{Alien} & \text{Serenity} & \text{Casablanca} & \text{Amelie}\end{array} \\
\begin{array}{c}\text{SciFi} \\ \\ \\ \\ \text{Romance} \\ \\ \\ \end{array} &
\begin{bmatrix}
\mathbf{1} & \mathbf{1} & \mathbf{1} & 0 & 0 \\
\mathbf{3} & \mathbf{3} & \mathbf{3} & 0 & 0 \\
\mathbf{4} & \mathbf{4} & \mathbf{4} & 0 & 0 \\
\mathbf{5} & \mathbf{5} & \mathbf{5} & 0 & 0 \\
0 & 0 & 0 & \mathbf{4} & \mathbf{4} \\
0 & 0 & 0 & \mathbf{5} & \mathbf{5} \\
0 & 0 & 0 & \mathbf{2} & \mathbf{2}
\end{bmatrix}
\end{array}
\end{array}
=
\underbrace{\begin{bmatrix}
\mathbf{0.14} & 0.00 \\
\mathbf{0.42} & 0.00 \\
\mathbf{0.56} & 0.00 \\
\mathbf{0.70} & 0.00 \\
0.00 & \mathbf{0.60} \\
0.00 & \mathbf{0.75} \\
0.00 & \mathbf{0.30}
\end{bmatrix}}_{\text{matrix } U}
\times
\underbrace{\begin{bmatrix}
\mathbf{12.4} & 0 \\
0 & \mathbf{9.5}
\end{bmatrix}}_{\text{matrix } \Sigma}
\times
\underbrace{\begin{bmatrix}
\mathbf{0.58} & \mathbf{0.58} & \mathbf{0.58} & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & \mathbf{0.71} & \mathbf{0.71}
\end{bmatrix}}_{\text{matrix } V^T}
$$

SVD – Example: Users-to-Movies

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

SciFi-concept

Romance-concept

U is “user-to-concept” similarity matrix

$$
\begin{array}{c c}
 & \begin{array}{ccccc} \text{Matrix} & \text{Alien} & \text{Serenity} & \text{Casablanca} & \text{Amelie} \end{array} \\
\begin{array}{c} \text{SciFi} \\ \\ \\ \\ \text{Romance} \\ \\ \\ \end{array} &
\begin{bmatrix}
\mathbf{1} & \mathbf{1} & \mathbf{1} & 0 & 0 \\
\mathbf{3} & \mathbf{3} & \mathbf{3} & 0 & 0 \\
\mathbf{4} & \mathbf{4} & \mathbf{4} & 0 & 0 \\
\mathbf{5} & \mathbf{5} & \mathbf{5} & 0 & 0 \\
0 & 0 & 0 & \mathbf{4} & \mathbf{4} \\
0 & 0 & 0 & \mathbf{5} & \mathbf{5} \\
0 & 0 & 0 & \mathbf{2} & \mathbf{2}
\end{bmatrix}
\end{array}
=
\begin{bmatrix}
\mathbf{0.14} & 0.00 \\
\mathbf{0.42} & 0.00 \\
\mathbf{0.56} & 0.00 \\
\mathbf{0.70} & 0.00 \\
0.00 & \mathbf{0.60} \\
0.00 & \mathbf{0.75} \\
0.00 & \mathbf{0.30}
\end{bmatrix}
\times
\begin{bmatrix}
\mathbf{12.4} & 0 \\
0 & \mathbf{9.5}
\end{bmatrix}
\times
\begin{bmatrix}
\mathbf{0.58} & \mathbf{0.58} & \mathbf{0.58} & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & \mathbf{0.71} & \mathbf{0.71}
\end{bmatrix}
$$

matrix U

matrix V^T

SVD – Example: Users-to-Movies

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

$$
\begin{array}{cc}
 & \begin{array}{ccccc} \text{Matrix} & \text{Alien} & \text{Serenity} & \text{Casablanca} & \text{Amelie} \end{array} \\
\begin{array}{c} \\ \\ \text{SciFi} \\ \\ \\ \text{Romance} \\ \\ \end{array} &
\begin{bmatrix}
\mathbf{1} & \mathbf{1} & 1 & 0 & 0 \\
3 & 3 & 3 & 0 & 0 \\
4 & 4 & 4 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 4 & 4 \\
0 & 0 & 0 & 5 & 5 \\
0 & 0 & 0 & 2 & 2
\end{bmatrix}
\end{array}
=
\begin{array}{c}
\begin{array}{cc} \text{SciFi-concept} & \text{Romance-concept} \end{array} \\
\begin{bmatrix}
\mathbf{0.14} & 0.00 \\
\mathbf{0.42} & 0.00 \\
\mathbf{0.56} & 0.00 \\
\mathbf{0.70} & 0.00 \\
0.00 & \mathbf{0.60} \\
0.00 & \mathbf{0.75} \\
0.00 & \mathbf{0.30}
\end{bmatrix}
\end{array}
\times
\begin{bmatrix}
12.4 & 0 \\
0 & 9.5
\end{bmatrix}
\times
\begin{array}{cc}
\begin{array}{c} \text{SciFi-concept} \\ \text{Romance-concept} \end{array} &
\begin{bmatrix}
\mathbf{0.58} & \mathbf{0.58} & \mathbf{0.58} & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & \mathbf{0.71} & \mathbf{0.71}
\end{bmatrix}
\end{array}
$$

