Computational Intelligence Laboratory

Lecture 2

Singular Value Decomposition

Thomas Hofmann

ETH Zurich – cil.inf.ethz.ch

March 10, 2017

Section 1

Collaborative Filtering

Collaborative Filtering

- **Recommender systems**
  - analyze patterns of interest in items (products, movies, ...)
  - provide personalized recommendations for users
- **Collaborative Filtering**
  - exploit collective data from many users
  - generalize across users and – possibly – across items
- Applications:
  - Amazon, Netflix, Pandora, online advertising, etc.
  - special case of **algorithmic selection**

Netflix Data

- Input: user-item preferences stored in a matrix
  - **rows = users**, **columns = items**

18,000 movies

480,000 users

x	1	1	x	...	x
x	x	x	5	...	x
x	x	3	x	...	x
x	4	3	x	...	2
...	x	x	x	...	x
x	5	x	1	...	x
x	x	3	3	...	x
x	1	x	x	...	2

  - 1-5 star rating of movies. x denotes a missing value.
  - predict missing values = matrix completion

Matrix Completion

- How can we fill in missing values?
- **Statistical model** with $k \ll m \cdot n$ parameters
  - $m \times n$: dimensionality of rating matrix
  - introduces coupling between entries
  - infer missing entries from observed ones
- **Low Rank** decomposition
  - find best approximation with low rank r
  - entries in decomposition: $k \leq r \cdot (m + n)$

Section 2

Singular Value Decomposition

Motivation

- **Singular Value Decomposition** (SVD)
  - widely used technique to decompose a matrix $\mathbf{A}$
    - least squares, matrix approximation, etc.
  - exposes useful and interesting properties of $\mathbf{A}$
    - rank, range, null-space, orthogonal basis of col/row space
  - closely related to eigen-decomposition
  - indespensable analysis tool

Singular Value Decomposition

- Any rectangular matrix $\mathbf{A}$ can be decomposed into

$$\underset{m \times n}{\mathbf{A}} = \underset{m \times m}{\mathbf{U}} \cdot \underset{m \times n}{\mathbf{D}} \cdot \underset{n \times n}{\mathbf{V}^\top}$$

  - with $\mathbf{U}$, $\mathbf{V}$ orthogonal
  - and with $\mathbf{D}$ diagonal, $s := \min\{m, n\}$

$$\mathbf{D} = \mathrm{diag}(\sigma_1, \ldots, \sigma_s), \quad \sigma_1 \geq \cdots \geq \sigma_s \geq 0$$

Singular Value Decomposition - cont'd

- Columns of $\mathbf{U}$ and $\mathbf{V}$: left/right **singular vectors**
- Entries of $\mathbf{D}$: **singular values**
  - number of distinct singular values $\leq s = \min\{n, m\}$
  - σ_i with two (or more) linearly independent left (or right) singular vectors = **degenerate**
  - singular vectors for non-degenerate σ_i: unique up to sign
  - singular vectors for degenerate σ_i: orthonormal basis (non-unique) of span (unique)

Range and Rank of a Matrix

- Range of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ (also: column space)

$$\text{range}(\mathbf{A}) = \{\mathbf{z} : \exists \mathbf{x} \text{ s.t. } \mathbf{z} = \mathbf{A}\mathbf{x}\} \; = \; \text{span}\left(\text{columns of } \mathbf{A}\right)$$

- **Rank** of a matrix

$$\text{rank}(\mathbf{A}) = \dim\left(\text{range}(\mathbf{A})\right)$$

  - rank of a diagonal matrix = number of non-zero entries

- Rank and SVD

$$\text{rank}(\mathbf{A}) = r \quad \Longleftrightarrow \quad \sigma_r > 0 \; \wedge \sigma_{r+1} = \sigma_{r+2} = \cdots = 0$$

Kernel of a Matrix

- Kernel of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$

$$\text{kernel}(\mathbf{A}) = \{\mathbf{x} \ \text{ s.t. } \ \mathbf{A}\mathbf{x} = \mathbf{0}\}$$

