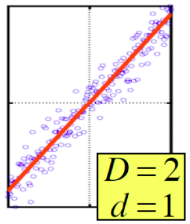
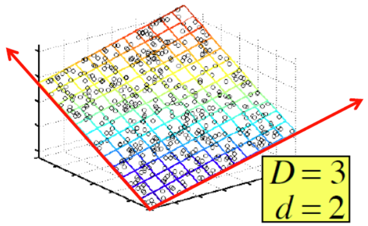
大 数 据 分 析

Scalable Machine Learning
Dimension Reduction

刘盛华

降维





- **Assumption: 数据分布于低维度子空间**
- **该子空间的基可以有效表征数据**

J. Leskovec, A. Rajaraman, J. Ullman: Mining of Massive Datasets, http://www.mmds.org

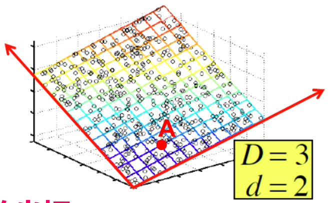
矩阵的秩

- **Q: 矩阵 A的秩(rank)是什么?**
- **A: A中线性无关的列的数量**
- **For example:**
  - A = $\begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix}$ 秩 r=2
    - 原因：1）A的前两行线性无关，所以秩至少为2 2）A第三行与前两行线性相关（row3=row1-row2），所以秩小于3
- **为什么需要关心矩阵的低秩?**
  - 矩阵A可以先写成两个“基”向量: [1 2 1] [-2 -3 1]
  - 然后通过二维坐标表征: [1 0] [0 1] [1 1]

秩就是“维度”

- **3D空间中的点云:**
  - 可以用矩阵表征点的位置:

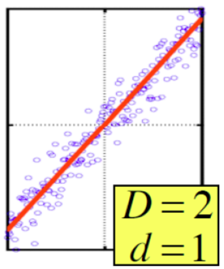
每行表示一个点: $\begin{array}{c} \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix} \end{array} \begin{array}{l} A \\ B \\ C \end{array}$



- **可以更简洁地重写每个点的坐标**
  - 旧的基向量: [1 0 0] [0 1 0] [0 0 1]
  - 新的基向量: [1 2 1] [-2 -3 1]
  - 三个点都有新的坐标 A: [1 0]. B: [0 1], C: [1 -1]
    - 注：我们已经减少了坐标的维度

降维

- **降维的目的是寻找数据的轴(axis)**



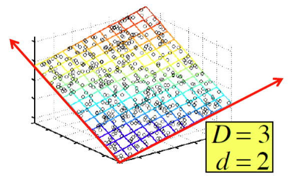
我们可以根据数据点在红线上的分布，使用一维坐标替代原始的二维坐标。

该方法会引入一定的偏差，因为数据并没有严格分布在红线上。

为什么降维？

为什么降维?

- **发现隐含的关联、主题**
  - **通常共现的词**
- **移除冗余的、有噪声的特征**
  - **并不是所有的词都有用**
- **数据的解释、可视化**
- **方便数据的存储和处理**



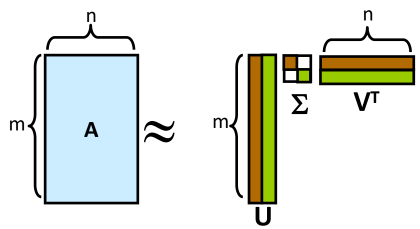
SVD - Definition

$$A_{[m \times n]} = U_{[m \times r]} \Sigma_{[r \times r]} (V_{[n \times r]})^T$$

- **A: 输入矩阵**
  - ***m* x *n* matrix (e.g., *m* 篇文档, *n* 个词项 terms)**
- **U: A的左奇异向量构成的矩阵**
  - ***m* x *r* matrix (*m* 篇文档, *r* 个概念concepts)**
- **Σ: 奇异值矩阵**
  - ***r* x *r* 对角阵 (每种主题的“强度” strength)**
    (*r* : 矩阵A的秩)
- **V: A的右奇异向量构成的矩阵**
  - ***n* x *r* matrix (*n* 个词项 terms, *r* 个概念concepts)**

