Applications of SVD

Linear Algebra Honours course

Jonathan Nilsson

Linköping University

Part I

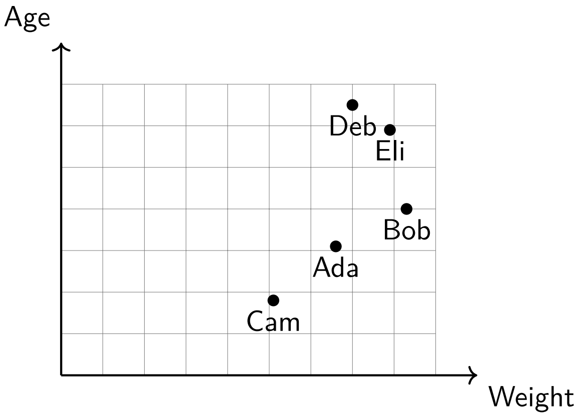
Data analysis with SVD

Cosine similarity

$$\text{cosSim}(u, v) := \frac{u \bullet v}{\|u\| \cdot \|v\|}.$$

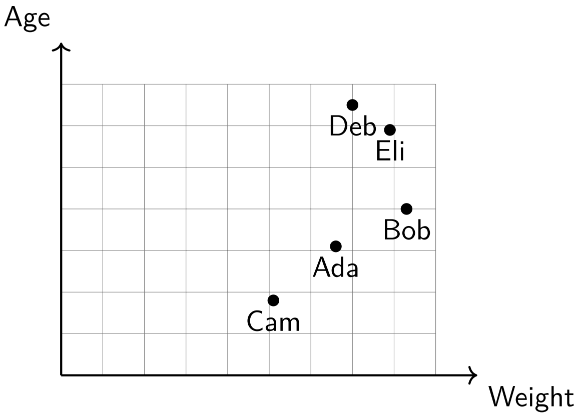
Weight vs age of Ada, Bob, Cam, Deb, Eli.

$$A = \begin{pmatrix} 66 & 31 \\ 83 & 40 \\ 51 & 18 \\ 70 & 65 \\ 79 & 59 \end{pmatrix}$$



Weight vs age of Ada, Bob, Cam, Deb, Eli.

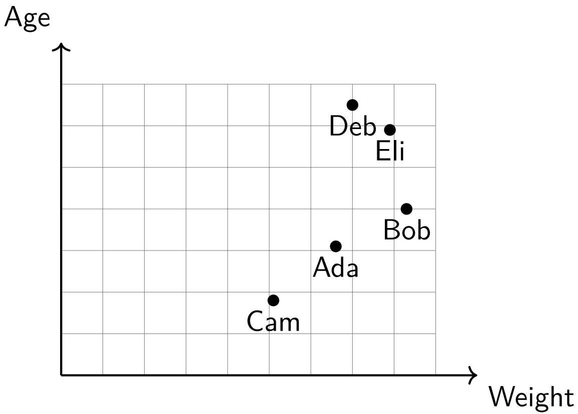
$$A = \begin{pmatrix} 66 & 31 \\ 83 & 40 \\ 51 & 18 \\ 70 & 65 \\ 79 & 59 \end{pmatrix}$$



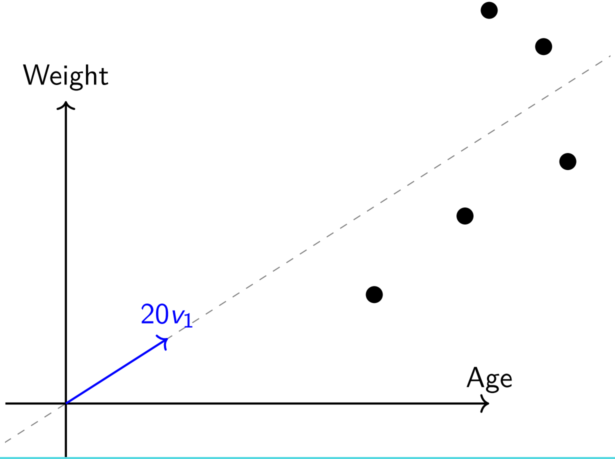
$$\text{cosSim}(A_1, A_2) = 0.955$$

Weight vs age of Ada, Bob, Cam, Deb, Eli.

