Dimensionality Reduction: SVD & CUR

CS345a: Data Mining
Jure Leskovec and Anand Rajaraman
Stanford University
Announcements

Homework 2 is out:
- Due Monday 15\textsuperscript{th} at midnight!
- Submit PDFs

Talk:
- [http://rain.stanford.edu](http://rain.stanford.edu)
- Wed at 12:30 in Terman 453
- Yehuda Koren – Winner of the Netflix challenge!
## Text - LSI: find ‘concepts’

<table>
<thead>
<tr>
<th>document</th>
<th>term</th>
<th>data</th>
<th>information</th>
<th>retrieval</th>
<th>brain</th>
<th>lung</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS-TR1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CS-TR2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CS-TR3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CS-TR4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MED-TR1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>MED-TR2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>MED-TR3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
- **Compress / reduce dimensionality**
  - $10^6$ rows; $10^3$ columns; no updates
  - random access to any cell(s); small error: OK

<table>
<thead>
<tr>
<th>customer</th>
<th>7/10/06</th>
<th>7/11/06</th>
<th>7/12/06</th>
<th>7/13/06</th>
<th>7/14/06</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC Inc.</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DEF Ltd.</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GHI Inc.</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>KLM Co.</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Smith</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Johnson</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Thompson</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
SVD - Motivation

day 2

day 1
SVD - Motivation
A_{[n \times m]} = U_{[n \times r]} \sum_{[r \times r]} (V_{[m \times r]})^T

- **A**: \(n \times m\) matrix  
  (eg., \(n\) documents, \(m\) terms)
- **U**: \(n \times r\) matrix  
  (\(n\) documents, \(r\) concepts)
- **\(\Sigma\)**: \(r \times r\) diagonal matrix  
  (strength of each ‘concept’)
  (\(r\) : rank of the matrix)
- **V**: \(m \times r\) matrix  
  (\(m\) terms, \(r\) concepts)
$A \approx U \Sigma V^T = \sum_i \sigma_i u_i \circ v_i$
\[ A \approx U\Sigma V^T = \sum_i \sigma_i u_i \circ v_i \]
THEOREM [Press+92]: always possible to decompose matrix $A$ into $A = U \Sigma V^T$, where

- $U$, $\Sigma$, $V$: unique
- $U$, $V$: column orthonormal:
  - $U^T U = I$; $V^T V = I$ ($I$: identity matrix)
  - (Cols. are orthogonal unit vectors)
- $\Sigma$: diagonal
  - Entries (singular values) are positive, and sorted in decreasing order
SVD - Example

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

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1 \\
\end{bmatrix}
= \begin{bmatrix}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\
0.90 & 0 \\
0 & 0.53 \\
0 & 0.80 \\
0 & 0.27 \\
\end{bmatrix}
\times \begin{bmatrix}
9.64 & 0 \\
0 & 5.29 \\
\end{bmatrix}
\times \begin{bmatrix}
0.58 & 0.58 & 0.58 & 0 & 0 \\
0 & 0 & 0 & 0.71 & 0.71 \\
\end{bmatrix}
\]
\[ A = U \Sigma V^T \] - example:

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1 
\end{bmatrix}
\begin{bmatrix}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\
0.90 & 0 \\
0 & 0.53 \\
0 & 0.80 \\
0 & 0.27 
\end{bmatrix}
\begin{bmatrix}
9.64 & 0 \\
0 & 5.29 \\
0.58 & 0.58 \\
0.58 & 0.58 \\
0.71 & 0.71 
\end{bmatrix}
\]
\[ A = U \Sigma V^T \] - example:

```
<table>
<thead>
<tr>
<th></th>
<th>retrieval</th>
<th>lung</th>
<th>brain inf.</th>
<th>data</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS</td>
<td>1 1 1 0 0</td>
<td>5 5 5 0 0</td>
<td>0 0 0 2 2</td>
<td>0 0 0 3 3</td>
</tr>
<tr>
<td>MD</td>
<td>2 2 2 0 0</td>
<td>1 1 1 0 0</td>
<td>0 0 0 2 2</td>
<td>0 0 0 3 3</td>
</tr>
</tbody>
</table>
```

```
\begin{bmatrix}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\
0.90 & 0 \\
0 & 0.53 \\
0 & 0.80 \\
0 & 0.27 \\
\end{bmatrix}
\times
\begin{bmatrix}
9.64 & 0 \\
0 & 5.29 \\
\end{bmatrix}
\times
\begin{bmatrix}
0.58 & 0.58 & 0.58 & 0 & 0 \\
0 & 0 & 0 & 0.71 & 0.71 \\
\end{bmatrix}
```
SVD - Example

