Data Pre-Processing-IV

(Data Reduction- SVD, LDA)

Dr. JASMEET SINGH

ASSISTANT PROFESSOR, CSED

TIET, PATIALA

Singular Valued Decomposition (SVD)

- In linear algebra, the singular value decomposition (SVD) is a factorization of a real or complex matrix.
- Formally, a matrix A of order m × n can be decomposed using SVD as follows:

$$A_{m\times n} = U_{m\times m}\Sigma_{m\times n}V^T_{n\times n}$$

- where U and V are column unit orthonormal vectors and Σ is a rectangular diagonal matrix whose diagonal entries are the singular values of matrix A.
- The number of non zero singular values is the rank of A.

Singular Valued Decomposition- Contd…

- U and V are orthonormal i.e.

$$UU^T = I \ or \ U^T = U^{-1}$$

$$\mathrm{V}V^T = I \ or \ V^T = V^{-1}$$

- Singular values of any matrix $M_{m\times n}$ is the positive square root of the eigen values of matrix $M^T M$ of order n×n.

How to Compute U, Σ, and V ?

- Σ is a rectangular diagonal matrix of singular values of A.
- So, **in order to compute *Σ*, calculate eigen value of A^TA or AA^T i.e.**
  - Find λ's such that $|A^TA - \lambda I| = 0$
  - Compute positive square root of λ's to find singular values of A (say $\sigma_1, \sigma_2, \sigma_3, \dots \dots .. \sigma_n$) such that $\sigma_1 > \sigma_2 > \sigma_3 \dots\dots\dots > \sigma_n$
  - The diagonal entries of Σ is $(\sigma_1, \sigma_2, \sigma_3, \dots \dots .. \sigma_n)$ and rest all entries are 0.

How to Compute U, Σ, and V ?

- V is the column normalized eigen vectors of A^TA as explained below:

$$A^TA = (U\Sigma V^T)^T\,(U\Sigma V^T)$$

$$= V\Sigma^T U^T \mathrm{U}\Sigma V^T$$

$$= V\Sigma\Sigma^T V^T \quad \text{(because U is orthonormal)}$$

$$= V\Sigma^2 V^T \quad \text{(because for diagonal matrix } AA^T = A^2\text{)}$$

Where, Σ^2 is the eigen value matrix of A^TA. So according to diagonalization process,

Therefore, V represents eigen vector of A^TA , since it is column unit vector so it must be normalized by each column.

How to Compute U, Σ, and V ?

- U is the column normalized eigen vectors of AA^T as explained below:

$$AA^T = (U\Sigma V^T)(U\Sigma V^T)^T$$

$$= \mathrm{U}\Sigma V^T V \Sigma^T U^T$$

$$= \mathrm{U}\Sigma\Sigma^T U^T \quad \text{(because V is orthonormal)}$$

$$= \mathrm{U}\Sigma^2 U^T \quad \text{(because for diagonal matrix } AA^T=A^2\text{)}$$

Where, Σ^2 is the eigen value matrix of AA^T. So according to diagonalization process,

Therefore, U represents eigen vector of AA^T , since it is column unit vector so it must be normalized by each column.

How to Compute U, Σ, and V ?

- Alternatively, we can find U or V (anyone) using column normalized eigen vector of AA^T or A^TA respectively and then other can be found as

$$u_i = \frac{1}{\sigma_i} A v_i \text{ (because AV=U } \Sigma\text{)}$$

$$\text{or } v_i = \frac{1}{\sigma_i} A^T u_i \text{ (because } A^TU = V\ \Sigma\text{)}$$

SVD Example

Find the SVD of A, $U\Sigma V^T$, where

$$A = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix}$$

SVD Example

Solution:

First we compute the singular values σ_i by finding the eigenvalues of AA^T

$$AA^T = \begin{pmatrix} 17 & 8 \\ 8 & 17 \end{pmatrix}$$

The characteristic polynomial is $\det(AA^T - \lambda I) = \lambda^2 - 34\lambda + 225 = (\lambda - 25)(\lambda - 9)$,

so the singular values are $\sigma_1 = \sqrt{25} = 5$ and $\sigma_2 = \sqrt{9} = 3$.

$$\text{Therefore } \Sigma = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix}$$

In case, we will A^TA , we will have a 3X3 matrix and three values of λ which will be 25, 9, and 0.

