

Singular value decomposition

Victor Kitov

Table of Contents

- 1 Definition of SVD
- 2 Reduced (truncated) SVD
- 3 Applications of SVD
- 4 Recommendation system with SVD

SVD decomposition¹²

Every matrix $X \in \mathbb{R}^{N \times D}$, $\text{rank } X = R$, can be decomposed into the product of three matrices:

$$X = U \Sigma V^T$$

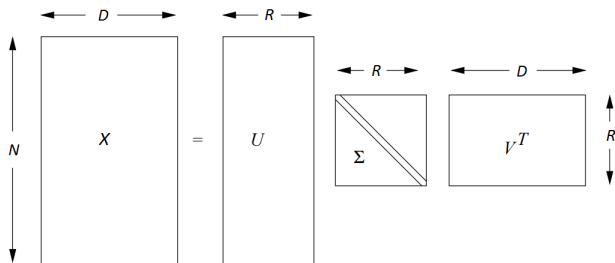
where

- $U \in \mathbb{R}^{N \times R}$, $\Sigma \in \mathbb{R}^{R \times R}$, $V^T \in \mathbb{R}^{R \times D}$
- $\Sigma = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_R\}$, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_R \geq 0$,
- $U^T U = I$, $V^T V = I$, where $I \in \mathbb{R}^{R \times R}$ is identity matrix.

¹Prove it

²Is it unique?

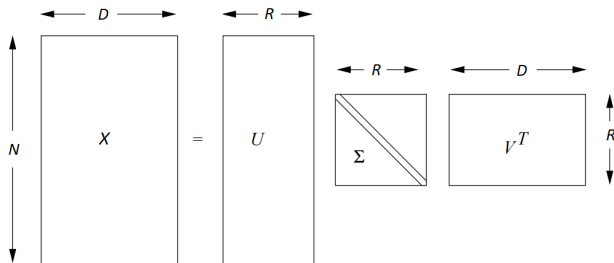
Interpretation of SVD



For X_{ij} let i denote objects and j denote properties.

- Columns of U - orthonormal basis of columns of X
- Rows of V^T - orthonormal basis of rows of X
- Σ - scaling.
- Efficient representations of low-rank matrix!

Interpretation of SVD



For X_{ij} let i denote objects and j denote properties.

- Rows of U are normalized coordinates of rows in V^T
- $\Sigma = \text{diag}\{\sigma_1, \dots, \sigma_R\}$ shows the magnitudes of presence of each row from V^T .

Finding U and V

- **Finding U :**

$$XX^T = U\Sigma V^T (U\Sigma V^T)^T = U\Sigma V^T V\Sigma U^T = U\Sigma^2 U^T. \text{ So}$$

$$XX^T U = U\Sigma^2 U^T U = U\Sigma^2.$$

So U consists of eigenvectors of XX^T with corresponding eigenvalues $\sigma_1^2, \sigma_2^2, \dots, \sigma_R^2$.

Finding U and V

- **Finding U :**

$XX^T = U\Sigma V^T (U\Sigma V^T)^T = U\Sigma V^T V\Sigma U^T = U\Sigma^2 U^T$. So

$$XX^T U = U\Sigma^2 U^T U = U\Sigma^2.$$

So U consists of eigenvectors of XX^T with corresponding eigenvalues $\sigma_1^2, \sigma_2^2, \dots, \sigma_R^2$.

- **Finding V**

$X^T X = (U\Sigma V^T)^T U\Sigma V^T = (V\Sigma U^T)U\Sigma V^T = V\Sigma^2 V^T$. It follows that

$$X^T X V = V\Sigma^2 V^T V = V\Sigma^2$$

So V consists of eigenvectors of $X^T X$ with corresponding eigenvalues $\sigma_1^2, \sigma_2^2, \dots, \sigma_R^2$ - **these are top R principal components!**

SVD: existence

Theorem 1

For any matrix $X \in \mathbb{R}^{N \times D}$ SVD decomposition exists.

Proof. Consider arbitrary $X = [x_1^T, \dots, x_N^T]^T \in \mathbb{R}^{N \times D}$ with $\text{rg } X = R$.

For rows x_1^T, \dots, x_N^T find principal components v_1, \dots, v_R .

Define $V^T = [v_1^T, \dots, v_R^T] \in \mathbb{R}^{R \times D}$. By definition of principal components $V^T V = I$. Consider B with rows=coordinates of x_1, \dots, x_N in principal components, then $X = BV^T$.