$$A = \begin{pmatrix} 66 & 31 \\ 83 & 40 \\ 51 & 18 \\ 70 & 65 \\ 79 & 59 \end{pmatrix}$$



$$\text{cosSim}(A_1, A_2) = 0.955 \qquad \text{cosSim}(\text{Bob}, \text{Deb}) = 0.953$$


Weight
Age
$20v_1$

Movie ratings example

	Interstellar	Mean Girls	The Matrix	Borat	Ghost Busters
Alex	5	2	4	2	4
Bert	5	1	4	1	4
Cleo	4	5	2	4	5
Dany	3	5	3	5	5
Elle	5	2	5	2	3
Fred	5	3	4	2	4

Movie ratings example

	Interstellar	Mean Girls	The Matrix	Borat	Ghost Busters
Alex	5	2	4	2	4
Bert	5	1	4	1	4
Cleo	4	5	2	4	5
Dany	3	5	3	5	5
Elle	5	2	5	2	3
Fred	5	3	4	2	4

Let A be this 6×5 matrix of coefficients.

SVD of the movie matrix

$$A = U\Sigma V^T$$

$$U = \begin{pmatrix} 0.39 & 0.26 & 0.12 & 0.17 & -0.46 \\ 0.36 & 0.45 & 0.41 & 0.52 & 0.15 \\ 0.44 & -0.49 & 0.48 & -0.38 & -0.35 \\ 0.45 & -0.57 & -0.42 & 0.49 & 0.22 \\ 0.39 & 0.37 & -0.64 & -0.30 & -0.29 \\ 0.41 & 0.16 & 0.09 & -0.46 & 0.71 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 20.1 & & & & \\ & 5.60 & & & \\ & & 1.57 & & \\ & & & 0.96 & \\ & & & & 0.35 \\ & & & & \end{pmatrix} \quad V = \begin{pmatrix} 0.54 & 0.45 & 0.33 & -0.42 & -0.47 \\ 0.38 & -0.56 & -0.05 & -0.61 & 0.41 \\ 0.44 & 0.47 & -0.67 & 0.14 & 0.36 \\ 0.34 & -0.50 & -0.41 & 0.28 & -0.63 \\ 0.51 & -0.13 & 0.53 & 0.60 & 0.30 \end{pmatrix}$$

Singular values correspond to "latent features"

$\sigma_1 = 20.1 \sim$ Movie popularity

Singular values correspond to "latent features"

$\sigma_1 = 20.1 \sim$ Movie popularity

$\sigma_2 = 5.60 \sim$ Movie genre

Right singular vectors in movie-space

$$v_1 = (0.54, 0.38, 0.44, 0.34, 0.51)^T \sim \text{Average popularity for each movie}$$

Right singular vectors in movie-space

$$v_1 = (0.54, 0.38, 0.44, 0.34, 0.51)^T \sim \text{Average popularity for each movie}$$

$$v_2 = (0.45, -0.56, 0.47, -0.50, -0.13)^T \sim \text{Genre for each movie}$$

Left singular vectors in user-space

$$u_1 = (0.39, 0.36, 0.44, 0.45, 0.39, 0.41)^T \sim \text{Average ratings of each user}$$