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

<table>
<thead>
<tr>
<th>Retrieval</th>
<th>Inf.</th>
<th>Brain</th>
<th>Lung</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1 1 0 0</td>
<td>2 2 2 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 5 5 0 0</td>
<td>0 0 0 2 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 0 0 3 3</td>
<td>0 0 0 1 1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\
0.90 & 0 \\
0 & 0.53 \\
0 & 0.80 \\
0 & 0.27
\end{bmatrix}
\begin{bmatrix}
9.64 & 0 \\
0 & 5.29
\end{bmatrix}
\begin{bmatrix}
0.58 & 0.58 & 0.58 & 0 & 0 \\
0 & 0 & 0 & 0.71 & 0.71
\end{bmatrix}
\]

'Strength' of CS-concept
SVD - Example

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

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\
0.90 & 0 \\
0 & 0.53 \\
0 & 0.80 \\
0 & 0.27
\end{bmatrix}
\begin{bmatrix}
9.64 & 0 \\
0 & 5.29 \\
\end{bmatrix}
\begin{bmatrix}
0.58 & 0.58 & 0.58 & 0 & 0 \\
0 & 0 & 0 & 0.71 & 0.71
\end{bmatrix}
\]
**SVD - Example**

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

  - Data retrieval
  - Inf. brain lung

  \[
  \begin{bmatrix}
  1 & 1 & 1 & 0 & 0 \\
  2 & 2 & 2 & 0 & 0 \\
  1 & 1 & 1 & 0 & 0 \\
  5 & 5 & 5 & 0 & 0 \\
  0 & 0 & 0 & 2 & 2 \\
  0 & 0 & 0 & 3 & 3 \\
  0 & 0 & 0 & 1 & 1 \\
  \end{bmatrix}
  \begin{bmatrix}
  0.18 & 0 \\
  0.36 & 0 \\
  0.18 & 0 \\
  0.90 & 0 \\
  0 & 0.53 \\
  0 & 0.80 \\
  0 & 0.27 \\
  \end{bmatrix}
  \begin{bmatrix}
  9.64 & 0 \\
  0 & 5.29 \\
  \end{bmatrix}
  \begin{bmatrix}
  0.58 & 0.58 & 0.58 & 0 & 0 \\
  0 & 0 & 0 & 0.71 & 0.71 \\
  \end{bmatrix}
  \]

  - Term-to-concept similarity matrix
SVD - Interpretation #1

‘documents’, ‘terms’ and ‘concepts’:
- $U$: document-to-concept similarity matrix
- $V$: term-to-concept sim. matrix
- $\Sigma$: its diagonal elements:
  ‘strength’ of each concept
SVD: gives best axis to project

- best axis to project on: ('best' = min sum of squares of projection errors)
- minimum RMS error
SVD - Interpretation #2

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

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1
\end{bmatrix}
= \begin{bmatrix}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\
0.90 & 0 \\
0 & 0.53 \\
0 & 0.80 \\
0 & 0.27
\end{bmatrix}
\times \begin{bmatrix}
9.64 & 0 \\
0 & 5.29
\end{bmatrix}
\times
\begin{bmatrix}
0.58 & 0.58 & 0.58 & 0 & 0 \\
0 & 0 & 0 & 0.71 & 0.71
\end{bmatrix}
\]
\[ A = U \Sigma V^T \] - example:

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1 \\
\end{bmatrix}
\begin{bmatrix}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\
0.90 & 0 \\
0 & 0.53 \\
0 & 0.80 \\
0 & 0.27 \\
\end{bmatrix}
\begin{bmatrix}
9.64 & 0 \\
0 & 5.29 \\
\end{bmatrix}
\begin{bmatrix}
0.58 & 0.58 & 0.58 & 0 & 0 \\
0 & 0 & 0 & 0.71 & 0.71 \\
\end{bmatrix}
\]

variance ('spread') on the \( v_1 \) axis
**SVD - Interpretation #2**

- \( A = U \Sigma V^T \) - example:
  - \( U\Sigma \): gives the coordinates of the points in the projection axis

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1 \\
\end{bmatrix}
= \begin{bmatrix}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\
0.90 & 0 \\
& 0.53 \\
& 0.80 \\
& 0.27 \\
\end{bmatrix}
\times \begin{bmatrix}
9.64 & 0 \\
0 & 5.29 \\
\end{bmatrix}
\times \begin{bmatrix}
0.58 & 0.58 & 0.58 & 0 & 0 \\
0 & 0 & 0 & 0.71 & 0.71 \\
\end{bmatrix}
\]
More details

- **Q:** how exactly is dim. reduction done?