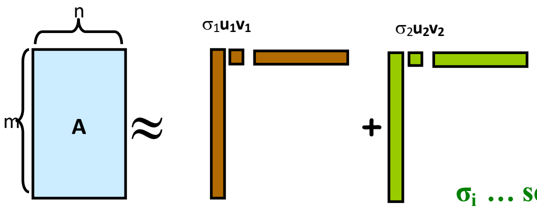
SVD

$$\mathbf{A} \approx \mathbf{U}\Sigma\mathbf{V}^T = \sum_i \sigma_i \mathbf{u}_i \circ \mathbf{v}_i^T$$



SVD

$$\mathbf{A} \approx \mathbf{U}\Sigma\mathbf{V}^T = \sum_i \sigma_i \mathbf{u}_i \circ \mathbf{v}_i^{\mathrm{T}}$$



σ_i … scalar
$\mathbf{u}_i$ … vector
$\mathbf{v}_i$ … vector

SVD - Properties

我们总能对一个实值矩阵A分解： $A = U\Sigma V^T$
分解时需要满足：

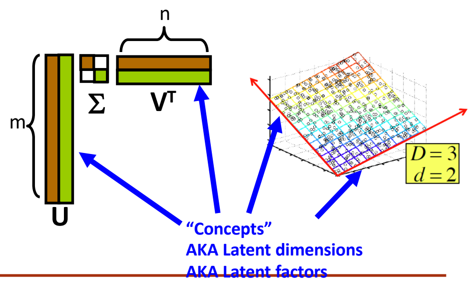
- U, Σ, V: 唯一
- U, V: 每个矩阵的列向量彼此正交
  - $U^T U = I$; $V^T V = I$ (I: 单位阵)
  - 每个矩阵的列都是彼此正交的单位向量
- Σ: 对角矩阵
  - 每个元素（奇异值）是正的，并且按照值大小降序排列($\sigma_1 \geq \sigma_2 \geq \ldots \geq 0$)

Nice proof of uniqueness: http://www.mpi-inf.mpg.de/~bast/ir-seminar-ws04/lecture2.pdf

SVD – Example: Users-to-Movies

- $A = U\,\Sigma\,V^T$ - example: 用户—电影评分矩阵

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



SVD – Example: Users-to-Movies

- $A = U\,\Sigma\,V^T$ - example: 用户—电影评分矩阵

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

SVD – Example: Users-to-Movies

- $A = U \Sigma V^T$ - example: 用户—电影评分矩阵

SciFi-concept

Romance-concept

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

SVD – Example: Users-to-Movies

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

U:用户—主题相似度矩阵

SciFi-concept

Romance-concept

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

SVD – Example: Users-to-Movies

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

SciFi-concept

"strength" of the SciFi-concept

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

SVD – Example: Users-to-Movies

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

V：电影—主题相似度矩阵

SciFi-concept

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

SciFi-concept

SVD – 解释 #1

'**电影**', '**用户**' and '**主题**':

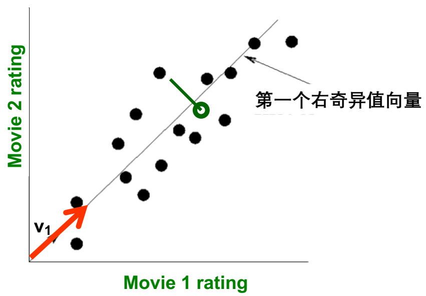
'**movies**', '**users**' and '**concepts**':

- *U*: 用户—主题相似度矩阵
- *V*: 电影—主题相似度矩阵
- Σ: 对角线的元素代表了每个主题的强度

17

使用SVD降维

SVD – 降维



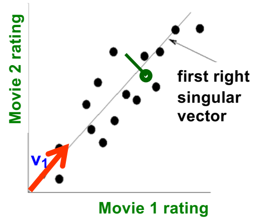
- 只使用一维坐标 (z) 替代原始的二维坐标(x, y)
- 点的坐标是其在向量v_1上的位置
- 如何选择 v_1? **最小化重构误差**