- Rank-nullity theorem

$$\dim(\text{kernel}(\mathbf{A})) + \dim(\text{range}(\mathbf{A})) = n$$

Example

$\mathbf{A}$ with $\text{rank}(\mathbf{A}) = 3$

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{5} & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{U} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \text{range}$$

$$\mathbf{V} = \begin{bmatrix} 0 & \frac{1}{\sqrt{5}} & 0 & 0 & -\frac{2}{\sqrt{5}} \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & \frac{2}{\sqrt{5}} & 0 & 0 & \frac{1}{\sqrt{5}} \end{bmatrix}, \quad \text{null space}$$

Reduced Singular Value Decomposition

- Assume w.l.o.g. that $n \geq m$ (otherwise transpose)
- Rightmost $(n-m)$ columns of $\mathbf{V}$ are multiplied by zeros in $\mathbf{D}$
  - (partial) basis of kernel of $\mathbf{A}$
  - need not be computed, if no interest in kernel
- **Reduced** SVD

$$\mathbf{A} = \underbrace{\mathbf{U}}_{m\times m} \cdot \underbrace{\mathrm{diag}(\sigma_1, \ldots, \sigma_m)}_{m\times m} \cdot \underbrace{\tilde{\mathbf{V}}^\top}_{m\times n}, \quad \tilde{\mathbf{V}} = \left[\mathbf{v}_1 \cdots \mathbf{v}_m\right] \in \mathbb{R}^{n\times m}$$

Existence of SVD: Construction

- (0) Initialize $\mathcal{U} = \emptyset$, $\mathcal{V} = \emptyset$, $\mathbf{B}_1 = \mathbf{A}$, $i = 1$
- (1) If $\mathbf{B}_i = \mathbf{0}$ extend $\mathcal{U}$, $\mathcal{V}$ to an orthonormal basis and stop
- (2) Pick $\mathbf{v}_i$ as below and $\mathcal{V} = \mathcal{V} \cup \{\mathbf{v}_i\}$

$$\|\mathbf{B}_i \mathbf{v}_i\| = \max_{\mathbf{w}:\|\mathbf{w}\|=1} \|\mathbf{B}_i \mathbf{w}\| = \sigma_i > 0,$$

- (3) Define $\mathbf{u}_i = \frac{1}{\sigma_i} \mathbf{B}_i \mathbf{v}_i$ and set $\mathcal{U} = \mathcal{U} \cup \{\mathbf{u}_i\}$
- (4) Set $\mathbf{B}_{i+1} = \mathbf{B}_i - \sigma_i \cdot \mathbf{u}_i \mathbf{v}_i^\top$, $i = i + 1$, goto (1)

Existence of SVD (cont'd)

Check conditions on SVD (sketch of proof)

- $\mathbf{v}_i \perp \text{span}(\mathcal{V}_{i-1})$ since $\text{range}(\mathbf{B}_i) \perp \text{span}(\mathcal{V}_{i-1})$
- $\mathbf{u}_i \perp \text{span}(\mathcal{U}_{i-1})$ – by symmetry as

$$\mathbf{u}_i = \underset{\mathbf{u}:\|\mathbf{u}\|=1}{\arg\max} \|\mathbf{B}_i^\top \mathbf{u}\|$$

- σ_i ordered: obvious from construction (projection matrices B_i)

From Orthogonal Vectors to Orthogonal Matrices

- Orthogonality between vectors

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0 \quad \Longleftrightarrow \quad \mathbf{u} \perp \mathbf{v}$$

- Matrix $\mathbf{U} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_m] \in \mathbb{R}^{n \times m}$ has **orthonormal columns**

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \delta_{ij} := \begin{cases} 1, & i = j \ \text{(normalized)} \\ 0, & i \neq j \ \text{(orthogonal)} \end{cases}$$

$$\Longrightarrow \text{left invertible} \quad \mathbf{U}^\top \mathbf{U} = \mathbf{I}_m, \quad (\text{not in general: } \mathbf{U}\mathbf{U}^\top = \mathbf{I}_n)$$