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1 \\
\end{bmatrix}
= \begin{bmatrix}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\
0.90 & 0 \\
0 & 0.53 \\
0 & 0.80 \\
0 & 0.27 \\
\end{bmatrix}
\times \begin{bmatrix}
9.64 & 0 \\
0 & 5.29 \\
\end{bmatrix}
\times \begin{bmatrix}
0.58 & 0.58 & 0.58 & 0 & 0 \\
0 & 0 & 0 & 0.71 & 0.71 \\
\end{bmatrix}
\]
More details

- **Q:** how exactly is dim. reduction done?
- **A:** set the smallest singular values to zero:

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1 \\
\end{bmatrix}
= \begin{bmatrix}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\
0.90 & 0 \\
0 & 0.53 \\
0 & 0.80 \\
0 & 0.27 \\
\end{bmatrix}
\times \begin{bmatrix}
9.64 & 0 \\
0 & 5.29 \\
\end{bmatrix}
\times \begin{bmatrix}
0.58 & 0.58 & 0.58 & 0 & 0 \\
0 & 0 & 0 & 0.71 & 0.71 \\
\end{bmatrix}
\]
SVD - Interpretation #2

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1
\end{bmatrix}
\approx
\begin{bmatrix}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\
0.90 & 0 \\
0 & 0.53 \\
0 & 0.80 \\
0 & 0.27
\end{bmatrix}
\times
\begin{bmatrix}
9.64 & 0 \\
0 & 0 \\
0.58 & 0.58 & 0.58 & 0 & 0 \\
0 & 0 & 0 & 0.71 & 0.71
\end{bmatrix}
\]
SVD - Interpretation #2

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1 \\
\end{bmatrix}
\sim
\begin{bmatrix}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\
0.90 & 0 \\
0 & 0.53 \\
0 & 0.80 \\
0 & 0.27 \\
\end{bmatrix}
\times
\begin{bmatrix}
9.64 & 0 \\
0 & 0 \\
\end{bmatrix}
\times
\begin{bmatrix}
0.58 & 0.58 & 0.58 & 0 & 0 \\
0 & 0 & 0 & 0.71 & 0.71 \\
\end{bmatrix}
\]
SVD - Interpretation #2

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1 \\
\end{bmatrix}
\sim
\begin{bmatrix}
0.18 \\
0.36 \\
0.18 \\
0.90 \\
0 \\
0 \\
0 \\
\end{bmatrix}
\times
\begin{bmatrix}
9.64 \\
0.58 \\
0.58 \\
0.58 \\
0 \\
0 \\
0 \\
\end{bmatrix}
\times
\]
SVD - Interpretation #2

\[
\begin{bmatrix}
1 & 1 & 0 & 0 \\
2 & 2 & 0 & 0 \\
1 & 1 & 0 & 0 \\
5 & 5 & 0 & 0 \\
0 & 0 & 2 & 2 \\
0 & 0 & 3 & 3 \\
0 & 0 & 1 & 1 \\
\end{bmatrix}
\sim
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
\]
Equivalent:
’spectral decomposition’ of the matrix:

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1 \\
\end{bmatrix} = \begin{bmatrix}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\
0.90 & 0 \\
0 & 0.53 \\
0 & 0.80 \\
0 & 0.27 \\
\end{bmatrix} \times \begin{bmatrix}
9.64 & 0 \\
0 & 5.29 \\
0.58 & 0.58 & 0.58 & 0 & 0 \\
0 & 0 & 0 & 0.71 & 0.71 \\
\end{bmatrix}
\]
SVD - Interpretation #2

Equivalent:
‘spectral decomposition’ of the matrix:

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1 \\
\end{bmatrix}
= \begin{bmatrix}
u_1 & u_2 \\
\end{bmatrix}
\times \begin{bmatrix}
\lambda_1 & \emptyset \\
\emptyset & \lambda_2 \\
\end{bmatrix}
\times \begin{bmatrix}
v_1 \\
v_2 \\
\end{bmatrix}
\]
Equivalent: ‘spectral decomposition’ of the matrix:

\[
\begin{pmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1 \\
\end{pmatrix} = \lambda_1 \ u_1 \ v_1^T + \lambda_2 \ u_2 \ v_2^T + \ldots
\]
Equivalent:
‘spectral decomposition’ of the matrix:

\[
\begin{pmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1 \\
\end{pmatrix}
= \sum_{i=1}^{r} \lambda_i \mathbf{u}_i \mathbf{v}_i^T + \ldots
\]
Approximation / dim. reduction:
by keeping the first few terms (Q: how many?)

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1 \\
\end{bmatrix}
= \lambda_1 \quad u_1 \quad v_1^T + \lambda_2 \quad u_2 \quad v_2^T + \ldots
\]

assume: \( \lambda_1 \geq \lambda_2 \geq \ldots \)
Answer: (heuristic-[Fukunaga]):
keep 80-90% of ‘energy’ \((= \sum \lambda_i^2)\)

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1 \\
\end{bmatrix}
\]

\[
= \lambda_1 \ u_1 \ v_1^T + \lambda_2 \ u_2 \ v_2^T + ...
\]

assume: \(\lambda_1 \geq \lambda_2 \geq ...
\]
SVD - Complexity

- $O(nm^2)$ or $O(n^2m)$ (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:
  - Linear algebra packages like: LINPACK, Matlab, SPlus, Mathematica ...
SVD - conclusions so far

- **SVD**: \( A = U \Sigma V^T \): unique (*)
- **U**: document-to-concept similarities
- **V**: term-to-concept similarities
- **\( \Sigma \)**: strength of each concept

- **Dim. reduction:**
  - keep the few largest singular values (80-90% of ‘energy’)
  - SVD: picks up linear correlations
Relation to Eigen-decomposition

- **SVD gives us:**
  - \( A = U \Sigma V^T \)

- **Eigen-decomposition:**
  - \( A = X \Lambda X^T \)
    - \( U, V, X \) are orthonormal (\( U^TU = I \)),
    - \( \Lambda, \Sigma \) are diagonal

- **What is:**
  - \( AA^T = \)
  - \( A^TA = \)
SVD: Properties

- \( AA^T = U \Sigma^2 U^T \)
- \( A^T A = V \Sigma^2 V^T \)

- \((A^T A)^k = V \Sigma^{2k} V^T \)

- \((A^T A)^k \sim v_1 \lambda_1^{2k} v_1^T \) for \( k >> 1 \)

- \((A^T A)^k x \sim \) (constant) \( v_1 \) for (almost) any \( x \)
Q: How to do queries with LSI?
Problem: e.g., find documents with ‘data’
Case study - LSI

Q: How to do queries with LSI?
A: map query vectors into ‘concept space’ – how?

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\
0.90 & 0 \\
0 & 0.53 \\
0 & 0.80 \\
0 & 0.27
\end{bmatrix}
\begin{bmatrix}
9.64 & 0 \\
0 & 5.29
\end{bmatrix}
\begin{bmatrix}
0.58 & 0.58 & 0.58 & 0 & 0 \\
0 & 0 & 0 & 0.71 & 0.71
\end{bmatrix}
\]
Q: How to do queries with LSI?
A: map query vectors into ‘concept space’ – how?

\[
q = \begin{bmatrix}
1 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

Data retrieval brain lung

\[
\text{term}_2
\]

\[
\text{term}_1
\]
Q: How to do queries with LSI?
A: map query vectors into ‘concept space’ – how?

q = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}

A: inner product (cosine similarity) with each ‘concept’ vector \( v_i \)

\[ q = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \text{inner product} \]
**Case study - LSI**

Q: How to do queries with LSI?