19

SVD降维

- **Goal: 最小化总重构误差**:

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

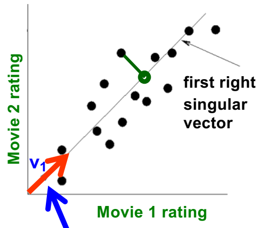


  - x_{ij} 是旧坐标 z_{ij} 是新坐标
- **SVD给出最适合原始数据投影的轴:**
  - 最适合：最小化重构误差
- 换言之，SVD给出最小的重构误差

20

SVD – 解释 #2

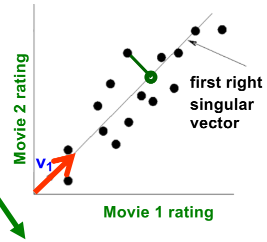
- $A = U \Sigma V^T$ - example:
  - V: “电影—主题” 矩阵
  - U: “用户—主题” 矩阵



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

SVD - Interpretation #2

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



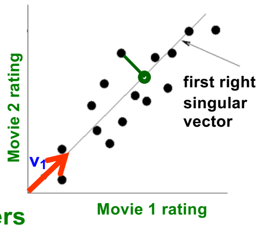
variance (‘spread’) on the v_1 axis

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

SVD - Interpretation #2

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

- U Σ: 给出投影轴以及数据点的投影值（坐标）



Projection of users on the “Sci-Fi” axis (U Σ) :

$$\begin{bmatrix} 1&1&1&0&0\\ 3&3&3&0&0\\ 4&4&4&0&0\\ 5&5&5&0&0\\ 0&2&0&4&4\\ 0&0&0&5&5\\ 0&1&0&2&2 \end{bmatrix} \quad \begin{bmatrix} 1.61&0.19&-0.01\\ 5.08&0.66&-0.03\\ 6.82&0.85&-0.05\\ 8.43&1.04&-0.06\\ 1.86&-5.60&0.84\\ 0.86&-6.93&-0.87\\ 0.86&-2.75&0.41 \end{bmatrix}$$

SVD - Interpretation #2

More details

- Q: SVD降维有多精确？

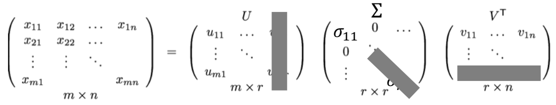
$$\begin{bmatrix} 1&1&1&0&0\\ 3&3&3&0&0\\ 4&4&4&0&0\\ 5&5&5&0&0\\ 0&2&0&4&4\\ 0&0&0&5&5\\ 0&1&0&2&2 \end{bmatrix} = \begin{bmatrix} \mathbf{0.13}&0.02&-0.01\\ \mathbf{0.41}&0.07&-0.03\\ \mathbf{0.55}&0.09&-0.04\\ \mathbf{0.68}&0.11&-0.05\\ 0.15&\mathbf{-0.59}&\mathbf{0.65}\\ 0.07&\mathbf{-0.73}&\mathbf{-0.67}\\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

SVD – 最优低秩近似.

- **Theorem:**
  Let $A = U \Sigma V^T$ and $B = U S V^T$ where
  S = diagonal $r \times r$ matrix with $s_i=\sigma_i$ ($i=1\ldots k$) else $s_i=0$
  then B is a best rank(B)=k approx. to A

What do we mean by “best”:

- B is a solution to $\min_B \|A-B\|_F$ where rank(B)=k



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

SVD - Interpretation #2

More details

- Q: SVD降维有多精确？
- A: 把最小的奇异值(一个或多个)设置为0.

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

SVD - Interpretation #2

More details

- Q: SVD降维有多精确？
- A: 把最小的奇异值(一个或多个)设置为0.

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

SVD - Interpretation #2

More details

- Q: SVD降维有多精确？
- A: 把最小的奇异值(一个或多个)设置为0.

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

SVD - Interpretation #2

More details

- **Q:** SVD降维有多精确？
- **A:** 把最小的奇异值(一个或多个)设置为0.