- **Orthogonal matrix** = square matrix s.t.

$$\mathbb{R}^{n \times n} \ni \mathbf{U}, \quad \mathbf{U}^\top \mathbf{U} = \mathbf{U}\mathbf{U}^\top = \mathbf{I}_n$$

  - **change of basis matrix between orthonormal bases**

Invariances

- Matrices with orthonormal columns: perserve inner products

$$\langle \mathbf{Au}, \mathbf{Av} \rangle = \mathbf{u}^\top \underbrace{\mathbf{A}^\top \mathbf{A}}_{=\mathbf{I}} \mathbf{v} = \langle \mathbf{u}, \mathbf{w} \rangle$$

- ... and hence also
  - norms, as: $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$
  - distances, as: $\|\mathbf{u} - \mathbf{v}\| = \sqrt{\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\langle \mathbf{u}, \mathbf{v} \rangle}$
  - angles, as: $\cos \angle(\mathbf{u}, \mathbf{v}) = \langle \mathbf{u}, \mathbf{v} \rangle / (\|\mathbf{u}\| \cdot \|\mathbf{v}\|)$

Permutation Matrices

- First example of orthogonal matrices: **permutation matrices**

$$\pi: \text{ permutation}, \quad \mathbf{A} = \mathbf{A}(\pi) \quad \text{s.t.} \quad a_{ij} = \begin{cases} 1 & \text{if } j = \pi_i \\ 0, & \text{otherwise} \end{cases}$$

  - exactly one 1 in each row and in each column
  - implies orthogonality as: $(\mathbf{A}^\top \mathbf{A})_{ij} = \sum_k a_{ki} a_{kj} = \delta_{ij}$
  - example: $\pi = (2, 4, 1, 3)$, $\pi^{-1} = (3, 1, 4, 2)$

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

Permutation Matrices

- example: re-ordering of diagonal elements

$$\underbrace{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}}_{\text{orthogonal}} \begin{bmatrix} \sigma_2 & 0 \\ 0 & \sigma_1 \end{bmatrix} \underbrace{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}}_{\text{orthogonal}} = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$$

Reflector Matrices

- Reflector matrix with direction $\mathbf{u}$ ($\|\mathbf{u}\| = 1$)

$$\mathbf{A} = \mathbf{I} - 2\mathbf{u}\mathbf{u}^\top$$

  - $\mathbf{A} = \mathbf{I} - \mathbf{u}\mathbf{u}^\top$ is projection to hyperplane $(\mathbb{R}\mathbf{u})^\perp$

- Orthogonality

$$\mathbf{A}\mathbf{A}^\top = \mathbf{I} - 4\mathbf{u}\mathbf{u}^\top\mathbf{I} + 4\mathbf{u}\underbrace{\mathbf{u}^\top\mathbf{u}}_{=1}\mathbf{u}^\top = \mathbf{I}$$

Reflector Matrices

- example: sign change in a diagonal matrix

$$\text{diag}(\sigma_1, \ldots, -\sigma_i, \ldots, \sigma_n) \cdot \underbrace{(\mathbf{I} - 2\mathbf{e}_i\mathbf{e}_i^\top)}_{\text{orthogonal}}$$

$$= \text{diag}(\sigma_1, \ldots, \sigma_i, \ldots, \sigma_n)$$

Rotation Matrices

- Rotation in 2d with angle θ

$$\mathbf{A} = \mathbf{A}(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \mathbf{I} - \begin{bmatrix} 1-\cos\theta & \sin\theta \\ -\sin\theta & 1-\cos\theta \end{bmatrix}$$

- Rotation matrix in (k,l) coordinate plane

$$\mathbf{A} = \mathbf{I} - \begin{bmatrix} 1-\cos\theta & \sin\theta \\ -\sin\theta & 1-\cos\theta \end{bmatrix}_{k,l}, \quad \text{where}$$

$$[\cdot]_{k,l} : \text{embeds } 2\times 2 \text{ matrix at } (k,k),(k,l),(l,k),(l,l)$$