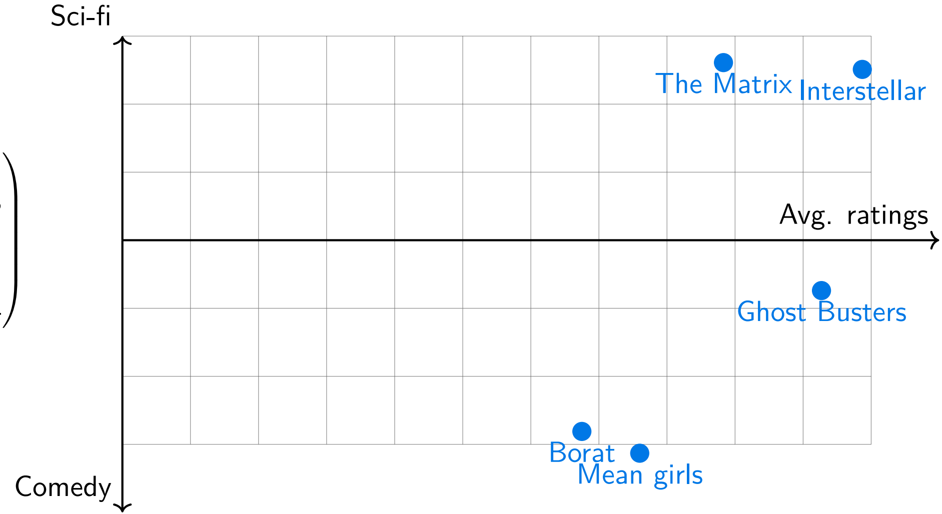
Left singular vectors in user-space

$$u_1 = (0.39, 0.36, 0.44, 0.45, 0.39, 0.41)^T \sim \text{Average ratings of each user}$$

$$u_2 = (0.26, 0.45, -0.49, -0.57, 0.37, 0.1)^T \sim \text{Genre preference for each user}$$

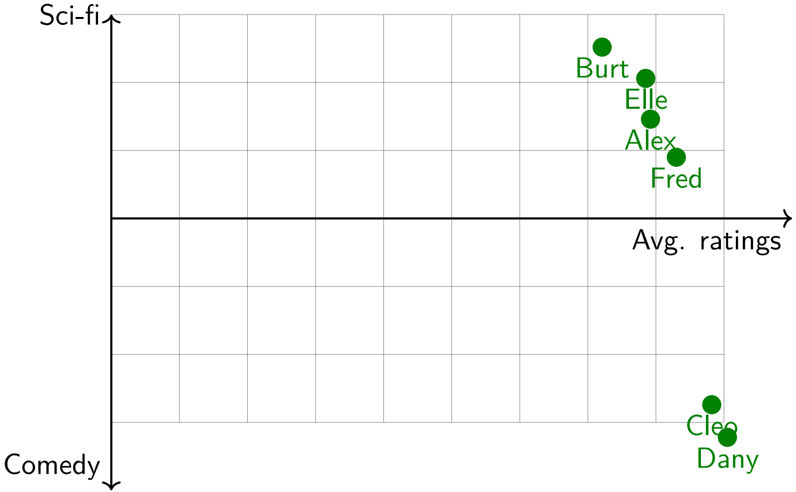
Movies in user-space

$$\begin{pmatrix} 10.87 & 2.51 \\ 7.60 & -3.13 \\ 8.83 & 2.61 \\ 6.75 & -2.81 \\ 10.27 & -0.74 \end{pmatrix}$$



Users in movie-space

$$\begin{pmatrix} 7.92 & 1.46 \\ 7.21 & 2.52 \\ 8.82 & -2.74 \\ 9.05 & -3.22 \\ 7.85 & 2.06 \\ 8.30 & 0.90 \end{pmatrix}$$



Part II

Facial recognition example

Faces of MAI

Input: 70 faces of employees at the math department MAI at LiU.

Faces of MAI

Input: 70 faces of employees at the math department MAI at LiU.

Original $\rightarrow$ Aligned $\rightarrow$ Resized and decolored $\rightarrow$ Mean subtracted

.

