# 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:

$$\begin{bmatrix} \mathbf{A} \\ \mathbf{D} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{D} \\ \mathbf{D} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{V}^{\top} \\ N \times N \end{bmatrix}$$

$$M \times N \qquad M \times M \qquad M \times N$$

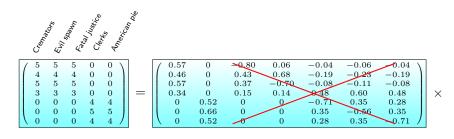
- $lackbox{U}$  is an M by M orthogonal matrix, such that  $\mathbf{U}^{\top}\mathbf{U} = \mathbf{I}_{M}$ .
- ightharpoonup D is an M by N diagonal matrix
- $ightharpoonup \mathbf{V}^{\top}$  is also an orthogonal matrix, N by N,  $\mathbf{V}^{\top}\mathbf{V} = \mathbf{I}_{N}$ .

# **SVD** Interpretation

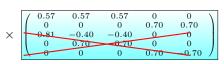
```
"Users", "Movies" and "Concepts":
```

- ▶ U: Users-to-concept affinity matrix
- ▶ V: Movies-to-concept similarity matrix
- ▶ **D**: The diagonal elements of **D** represent the "expressiveness" of each concept in the data.

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

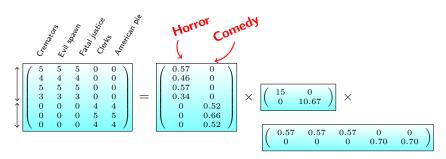


| 1  | 15 | 0     | 0  | 0           | 0 \ |   |
|----|----|-------|----|-------------|-----|---|
| 11 | 0  | 10.67 | 0  | 0           | 0   | ١ |
|    | 0  | 0     | Q  | 0           | 0/  | П |
|    | 0  | 0     | 0  | 0           | /0  | П |
|    | 0  | 0     | 0  | <b>%</b> /  | 0   | П |
| Н  | 0  | 0     | 0  | <b>∕</b> 0` | 0   | Н |
| /  | 0  | 0     | 9/ | 0           | 0   | ' |
|    |    |       |    |             |     |   |



#### Concepts: Horror, Comedy

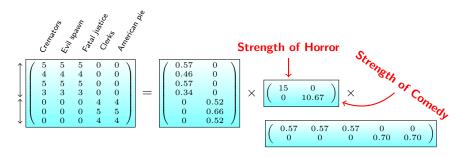
U: Users-to-concept affinity matrix.



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

Concepts: Horror, Comedy

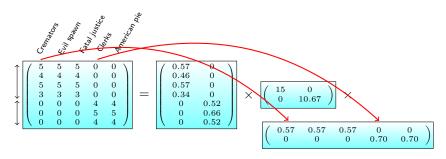
D: Expression level of the different concepts in the data.



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

Concepts: Horror, Comedy

V: Movies-to-concept similarity matrix.



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

Home Exercise: Prove that the above formula is equivalent to  $\mathbf{A}_k = \mathbf{U}_k \mathbf{D}_k \mathbf{V}_k^{\mathsf{T}}$ , 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.

**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  $A_k$  is the closest k-rank matrix to matrix A, under both the frobenius norm and the 2-norm (spectral norm).

#### Comparison with the euclidian norm:

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

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.

#### 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:

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

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

### 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 Q contain user embeddings, columns of 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