- Orthogonality essentially follows from 2D case

$$\mathbf{A}\mathbf{A}^\top = \begin{bmatrix} \cos^2\theta + \sin^2\theta & 0 \\ 0 & \cos^2\theta + \sin^2\theta \end{bmatrix} = \mathbf{I}$$

Multiplication of Orthogonal Matrices

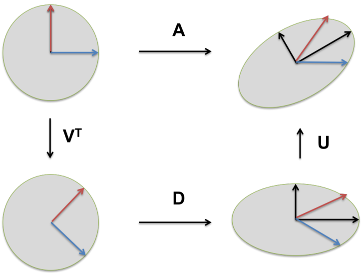
- Product of orthogonal matrices

$$
\begin{aligned}
&(\mathbf{A}_1 \cdot \cdots \cdot \mathbf{A}_r) \cdot (\mathbf{A}_1 \cdot \cdots \cdot \mathbf{A}_r)^\top \\
&= (\mathbf{A}_1 \cdot \cdots \cdot \mathbf{A}_{r-1}) \cdot \underbrace{\mathbf{A}_r \mathbf{A}_r^\top}_{=\mathbf{I}} \cdot (\mathbf{A}_1 \cdot \cdots \cdot \mathbf{A}_{r-1})^\top \overset{\text{induct}}{=} \mathbf{I}
\end{aligned}
$$

- Represent orthogonal matrices via products of $\{$permutation, reflection, planar rotation$\}$ matrices

SVD View of Linear Maps

- Decompose linear map:
  - orthonormal basis transformation in input space
  - perform linear mapping: **diagonal matrix**
  - orthonormal basis transformation in output space



- special case: square matrix $n = m$
- very special case: symmetric matrix, then $\mathbf{U} = \mathbf{V}$ (cf. PCA)

Section 3

SVD and PCA

PCA via SVD (1 of 3)

- Can compute **eigen-decomposition** of $\mathbf{A}\mathbf{A}^\top$ via SVD
  - straightforward calculation

$$\begin{aligned}\mathbf{A}\mathbf{A}^\top &= \left(\mathbf{U}\mathbf{D}\mathbf{V}^\top\right)\left(\mathbf{V}\mathbf{D}^\top\mathbf{U}^\top\right)\\ &= \mathbf{U}\underbrace{\mathbf{D}\cdot\mathbf{I_n}\cdot\mathbf{D}^\top}_{\text{diag}(\lambda_1,\ldots,\lambda_m)}\mathbf{U}^\top = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top\end{aligned}$$

where **eigenvalues** relate to **singular values**

$$\lambda_i = \begin{cases}\sigma_i^2 & \text{for } 1 \le i \le \min\{m,n\}\\ 0 & \text{for } n < i \le m\end{cases}$$

PCA via SVD (2 of 3)

- Similarly $\mathbf{A}^\top \mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}'\mathbf{V}^\top$, where

$$\boldsymbol{\Lambda}' = \text{diag}(\lambda'_1, \ldots, \lambda'_n), \quad \lambda'_i = \begin{cases} \lambda_i & \text{for } 1 \leq i \leq \min\{m, n\} \\ 0 & \text{for } m < i \leq n \end{cases}$$

- Interpretation
  - columns of $\mathbf{U}$: eigenvectors of $\mathbf{A}\mathbf{A}^\top$
  - columns of $\mathbf{V}$: eigenvectors of $\mathbf{A}^\top\mathbf{A}$
  - eigenvalues: $\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda}'$ (identical up to zero padding)
    - $\boldsymbol{\Lambda} = \mathbf{D}\mathbf{D}^\top \in \mathbb{R}^{m \times m}$
    - $\boldsymbol{\Lambda}' = \mathbf{D}^\top\mathbf{D} \in \mathbb{R}^{n \times n}$

PCA via SVD (3 of 3)

- Assume that $\mathbf{X}$ is a **centered data matrix**
- SVD of $\mathbf{X}$ can be used to compute eigendecomposition of $\mathbf{\Sigma}$
  - variance-covariance matrix: $\mathbf{\Sigma} = \frac{1}{n}\mathbf{X}\mathbf{X}^\top$
  - often $n \gg m$: reduced SVD sufficient

Section 4

SVD for Collaborative Filtering

SVD of Rating Matrix: Interpretation

$\mathbf{A}$ = rating matrix, then ...