A: map query vectors into ‘concept space’ – how?

$$q = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

A: inner product (cosine similarity) with each ‘concept’ vector $v_i$
Case study - LSI

Compactly, we have:
\[ q_{\text{concept}} = q \, V \]

E.g.:

\[ q = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix} \]

\[ \begin{bmatrix} 0.58 & 0 \\ 0.58 & 0 \\ 0.58 & 0 \\ 0 & 0.71 \\ 0 & 0.71 \end{bmatrix} \]

\[ = \begin{bmatrix} 0.58 & 0 \end{bmatrix} \]

term-to-concept similarities

CS-concept
Case study - LSI

How would the document (‘information’, ‘retrieval’) be handled by LSI?

\[ d_{\text{concept}} = d \cdot V \]

E.g.:

\[
\begin{bmatrix}
\text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung}
\end{bmatrix}
\begin{bmatrix}
0.58 & 0 & 0.58 & 0 & 0.58 \\
0.58 & 0 & 0.58 & 0 & 0.58 \\
0 & 0.71 & 0 & 0.71 & 0.71 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

\[ = \begin{bmatrix} 1.16 & 0 \end{bmatrix} \]

CS-concept
Observation: document (‘information’, ‘retrieval’) will be retrieved by query (‘data’), although it does not contain ‘data’!

\[
\begin{bmatrix}
data & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \\
0 & 1 & 1 & 0 & 0
\end{bmatrix}
\rightarrow
\begin{bmatrix}
\text{CS-concept} \\
1.16 & 0
\end{bmatrix}
\]

\[
\begin{bmatrix}
data \\
1 & 0 & 0 & 0 & 0
\end{bmatrix}
\rightarrow
\begin{bmatrix}
0.58 & 0
\end{bmatrix}
\]
SVD: Drawbacks

+ Optimal low-rank approximation:
  • in L2 norm
- Interpretability problem:
  ▪ A singular vector specifies a linear combination of all input columns or rows.
- Lack of Sparsity:
  ▪ Singular vectors are dense
CUR Decomposition

- **Goal:**
  Make $\|A-CUR\|$ small

\[
\begin{pmatrix}
\text{Blue}
\end{pmatrix}
\approx
\begin{pmatrix}
\text{Red}
\end{pmatrix}
\cdot
\begin{pmatrix}
U
\end{pmatrix}
\cdot
\begin{pmatrix}
R
\end{pmatrix}
\]
Goal:
Make $||A\text{-CUR}||$ small

$A \approx C \cdot U \cdot R$

Pseudo-inverse of the intersection of $C$ and $R$
Let:
\[ A_k \] be the “best” rank k approximation to A (e.i., SVD)

**Theorem** [Drineas et al.]
CUR in O(mn) time achieves
- \[ \| A - CUR \| \leq \| A - A_k \| + \varepsilon \| A \| \]
with probability at least 1-\( \delta \), by picking
- \( O(k \log(1/\delta)/\varepsilon^2) \) columns, and
- \( O(k^2 \log^3(1/\delta)/\varepsilon^6) \) rows
Sample columns (similarly for rows):

Input: matrix $A \in \mathbb{R}^{m \times n}$, sample size $c$

Output: $C_d \in \mathbb{R}^{m \times c}$

1. for $x = 1 : n$ [column distribution]
2. $P(x) = \sum_i A(i, x)^2 / \sum_{i,j} A(i, j)^2$
3. for $i = 1 : c$ [sample columns]
4. Pick $j \in 1 : n$ based on distribution $P(x)$
5. Compute $C_d(:, i) = A(:, j)/\sqrt{cP(j)}$
Let \( W \) be the “intersection” of sampled columns \( C \) and rows \( R \).

Then: \[
U = W^+ = X \Sigma^+ Y^T
\]

- \( \Sigma^+ \): reciprocals of non-zero singular values: \( \Sigma^+_{ii} = 1 / \Sigma_{ii} \)

i.e., Moore–Penrose pseudoinverse

\[
U = W^+
\]
CUR: Pros & Cons

+ Easy interpretation
  - Since the basis vectors are actual columns and rows

+ Sparse basis
  - Since the basis vectors are actual columns and rows

- Duplicate columns and rows
  - Columns of large norms will be sampled many times
If we want to get rid of the duplicates:

- Throw them away
- Scale the columns/rows by the square root of the number of duplicates
SVD vs. CUR

SVD: \( A = U \Sigma V^T \)

- Huge but sparse
- Big and dense

CUR: \( A = CUR \)

- Dense but small
- Huge but sparse
- Big but sparse
References: CUR


- J. Sun, Y. Xie, H. Zhang, C. Faloutsos: Less is More: Compact Matrix Decomposition for Large Sparse Graphs, SDM 2007


Acknowledgements

- Slides borrowed from Jimeng Sun, Christos Faloutsos, Michael Mahoney, Hanghang Tong, and Petros Drineas