Singular Value Decomposition

March 16-17, 2017

SVD Exercise 1

- Refresher on class material

SVD Theorem

Any real M by N matrix, $\mathbf{A} \in \mathbb{R}^{M \times N}$, can be decomposed as:

$$\underset{M \times N}{\mathbf{A}} = \underset{M \times M}{\mathbf{U}} \cdot \underset{M \times N}{\mathbf{D}} \cdot \underset{N \times N}{\mathbf{V}^\top}$$

- $\mathbf{U}$ is an M by M orthogonal matrix, such that $\mathbf{U}^\top \mathbf{U} = \mathbf{I}_M$.
- $\mathbf{D}$ is an M by N diagonal matrix
- $\mathbf{V}^\top$ is also an orthogonal matrix, N by N, $\mathbf{V}^\top \mathbf{V} = \mathbf{I}_N$.

SVD Interpretation

"Users", "Movies" and "Concepts":

- $\mathbf{U}$: Users-to-concept affinity matrix
- $\mathbf{V}$: Movies-to-concept similarity matrix
- $\mathbf{D}$: The diagonal elements of $\mathbf{D}$ represent the "expressiveness" of each concept in the data.

SVD Example

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

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

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

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

SVD Example

Concepts: **Horror, Comedy**

$\mathbf{U}$: Users-to-concept affinity matrix.

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

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

SVD Example

Concepts: Horror, Comedy

$\mathbf{D}$: Expression level of the different concepts in the data.

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

$$\begin{pmatrix} 5 & 5 & 5 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 4 & 4 \end{pmatrix} = \begin{pmatrix} 0.57 & 0 \\ 0.46 & 0 \\ 0.57 & 0 \\ 0.34 & 0 \\ 0 & 0.52 \\ 0 & 0.66 \\ 0 & 0.52 \end{pmatrix} \times \begin{pmatrix} 15 & 0 \\ 0 & 10.67 \end{pmatrix} \times \begin{pmatrix} 0.57 & 0.57 & 0.57 & 0 & 0 \\ 0 & 0 & 0 & 0.70 & 0.70 \end{pmatrix}$$

Strength of Horror

Strength of Comedy

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

SVD Example

Concepts: Horror, Comedy

$\mathbf{V}$: Movies-to-concept similarity matrix.

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

$$\begin{pmatrix} 5 & 5 & 5 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 4 & 4 \end{pmatrix} = \begin{pmatrix} 0.57 & 0 \\ 0.46 & 0 \\ 0.57 & 0 \\ 0.34 & 0 \\ 0 & 0.52 \\ 0 & 0.66 \\ 0 & 0.52 \end{pmatrix} \times \begin{pmatrix} 15 & 0 \\ 0 & 10.67 \end{pmatrix} \times \begin{pmatrix} 0.57 & 0.57 & 0.57 & 0 & 0 \\ 0 & 0 & 0 & 0.70 & 0.70 \end{pmatrix}$$

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

What is the similarity between Clerks and Comedy? 0.7

Closest matrix approximation

Let the SVD of $\mathbf{A} \in \mathbb{R}^{M \times N}$ be given by $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$

Define $\mathbf{A}_k$ as

$$\mathbf{A}_k = \sum_{i=1}^{k} d_i \mathbf{u}_i \mathbf{v}_i^\top$$

Where $k < r = \text{Rank}(\mathbf{A})$

Home Exercise: Prove that the above formula is equivalent to $\mathbf{A}_k = \mathbf{U}_k \mathbf{D}_k \mathbf{V}_k^\top$, where $\mathbf{U}_k, \mathbf{D}_k, \mathbf{V}_k$ are obtained by keeping only the first k columns and rows from the matrices U, D, V.

Frobenius norm

Def. The Frobenius norm is matrix norm, defined as the square root of the sum of the absolute squares of its elements.
For $\mathbf{A} \in \mathbb{R}^{M \times N}$:

$$\|\mathbf{A}\|_F := \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} |A_{i,j}|^2}$$

Eckart-Young theorem:

- The matrix $\mathbf{A}_k$ is the closest k-rank matrix to matrix $\mathbf{A}$, under both the frobenius norm and the 2-norm (spectral norm).

Frobenius norm

Comparison with the euclidian norm:

- $\|\mathbf{A}\|_2 = d_1$
- $\|\mathbf{A}\|_F^2 = d_1^2 + \ldots + d_r^2$

Frobenius norm

Comparison with the euclidian norm:

- $\|\mathbf{A}\|_2 = d_1$
- $\|\mathbf{A}\|_F^2 = d_1^2 + \ldots + d_r^2$

Exercise: prove the above formula.
Hint: use the fact that $\|\mathbf{A}\|_F^2 = \|\mathbf{UA}\|_F^2$ for any orthogonal matrix U.

Frobenius norm

Comparison with the euclidian norm:

- $\|\mathbf{A}\|_2 = d_1$
- $\|\mathbf{A}\|_F^2 = d_1^2 + \ldots + d_r^2$

Exercise: prove the above formula.
Hint: use the fact that $\|\mathbf{A}\|_F^2 = \|\mathbf{U}\mathbf{A}\|_F^2$ for any orthogonal matrix U.

Eckart-Young theorem:

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

$$\min_{\text{Rank}(\mathbf{B})=k} \|\mathbf{A}-\mathbf{B}\|_F^2 = \|\mathbf{A}-\mathbf{A_k}\|_F^2 = d_{k+1}^2 + \ldots + d_r^2$$

Stochastic Gradient Descent for CF

- Problem of SVD for CF: initializing missing values, low number of initial ratings
- Low rank approximation $\mathbf{A} = \mathbf{Q} \cdot \mathbf{P}$, where $\mathbf{A} \in \mathbb{R}^{M \times N}, \mathbf{Q} \in \mathbb{R}^{M \times K}, \mathbf{P} \in \mathbb{R}^{K \times N}$
- Rows of $\mathbf{Q}$ contain user embeddings, columns of $\mathbf{P}$ item embeddings
- Optimize only over known ratings (no need to input missing values):
  - $$\min_{\mathbf{Q},\mathbf{P}} \sum_{(u,i) \in \Omega(A)} (a_{ui} - q_u^\top p_i)^2 + \lambda(\|q_u\|^2 + \|p_i\|^2)$$
- Optimize this function using SGD : update embeddings for one random rating at a time.
- More about this in the Optimization lecture
- Tutorial Link

SVD Exercise 1

- Solve the "Pen and Paper" exercise