Let b_1, \dots, b_D be columns of B , satisfying $b_i = Xv_i$. Then $b_i^T b_j = v_i^T X^T X v_j = \lambda_j v_i^T v_j = \lambda_j \mathbb{I}[i = j]$, because v_j is an eigenvector of $X^T X$ with eigenvalue λ_j . Also $\lambda_j \geq 0$ because $X^T X \succeq 0$. So b_1, \dots, b_D are orthogonal.

If we consider $\Sigma = \text{diag}\{\sqrt{\lambda_1}, \dots, \sqrt{\lambda_D}\}$ $B = U\Sigma$ we will obtain that $U^T U = I$. So SVD decomposition $X = U\Sigma V^T$ exists. \square

SVD: uniqueness

Theorem

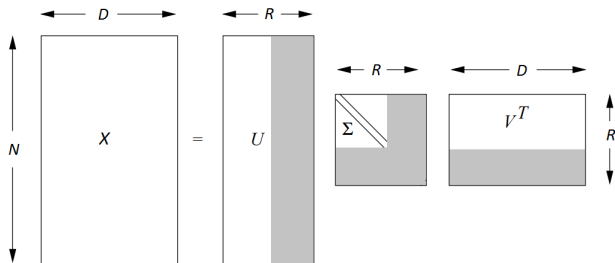
SVD decomposition is unique if $X^T X \in \mathbb{R}^{D \times D}$ has a set of D unique eigenvalues.

- Unique set of eigenvalues mean that eigenvectors are uniquely defined (up to multiplicative constant).
- If two eigenvalues are equal respective eigenvectors are not uniquely defined.
- Sometimes condition $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_R \geq 0$ is not required.
 - Then we can freely change ordering simultaneously of u_1, \dots, u_R ; $\sigma_1, \dots, \sigma_R$; v_1, \dots, v_R .
 - So SVD is not unique

Table of Contents

- 1 Definition of SVD
- 2 Reduced (truncated) SVD**
- 3 Applications of SVD
- 4 Recommendation system with SVD

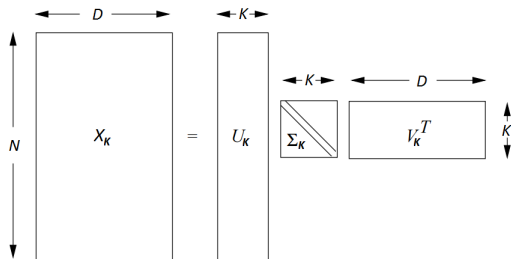
Reduced SVD decomposition



$$\Sigma = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_K, \sigma_{K+1}, \dots, \sigma_R\} \longrightarrow$$

$$\text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_K, 0, 0, \dots, 0\} = \Sigma_K$$

Reduced SVD decomposition



Simplification to rank $K \leq R$:

$$X_K = U_K \Sigma_K V_K$$

$$\Sigma = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_K, \sigma_{K+1}, \dots, \sigma_R\} \longrightarrow \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_K\} = \Sigma_K$$

$$U = [u_1, u_2, \dots, u_K, u_{K+1}, \dots, u_R] \longrightarrow [u_1, u_2, \dots, u_K] = U_K$$

$$V = [v_1, v_2, \dots, v_K, v_{K+1}, \dots, v_R] \longrightarrow [v_1, v_2, \dots, v_K] = V_K$$

- Now rows of U give reduced representation of rows of X .

Properties of reduced SVD decomposition

Frobenius norm of matrix

$$\|X\|_F^2 = \sum_{n=1}^N \sum_{d=1}^D x_{nd}^2$$

- For matrix X and its approximation \hat{X} we can measure

$$\text{approximation error} = \left\| \hat{X} - X \right\|_F^2$$

Theorem 2

Suppose $X \in \mathbb{R}^{N \times D}$, is approximated with $\hat{X}_K = U_K \Sigma_K V_K$. Then:

- 1 rank $X_K = K$.
- 2 $X_K = \arg \min_{B: \text{rank } B \leq K} \|X - B\|_F^2$