matrix V^T

V is "movie-to-concept" similarity matrix

SVD – Example: Users-to-Movies

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

SciFi-concept

Romance-concept

$$
\begin{array}{c|ccccc}
 & \text{Matrix} & \text{Alien} & \text{Serenity} & \text{Casablanca} & \text{Amelie} \\
\hline
\text{SciFi} & 1 & 1 & 1 & 0 & 0 \\
 & 3 & 3 & 3 & 0 & 0 \\
 & 4 & 4 & 4 & 0 & 0 \\
 & 5 & 5 & 5 & 0 & 0 \\
\text{Romance} & 0 & 0 & 0 & 4 & 4 \\
 & 0 & 0 & 0 & 5 & 5 \\
 & 0 & 0 & 0 & 2 & 2
\end{array}
=
\begin{bmatrix}
\mathbf{0.14} & 0.00 \\
\mathbf{0.42} & 0.00 \\
\mathbf{0.56} & 0.00 \\
\mathbf{0.70} & 0.00 \\
0.00 & \mathbf{0.60} \\
0.00 & \mathbf{0.75} \\
0.00 & \mathbf{0.30}
\end{bmatrix}
\times
\begin{bmatrix}
\mathbf{12.4} & 0 \\
0 & \mathbf{9.5}
\end{bmatrix}
\times
\begin{bmatrix}
\mathbf{0.58} & \mathbf{0.58} & \mathbf{0.58} & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & \mathbf{0.71} & \mathbf{0.71}
\end{bmatrix}
$$

Σ is the "concept strength" matrix

"strength" of the SciFi-concept

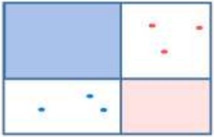
SciFi-concept

Romance-concept

matrix V^T

A More Realistic Example

- User-Movie matrix

A = 

- There are two prototype users and movies but they are noisy

SVD – Example: Users-to-Movies

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

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

SVD – Example: Users-to-Movies

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

SciFi-concept

Romance-concept

The first two vectors are
more or less unchanged

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

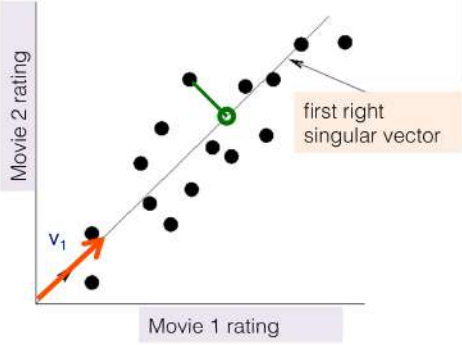
3rd concept, but of low strength

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

How to do Dimensionality Reduction with SVD?



Example of different users based on the rating on movies 1 and 2

How to chose v_1?

Minimize reconstruction error

- 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

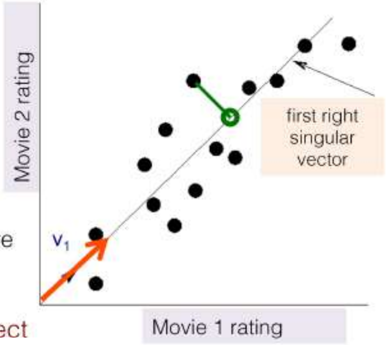
SVD and Dimensionality Reduction

- Goal: minimize the sum of reconstruction errors

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

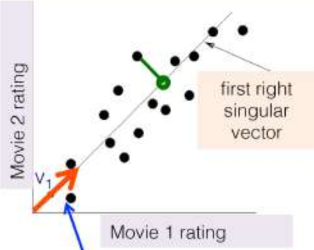
  - where $\mathbf{x}_{ij}$ are the 'old' and $\mathbf{z}_{ij}$ are the 'new' coordinates

- SVD gives the best axis to project on
  - 'best' = minimum sum of squares of projection errors



SVD - Interpretation #2

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

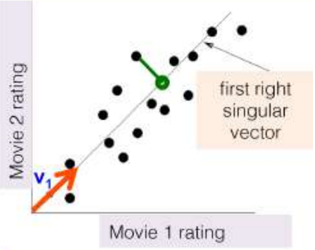


$$
\begin{bmatrix}
\mathbf{1} & \mathbf{1} & \mathbf{1} & 0 & 0 \\
\mathbf{3} & \mathbf{3} & \mathbf{3} & 0 & 0 \\
\mathbf{4} & \mathbf{4} & \mathbf{4} & 0 & 0 \\
\mathbf{5} & \mathbf{5} & \mathbf{5} & 0 & 0 \\
0 & \mathbf{2} & 0 & \mathbf{4} & \mathbf{4} \\
0 & 0 & 0 & \mathbf{5} & \mathbf{5} \\
0 &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

Defines the new projection space v_1

SVD - Interpretation #2

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



variance ('spread') on the v_1 axis

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

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

Example

More details

- Q: How exactly is dimensionality reduction done?
- A: Compute SVD

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

Example

More details

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

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

Example

More details

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

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

Example

More details

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

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

Example

More details

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

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

Example

More details

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

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

$$\| \mathbf{A} \|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |A_{ij}|^2}$$

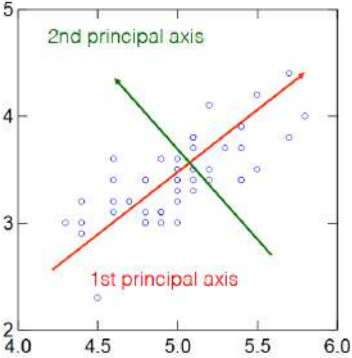
Frobenius norm:

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

PCA: Introduction

- **(Almost) like SVD:** PCA is closely related to the singular value decomposition.
- **Goal:** find a low-dimensional subspace so that projecting the data loses as little information as possible.
- **Principle:** choose directions (principal components) that **maximize variance**.
- **Standardize features** (zero-mean, unit-variance) when scales differ.

PCA



Input: 2-d dimensional points

Output:

1st (right) principal axis (vector): direction of maximal variance

2nd (right) principal axis (vector): direction of maximal variance, after removing the projection of the data along the first principal axis