- k dimensional ($k \leq \text{rank}(\mathbf{A})$) number of latent factors
- $\mathbf{U}$: users-to-factor association matrix
- $\mathbf{V}$: items-to-factor association matrix
- $\mathbf{D}$: level of strength of each factor

SVD For Collaborative Filtering

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^{\top}:$$

$$
\begin{array}{ccccc}
\text{Cremators} & \text{Evil spawn} & \text{Fatal justice} & \text{Clerks} & \text{American pie}
\end{array}
$$

$$
\begin{pmatrix}
5 & 5 & 5 & 0 & 0 \\
4 & 4 & 4 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
3 & 3 & 3 & 0 & 0 \\
0 & 0 & 0 & 4 & 4 \\
0 & 0 & 0 & 5 & 5 \\
0 & 0 & 0 & 4 & 4
\end{pmatrix}
=
\begin{pmatrix}
0.57 & 0 & -0.80 & 0.06 & -0.04 & -0.06 & -0.04 \\
0.46 & 0 & 0.43 & 0.68 & -0.19 & -0.23 & -0.19 \\
0.57 & 0 & 0.37 & -0.70 & -0.08 & -0.11 & -0.08 \\
0.34 & 0 & 0.15 & 0.14 & 0.48 & 0.60 & 0.48 \\
0 & 0.52 & 0 & 0 & -0.71 & 0.35 & 0.28 \\
0 & 0.66 & 0 & 0 & 0.35 & -0.56 & 0.35 \\
0 & 0.52 & 0 & 0 & 0.28 & 0.35 & -0.71
\end{pmatrix}
\times
$$

$$
\begin{pmatrix}
15 & 0 & 0 & 0 & 0 \\
0 & 10.67 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{pmatrix}
\times
\begin{pmatrix}
0.57 & 0.57 & 0.57 & 0 & 0 \\
0 & 0 & 0 & 0.70 & 0.70 \\
0.81 & -0.40 & -0.40 & 0 & 0 \\
0 & 0.70 & -0.70 & 0 & 0 \\
0 & 0 & 0 & 0.70 & -0.70
\end{pmatrix}
$$

SVD For Collaborative Filtering

Factors: **Horror, Comedy**

$\mathbf{U}$: users-to-factors association matrix.

$$
\begin{array}{c}
\begin{array}{ccccc} \text{Cremators} & \text{Evil spawn} & \text{Fatal justice} & \text{Clerks} & \text{American pie} \end{array} \\
\begin{pmatrix}
5 & 5 & 5 & 0 & 0 \\
4 & 4 & 4 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
3 & 3 & 3 & 0 & 0 \\
0 & 0 & 0 & 4 & 4 \\
0 & 0 & 0 & 5 & 5 \\
0 & 0 & 0 & 4 & 4
\end{pmatrix}
\end{array}
=
\begin{array}{c}
\begin{array}{cc} \textbf{Horror} & \textbf{Comedy} \end{array} \\
\begin{pmatrix}
0.57 & 0 \\
0.46 & 0 \\
0.57 & 0 \\
0.34 & 0 \\
0 & 0.52 \\
0 & 0.66 \\
0 & 0.52
\end{pmatrix}
\end{array}
\times
\begin{pmatrix}
15 & 0 \\
0 & 10.67
\end{pmatrix}
\times
\begin{pmatrix}
0.57 & 0.57 & 0.57 & 0 & 0 \\
0 & 0 & 0 & 0.70 & 0.70
\end{pmatrix}
$$

Q: What is the affinity between user1 and horror? 0.57

SVD For Collaborative Filtering

Factors: Horror, Comedy

$\mathbf{D}$: weight of different factors in the data.