The average face

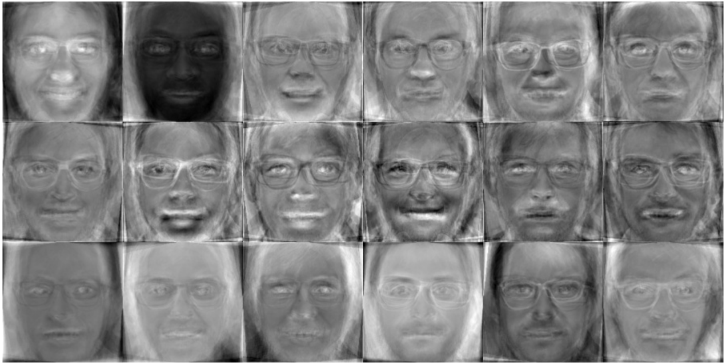
$$c = \frac{1}{70}\sum_{i=1}^{70} z_i =$$ 

Performing SVD

A is now an 70×22500-matrix. We find the eigenvectors of $A^T A$.

Performing SVD

A is now an 70×22500-matrix. We find the eigenvectors of $A^T A$.



The first 18 eigenfaces.

Projection onto the principal components

$k = 10$

Projection onto the principal components

$k = 10$

$k = 25$

Projection onto the principal components

$k = 10$

$k = 25$

$k = 40$

Projection onto the principal components

$k = 10$

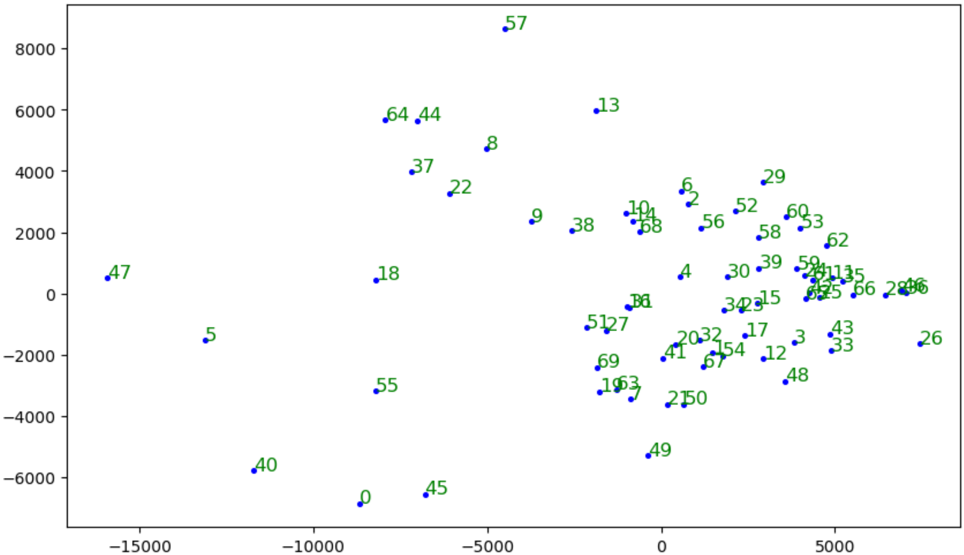
$k = 25$

$k = 40$

$k = 55$

Projection onto the principal components

$k = 10$ $k = 25$ $k = 40$ $k = 55$ $k = 70$


8000
6000
4000
2000
0
−2000
−4000
−6000
−15000
−10000
−5000
0
5000
57
64
44
13
8
37
22
29
6
2
52
60
53
9
38
56
58
62
47
18
4
30
39
66
26
15
51
27
5
17
3
43
33
12
41
67
69
48
55
7
50
49
40
45
0

Further applications

Here are a few examples of- applications where similar techniques can be used:

- Ancestry analysis
- Latent semantic analysis
- Recommender systems
- Signal processing
- Climate data analysis
- Word embeddings

Part III

Total least squares

Fitting a k-plane to a set of points

The total least square problem

Given $x_i, \ldots, x_m \in \mathbb{R}^n$, what **affine subspace** S of dimension k lies closest to the points in the sense that

$$\sum_{i=1}^{m} d(x_i, S)^2 \quad \text{is minimal?}$$

Fitting a k-plane to a set of points

The total least square problem

Given $x_i, \ldots, x_m \in \mathbb{R}^n$, what **affine subspace** S of dimension k lies closest to the points in the sense that

$$\sum_{i=1}^{m} d(x_i, S)^2 \quad \text{is minimal?}$$

Here $d(x_i, S) = \min_{s \in S} \|x_i - S\|$.