Proof of theorem 2

- ① $\text{rg } U_K = \text{rg } U_K \Sigma_K = K$, $\text{rg } V_K = K$, so
 $\text{rg } \hat{X}_K = \text{rg } [U_K \Sigma_K V_K] = K$
- ② Let $X = [x_1, \dots, x_N]^T$, $B = [b_1, \dots, b_N]^T$, $D = U \Sigma$,
 $D_K = U_K \Sigma_K$, so $X = DV$, $\hat{X}_K = D_K V_K$
- ① consider subspace L spanned by b_1, \dots, b_N . Since
 $\text{rg } B \leq K, \dim(L) \leq K$.
- ② $\|X - B\|_F^2 = \sum_{n=1}^N \|x_n - b_n\|^2 \leq \sum_{n=1}^N \|x_n - \tilde{b}_n\|^2$, where \tilde{b}_n
 is projection of x_n on L .
- ③ Since rows of V_K are top K principal components, rows of D_K
 are coordinates in first K principal components, and
 $\hat{X}_K = [p_1, \dots, p_N]^T$ consists of projections onto K best fit
 subspace.
- ④ $\|X - \hat{X}_K\|_F^2 = \|[x_1 - p_1, \dots, x_N - p_N]^T\|_F^2 =$
 $\sum_{n=1}^N \|x_n - p_n\|^2 \leq \sum_{n=1}^N \|x_n - \tilde{b}_n\|^2 \leq \sum_{n=1}^N \|x_n - b_n\|^2 =$
 $\|X - B\|_F^2$

Which K to choose for approximation?

- Suppose $X = U\Sigma V^T$, $\Sigma = \text{diag}\{\sigma_1, \dots, \sigma_R\}$
- Approximation $\hat{X}_K = U\Sigma_K V^T$, $\Sigma = \text{diag}\{\sigma_1, \dots, \sigma_K, 0, 0, \dots, 0\}$.
- Then error of approximation $E_K = X - \hat{X}_K = U\tilde{\Sigma}V^T$, where $\tilde{\Sigma} = \text{diag}\{0, 0, \dots, 0, \sigma_{K+1}, \dots, \sigma_R\}$

Which K to choose for approximation?

Select K giving relative error below some threshold t :

$$K = \arg \min_K \left\{ \frac{\|E_K\|_F^2}{\|X\|_F^2} = \frac{\sum_{i=K+1}^R \sigma_i^2}{\sum_{i=1}^R \sigma_i^2} < t \right\}$$

We used theorem 3 for calculation of Frobenius matrix norm.

Frobenius norm

Theorem 3

for any matrix X and its singular value decomposition $A = U\Sigma V^T$, $\Sigma = \text{diag}\{\sigma_1, \dots, \sigma_R\}$:

$$\|X\|_F^2 = \sum_{i=1}^R \sigma_i^2$$

Proof. Using lemmas 1 and 2, we obtain:

$$\begin{aligned}\|X\|_F^2 &= \text{tr}[U\Sigma V^T V\Sigma U^T] = \text{tr}[U(\Sigma^2 U^T)] = \\ &= \text{tr}[(\Sigma^2 U^T)U] = \text{tr}[\Sigma^2] = \sum_{r=1}^R \sigma_r^2\end{aligned}$$



Lemmas

Lemma 1

For any $X \in \mathbb{R}^{N \times D}$ $\|X\|_F^2 = \text{tr} XX^T$

Proof. $\{XX^T\}_{i,j} = \sum_{k=1}^D x_{ik}x_{kj}^t = \sum_{k=1}^D x_{ik}x_{jk}$. So

$$\text{tr} XX^T = \sum_{i=1}^N \{XX^T\}_{i,i} = \sum_{i=1}^N \sum_{k=1}^D x_{ik}x_{ik} = \|X\|_F^2$$



Lemmas

Lemma 2

For any $A \in \mathbb{R}^{N \times D}$ and $B \in \mathbb{R}^{D \times N}$

$$\text{tr } AB = \text{tr } BA$$

Proof. $\{AB\}_{n,n} = \sum_{d=1}^D a_{n,d} b_{d,n}$, so

$$\text{tr } AB = \sum_{n=1}^N \{AB\}_{n,n} = \sum_{n=1}^N \sum_{d=1}^D a_{n,d} b_{d,n}$$