Strength of Horror

Strength of Comedy

$$\begin{array}{c} \begin{array}{ccccc} \text{Cremators} & \text{Evil spawn} & \text{Fatal justice} & \text{Clerks} & \text{American pie} \end{array} \\ \begin{pmatrix} 5 & 5 & 5 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 4 & 4 \end{pmatrix} \end{array} = \begin{pmatrix} 0.57 & 0 \\ 0.46 & 0 \\ 0.57 & 0 \\ 0.34 & 0 \\ 0 & 0.52 \\ 0 & 0.66 \\ 0 & 0.52 \end{pmatrix} \times \begin{pmatrix} 15 & 0 \\ 0 & 10.67 \end{pmatrix} \times \begin{pmatrix} 0.57 & 0.57 & 0.57 & 0 & 0 \\ 0 & 0 & 0 & 0.70 & 0.70 \end{pmatrix}$$

Q: What is the expression of the comedy concept in the data? 10.67

SVD For Collaborative Filtering

Factors: Horror, Comedy

$\mathbf{V}$: Movies-to-factor association matrix.

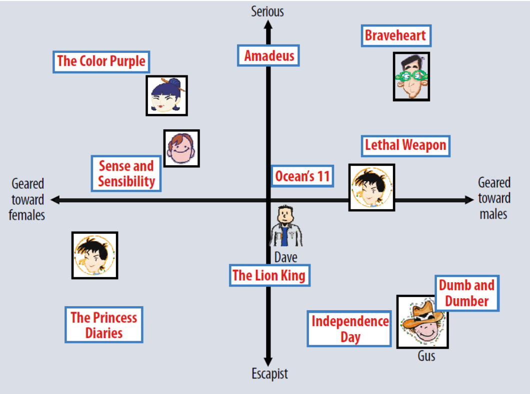
$$
\begin{array}{c}
\begin{array}{ccccc} \text{Cremators} & \text{Evil spawn} & \text{Fatal justice} & \text{Clerks} & \text{American pie} \end{array} \\
\begin{pmatrix}
5 & 5 & 5 & 0 & 0 \\
4 & 4 & 4 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
3 & 3 & 3 & 0 & 0 \\
0 & 0 & 0 & 4 & 4 \\
0 & 0 & 0 & 5 & 5 \\
0 & 0 & 0 & 4 & 4
\end{pmatrix}
\end{array}
=
\begin{pmatrix}
0.57 & 0 \\
0.46 & 0 \\
0.57 & 0 \\
0.34 & 0 \\
0 & 0.52 \\
0 & 0.66 \\
0 & 0.52
\end{pmatrix}
\times
\begin{pmatrix}
15 & 0 \\
0 & 10.67
\end{pmatrix}
\times
\begin{pmatrix}
0.57 & 0.57 & 0.57 & 0 & 0 \\
0 & 0 & 0 & 0.70 & 0.70
\end{pmatrix}
$$

Q: What is the similarity between Clerks and Horror? 0

What is the similarity between Clerks and Comedy? 0.7

Collaborative Filtering Example II

Characterization of the users and movies using two axes - male vs. female and serious vs. escapist.



* Ref: "Matrix factorization techniques for recommender systems"

`http://www2.research.att.com/~volinsky/papers/ieeecomputer.pdf`.