SVD Example

- Now we find the columns of V by finding an orthonormal set of eigenvectors of A^TA. The eigenvalues of A^TA are 25, 9, and 0.
- For λ = 25, we have, $A^TA - 25I = \begin{pmatrix} -12 & 12 & 2 \\ 12 & -12 & -2 \\ 2 & -2 & -17 \end{pmatrix}$

The column normalized eigen vector of the above matrix is $v_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix}$

- For λ = 9, we have, $A^TA - 9I = \begin{pmatrix} 4 & 12 & 2 \\ 12 & 4 & -2 \\ 2 & -2 & -1 \end{pmatrix}$

The column normalized eigen vector of the above matrix is $v_2 = \begin{pmatrix} 1/\sqrt{18} \\ -1/\sqrt{18} \\ 4/\sqrt{18} \end{pmatrix}$

SVD Example

- For λ = 0, we have, $A^T A = \begin{pmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{pmatrix}$

The column normalized eigen vector of the above matrix is $v_2 = \begin{pmatrix} 2/3 \\ -2/3 \\ 1/3 \end{pmatrix}$

Therefore, $V = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{18} & 2/3 \\ 1/\sqrt{2} & -1/\sqrt{18} & -2/3 \\ 0 & 4/\sqrt{18} & 1/3 \end{pmatrix}$

SVD Example

Finally, we can compute U by the formula $u_i = \frac{1}{\sigma_i} A v_i$

$u_1 = \frac{1}{\sigma_1} A v_1 \quad u_2 = \frac{1}{\sigma_2} A v_2$

This gives $U = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$

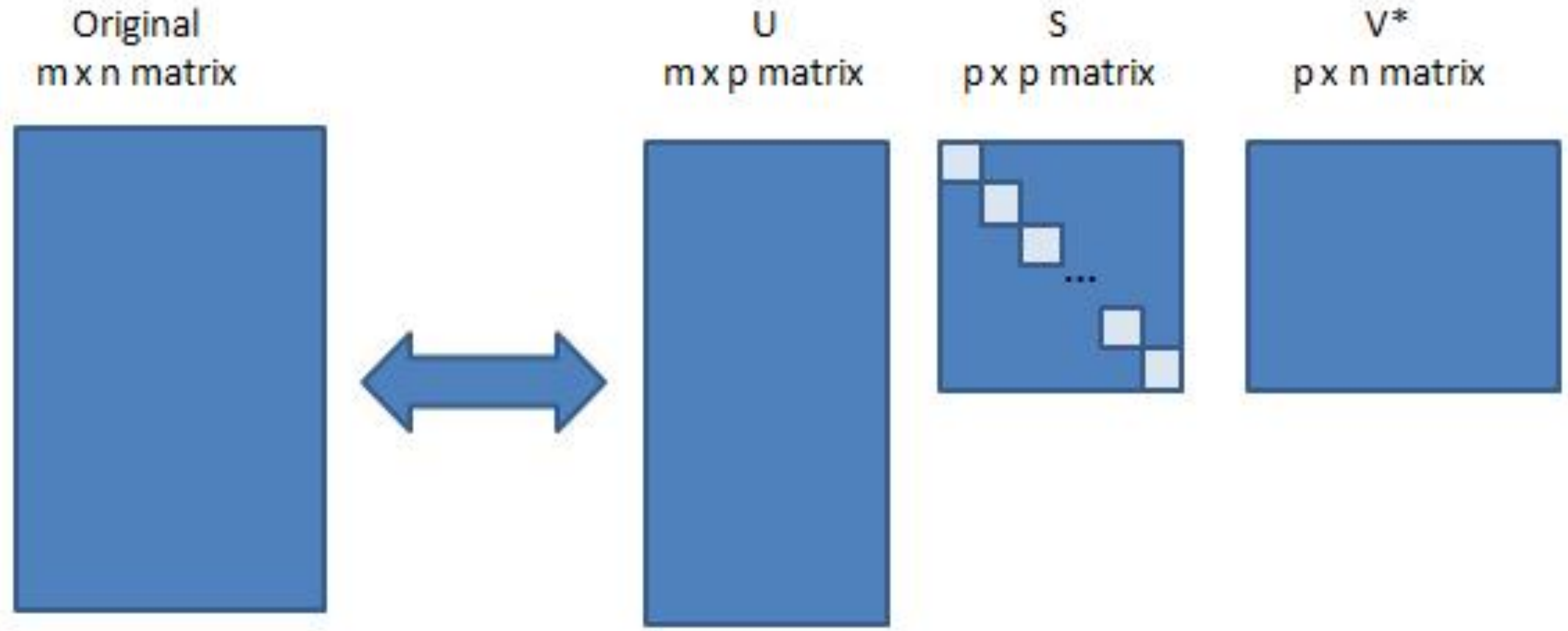
So in its full glory the SVD is:

A = UVΣ ^{T} = $\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{18} & 2/3 \\ 1/\sqrt{2} & -1/\sqrt{18} & -2/3 \\ 0 & 4/\sqrt{18} & 1/3 \end{pmatrix}$

SVD for Dimensionality Reduction

- SVD is used for dimensionality reduction by using **compressed SVD**.
- In compressed SVD, dimensionality reduction is done by neglecting small singular values in the diagonal matrix Σ.
- In compressed SVD, the factorization has the form $U \Sigma V^T$. U is an $m \times p$ matrix. Σ is a $p \times p$ diagonal matrix. V is an $n \times p$ matrix, with V^T being the transpose of V, a $p \times n$ matrix, or the conjugate transpose if M contains complex values. The value p is called the rank.

SVD for Dimensionality Reduction



Applications of SVD

- SVD, might be the most popular technique for dimensionality reduction when data is sparse.
- Sparse data refers to rows of data where many of the values are zero.
- This is often the case in some problem domains like recommender systems where a user has a rating for very few movies or songs in the database and zero ratings for all other cases.
- Another common example is a bag of words model of a text document, where the document has a count or frequency for some words and most words have a 0 value.

Applications of SVD

Examples of sparse data appropriate for applying SVD for dimensionality reduction:

- Recommender Systems
- Customer-Product purchases
- User-Song Listen Counts
- User-Movie Ratings
- Text Classification
- One Hot Encoding
- Bag of Words Counts
- TF/IDF

Applications of SVD

- $A = U \Sigma V^T$ - example: **Users to Movies**

$$\begin{array}{cc}
 & \begin{array}{ccccc} \text{Matrix} & \text{Alien} & \text{Serenity} & \text{Casablanca} & \text{Amelie} \end{array} \\
\begin{array}{c} \\ \text{SciFi} \\ \\ \\ \\ \text{Romance} \\ \\ \end{array} & \left[\begin{array}{ccccc} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 2 & 2 \end{array}\right]
\end{array}
=
\begin{array}{c}
\begin{array}{cc} \text{SciFi-concept} & \text{Romance-concept} \end{array} \\
\begin{bmatrix} \mathbf{0.14} & 0.00 \\ \mathbf{0.42} & 0.00 \\ \mathbf{0.56} & 0.00 \\ \mathbf{0.70} & 0.00 \\ 0.00 & \mathbf{0.60} \\ 0.00 & \mathbf{0.75} \\ 0.00 & \mathbf{0.30} \end{bmatrix}
\end{array}
\times
\begin{bmatrix} 12.4 & 0 \\ 0 & 9.5 \end{bmatrix}
\times
\begin{bmatrix} \mathbf{0.58} & \mathbf{0.58} & \mathbf{0.58} & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & \mathbf{0.71} & \mathbf{0.71} \end{bmatrix}$$