$\{BA\}_{d,d} = \sum_{n=1}^N b_{d,n} a_{n,d}$, so

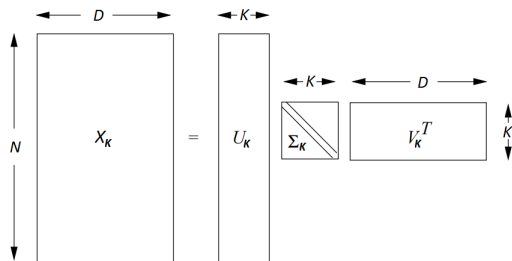
$$\begin{aligned} \text{tr } BA &= \sum_{d=1}^D \{BA\}_{d,d} = \sum_{d=1}^D \sum_{n=1}^N b_{d,n} a_{n,d} = \\ &= \sum_{n=1}^N \sum_{d=1}^D a_{n,d} b_{d,n} = \text{tr } AB \end{aligned}$$



Table of Contents

- 1 Definition of SVD
- 2 Reduced (truncated) SVD
- 3 Applications of SVD**
- 4 Recommendation system with SVD

Dimensionality reduction



- rows of U give reduced representation of rows of X .
- $x_n \in \mathbb{R}^D \longrightarrow u_n \in \mathbb{R}^K$

Memory efficiency

Storage costs of $X \in \mathbb{R}^{N \times D}$, assuming $N \geq D$ and each element taking 1 byte:

Memory storage costs

representation of X	memory requirements
original X	?
fully SVD decomposed	?
reduced SVD to rank K	?

Performance efficiency

- Multiplication Xq
 - X - normalized documents representation
 - q - normalized search query

representation of X	Xq complexity
original X	?
reduced SVD to rank K	?

Table of Contents

- 1 Definition of SVD
- 2 Reduced (truncated) SVD
- 3 Applications of SVD
- 4 Recommendation system with SVD**

Example

	Terminator	Gladiator	Rambo	Titanic	Love story	A walk to remember
Andrew	4	5	5	0	0	0
John	4	4	5	0	0	0
Matthew	5	5	4	0	0	0
Anna	0	0	0	5	5	5
Maria	0	0	0	5	5	4
Jessika	0	0	0	4	5	4

Example

$$U = \begin{pmatrix} 0. & 0.6 & -0.3 & 0. & 0. & -0.8 \\ 0. & 0.5 & -0.5 & 0. & 0. & 0.6 \\ 0. & 0.6 & 0.8 & 0. & 0. & 0.2 \\ 0.6 & 0. & 0. & -0.8 & -0.2 & 0. \\ 0.6 & 0. & 0. & 0.2 & 0.8 & 0. \\ 0.5 & 0. & 0. & 0.6 & -0.6 & 0. \end{pmatrix}$$

$$\Sigma = \text{diag}\{(14. \quad 13.7 \quad 1.2 \quad 0.6 \quad 0.6 \quad 0.5)\}$$

$$V^T = \begin{pmatrix} 0. & 0. & 0. & 0.6 & 0.6 & 0.5 \\ 0.5 & 0.6 & 0.6 & 0. & 0. & 0. \\ 0.5 & 0.3 & -0.8 & 0. & 0. & 0. \\ 0. & 0. & 0. & -0.2 & 0.8 & -0.6 \\ -0. & -0. & -0. & 0.8 & -0.2 & -0.6 \\ 0.6 & -0.8 & 0.2 & 0. & 0. & 0. \end{pmatrix}$$

Example (excluded insignificant concepts)

$$U_2 = \begin{pmatrix} 0. & 0.6 \\ 0. & 0.5 \\ 0. & 0.6 \\ 0.6 & 0. \\ 0.6 & 0. \\ 0.5 & 0. \end{pmatrix}$$

$$\Sigma_2 = \text{diag}\{(14. \quad 13.7)\}$$

$$V_2^T = \begin{pmatrix} 0. & 0. & 0. & 0.6 & 0.6 & 0.5 \\ 0.5 & 0.6 & 0.6 & 0. & 0. & 0. \end{pmatrix}$$

Concepts may be

- patterns among movies (along j) - action movie / romantic movie
- patterns among people (along i) - boys / girls

Dimensionality reduction case: patterns along j axis.

Applications

- Example: new movie rating by new person

$$x = (5 \ 0 \ 0 \ 0 \ 0 \ 0)$$

- **Dimensionality reduction:** map x into concept space:

$$y = V_2^T x = (0 \ 2.7)$$

- **Recommendation system:** map y back to original movies space:

$$\hat{x} = yV_2^T = (1.5 \ 1.6 \ 1.6 \ 0 \ 0 \ 0)$$