Section 5

Matrix Approximation via SVD

Frobenius Norm

- Definition: **Frobenis norm**

$$\|\mathbf{A}\|_F := \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N} a_{ij}^2} = \sqrt{\text{trace}(\mathbf{A}^\top \mathbf{A})}$$

  - only depends on singular values of $\mathbf{A}$

$$\|\mathbf{A}\|_F^2 = \sum_{i=1}^{k} \sigma_i^2, \quad k = \min\{m, n\}$$

  - follows from cyclic rule: $\text{trace}(\mathbf{XYZ}) = \text{trace}(\mathbf{ZXY})$

$$\begin{aligned}\text{trace}(\mathbf{A}^\top \mathbf{A}) &= \text{trace}(\mathbf{VD}^\top \mathbf{DV}^\top) = \text{trace}(\mathbf{D}^\top \mathbf{D}) \\ &= \text{trace}(\text{diag}(\sigma_1^2 \ldots, \sigma_k^2)) = \sum_{i=1}^{k} \sigma_i^2\end{aligned}$$

Singular Values and Matrix Norms

- **Induced p-norms**

$$\|\mathbf{A}\|_p := \sup\{\|\mathbf{A}\mathbf{x}\|_p : \|\mathbf{x}\|_p = 1\}$$

  - matrix 2-norm $=$ largest singular value

$$\|\mathbf{A}\|_2 = \sup\{\|\mathbf{A}\mathbf{x}\|_2 : \|\mathbf{x}\|_2 = 1\} = \sigma_1$$

  - proof: for $\mathbf{x}$ with $\|\mathbf{x}\|_2 = 1$

$$\|\mathbf{A}\mathbf{x}\|_2 = \|\mathbf{U}\mathbf{D}\mathbf{V}^\top\mathbf{x}\|_2 = \|\mathbf{D}\mathbf{y}\|_2 = \sqrt{\sum_{i=1}^{k} \sigma_i^2 y_i^2}$$

    - here $\mathbf{y} := \mathbf{V}^\top\mathbf{x}$, hence $\|\mathbf{y}\|_2 = \|\mathbf{x}\|_2$
    - maximized for $\mathbf{y} = (1, 0, \ldots, 0)^\top$, maximum σ_1.

Eckart–Young Theorem

- **Reduced rank SVD**:
  optimal low rank approximation in Frobenius norm
  - SVD of $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$, define for $k \leq \text{rank}(\mathbf{A})$
  $$\mathbf{A}_k := \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top, \quad \text{rank}(\mathbf{A}_k) = k$$
  - Then
  $$\min_{\text{rank}(\mathbf{B})=K} \|\mathbf{A} - \mathbf{B}\|_F^2 = \|\mathbf{A} - \mathbf{A}_\mathbf{K}\|_F^2 = \sum_{r=k+1}^{\text{rank}(\mathbf{A})} \sigma_r^2$$

Euclidean Norm Approximation

- $\mathbf{A}_K$: optimal approximation in the sense of the matrix 2-norm

$$\min_{\text{rank}(\mathbf{B})=k} \|\mathbf{A}-\mathbf{B}\|_2 = \|\mathbf{A}-\mathbf{A_k}\|_2 = \sigma_{k+1}$$

Section 6

SVD for Image Compression

SVD for Image Compression

- SVD can be applied to images
  - represented as a matrix of pixels) = **compression**
  - matrix = one image (not set of observation vectors)
- SVD: eliminate small singular values, compute reduced rank image
  - optimal approximation guarantees
  - least squares of individual pixel value differences: relevant error metric on images

SVD for Image Compression

* http://en.wikipedia.org/wiki/Lenna.

- 512×512 pixel image
- matrix representation $\mathbf{I}$ of the image is of full-rank
- $\implies \text{rank}(\mathbf{I}) = 512.$

SVD For Image Compression

Rank$(\mathbf{I}) = 512$

Rank$(\mathbf{I}) = 256$

SVD For Image Compression

$\text{Rank}(\mathbf{I}) = 128$

$\text{Rank}(\mathbf{I}) = 64$

SVD For Image Compression

$\text{Rank}(\mathbf{I}) = 32$

$\text{Rank}(\mathbf{I}) = 8$

SVD For Image Compression

Why Image Compression?

- Reduce the overhead for network transmission
- Reduce the size for efficient storage

Q: For a given image of size $m \times n$ pixels, what is the compression ratio? assuming the compressed image is of rank r.