The answer

To fit a k-plane to points $x_1, \ldots, x_m \in \mathbb{R}^n$ do:

The answer

To fit a k-plane to points $x_1, \ldots, x_m \in \mathbb{R}^n$ do:

- Let $c = \frac{1}{m} \sum_{i=1}^{m} x_i$ be the average of the points.

The answer

To fit a k-plane to points $x_1, \ldots, x_m \in \mathbb{R}^n$ do:

- Let $c = \frac{1}{m} \sum_{i=1}^{m} x_i$ be the average of the points.
- Put the normalized data-points $x_i - c$ as columns in an $m \times n$-matrix A.

The answer

To fit a k-plane to points $x_1, \ldots, x_m \in \mathbb{R}^n$ do:

- Let $c = \frac{1}{m} \sum_{i=1}^{m} x_i$ be the average of the points.
- Put the normalized data-points $x_i - c$ as columns in an $m \times n$-matrix A.
- Find the SVD of A, let $v_1, \ldots, v_r$ be the right singular vectors.

The answer

To fit a k-plane to points $x_1, \ldots, x_m \in \mathbb{R}^n$ do:

- Let $c = \frac{1}{m} \sum_{i=1}^{m} x_i$ be the average of the points.
- Put the normalized data-points $x_i - c$ as columns in an $m \times n$-matrix A.
- Find the SVD of A, let $v_1, \ldots, v_r$ be the right singular vectors.
- The affine k-dimensional subspace that best approximates the data in the total least squares sense is

$$S = c + \mathrm{span}(v_1, \ldots, v_k)$$

Which point in $\mathbb{R}^3$ lies closest to $(1, 5)$, $(1, 2)$, $(4, 2)$?

Which point in $\mathbb{R}^3$ lies closest to $(1, 5)$, $(1, 2)$, $(4, 2)$?

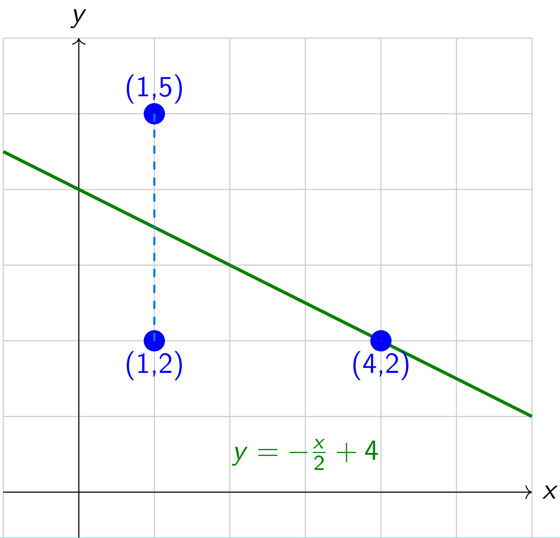
Old least squares method:
Ansatz $y = kx + m$ $\rightarrow$ linear system $\rightarrow$ least square solution.

Which point in $\mathbb{R}^3$ lies closest to $(1, 5)$, $(1, 2)$, $(4, 2)$?

Old least squares method:
Ansatz $y = kx + m$ $\rightarrow$ linear system $\rightarrow$ least square solution.

Total least squares:
Recenter points at origin $\rightarrow$ SVD on normalized data-matrix $\rightarrow$ singular vectors gives directions of affine subspace

Regular least squares method



Total least squares method