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

29

SVD - Interpretation #2

More details

- **Q:** SVD降维有多精确？
- **A:** 把最小的奇异值(一个或多个)设置为0.

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

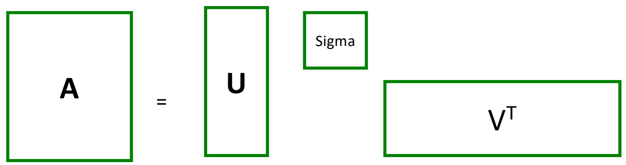
Frobenius norm:

$\|M\|_F = \sqrt{\Sigma_{ij} M_{ij}^2}$

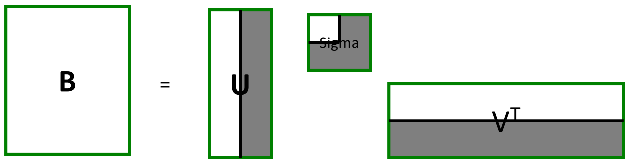
$\|A-B\|_F = \sqrt{\Sigma_{ij} (A_{ij}-B_{ij})^2}$ is “small”

30

SVD – 最优低秩近似.



B 是 A 的最优近似

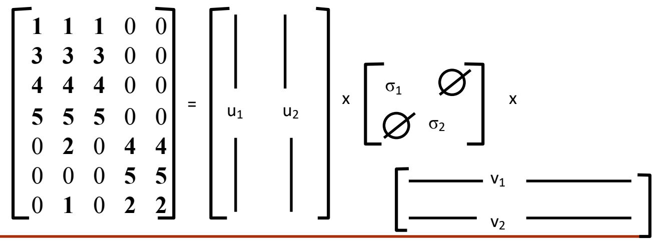


31

SVD - Interpretation #2

等价于:

矩阵的谱分解



32

SVD - Interpretation #2

等价于:

矩阵的谱分解

$$\underbrace{\begin{bmatrix} 1&1&1&0&0\\ 3&3&3&0&0\\ 4&4&4&0&0\\ 5&5&5&0&0\\ 0&2&0&4&4\\ 0&0&0&5&5\\ 0&1&0&2&2 \end{bmatrix}}_{n \times m} = \overbrace{\sigma_1 \underbrace{u_1}_{n \times 1} \underbrace{v^T_1}_{1 \times m} + \sigma_2 u_2 v^T_2 + \ldots}^{k \text{ terms}}$$

Assume: $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \ldots \geq 0$

为什么把最小的奇异值σ_i 设置为0是正确的？

向量 u_i v_i,是单位向量，所以 σ_i 起到缩放的作用。把小的奇异值 σ_i 设为0只会引入少量误差.

33

SVD - Interpretation #2

Q: 保留多少奇异值σ_s?

A: 经验:

keep 80-90% of ‘energy’ $= \sum_i \sigma_i^2$

$$\underbrace{\begin{bmatrix} 1&1&1&0&0\\ 3&3&3&0&0\\ 4&4&4&0&0\\ 5&5&5&0&0\\ 0&2&0&4&4\\ 0&0&0&5&5\\ 0&1&0&2&2 \end{bmatrix}}_{n \times m} = \sigma_1 u_1 v^T_1 + \sigma_2 u_2 v^T_2 + \ldots$$

Assume: $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \ldots$

34

SVD – 复杂度分析

- 计算SVD:
  - $O(nm^2)$ or $O(n^2m)$ (whichever is less)
- 但是:
  - 可以通过更少的计算得到
    - 如果只需要奇异值，
    - 或者只需要前K个特征向量
    - 或者矩阵是稀疏的
- 已经集成在许多线性代数计算包中：
  - LINPACK, Matlab, Python, SPlus, Mathematica ...

35

SVD – 总结

- SVD: A= U Σ V^T: 分解方法是唯一的
  - U: 用户—主题相似度
  - V: 电影—主题相似度
  - Σ : 每个主题的强度
- 降维:
  - 保存几个最大的奇异值
    (80-90% of ‘energy’)
  - SVD: 通过数据的轴的线性组合重构数据

36