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:

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

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 $u_i = \frac{1}{\sigma_i} A v_i$

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:

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

Relation between PCA and SVD

- Let the data matrix X be of n×p size, where n is the number of samples and p is the number of variables.
- Then the p×p covariance matrix C is a symmetric matrix and so it can be diagonalized:

$$C=VLV^{\top},$$

- where V is a matrix of eigenvectors (each column is an eigenvector) and L is a diagonal matrix with eigenvalues λi in the decreasing order on the diagonal.
- The eigenvectors are called *principal axes* or *principal directions* of the data.
- The coordinates of the i-th data point in the new PC space are given by the i-th row of XV.

Relation between PCA and SVD- Contd….

- If, we perform SVD on X we will get $X = U \sum V^T$
- If X is *centered*, i.e. column means have been subtracted and are now equal to zero, then the covariance matrix C is given by:

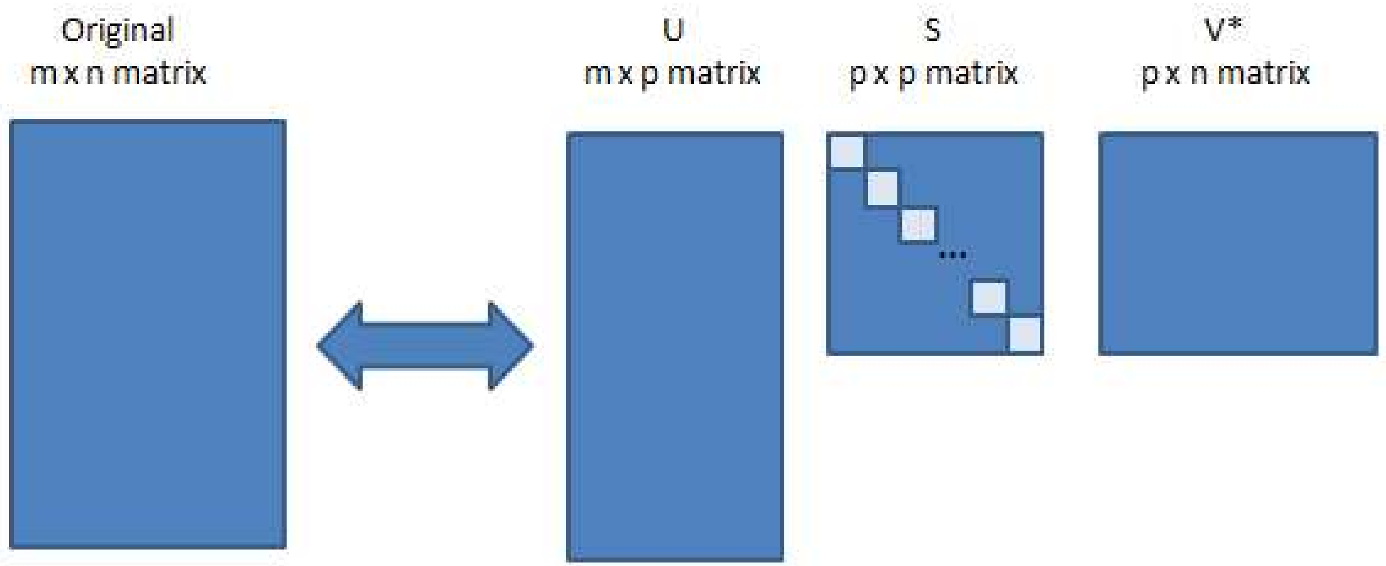
$$C = \frac{X^T X}{n-1} = \frac{V\Sigma^T U^T U \Sigma V^T}{n-1} = \frac{V\Sigma^2 V^T}{n-1} = V \frac{\Sigma^2}{n-1} V^T$$

- V are principal directions and that singular values are related to the eigenvalues of covariance matrix via $\lambda = \Sigma^2/(n-1)$.
- Transformed dataset are given by $XV=U \sum V^T V = U \sum$

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

SciFi-concept

Romance-concept

SciFi

Romance

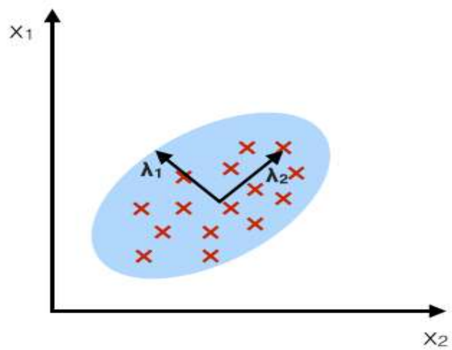
$$\begin{array}{c} \begin{matrix} \text{Matrix} & \text{Alien} & \text{Serenity} & \text{Casablanca} & \text{Amelie} \end{matrix} \\ \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 2 & 2 \end{bmatrix} \end{array} = \begin{bmatrix} \mathbf{0.14} & 0.00 \\ \mathbf{0.42} & 0.00 \\ \mathbf{0.56} & 0.00 \\ \mathbf{0.70} & 0.00 \\ 0.00 & \mathbf{0.60} \\ 0.00 & \mathbf{0.75} \\ 0.00 & \mathbf{0.30} \end{bmatrix} \times \begin{bmatrix} 12.4 & 0 \\ 0 & 9.5 \end{bmatrix} \times \begin{bmatrix} \mathbf{0.58} & \mathbf{0.58} & \mathbf{0.58} & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & \mathbf{0.71} & \mathbf{0.71} \end{bmatrix}$$

Linear Discriminant Analysis (LDA)

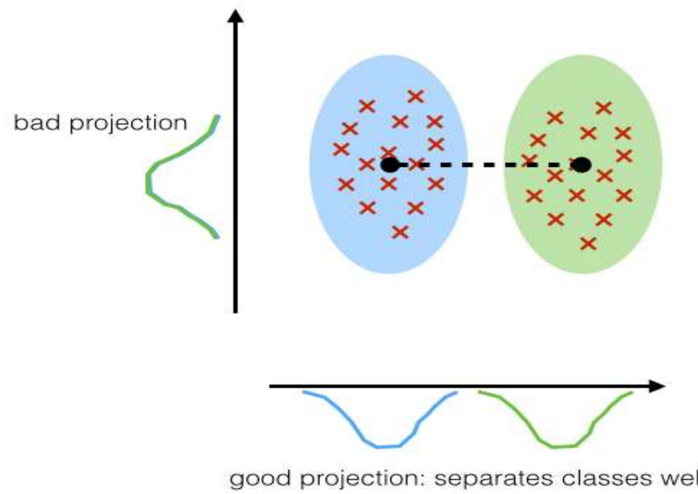
- Both Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) are linear transformation techniques that are commonly used for dimensionality reduction.
- PCA can be described as an **“unsupervised”** algorithm, since it “ignores” class labels and its goal is to find **the directions (the so-called principal components) that maximize the variance in a dataset.**
- In contrast to PCA, **LDA is “supervised” and computes the directions (“linear discriminants”) that will represent the axes that that maximize the separation between multiple classes.**
- Unlike **PCA that calculates eigenvalues and eigenvectors of the covariance matrix** of the data, **LDA calculates eigen value and eigen vectors using the within-class and between class scatter (variance) matrix**.

Linear Discriminant Analysis (LDA)

PCA:
component axes that maximize the variance



LDA:
maximizing the component axes for class-separation



Step-by-Step Working of LDA

- Consider a labelled dataset with n datapoints that is classified into p classes. Let each class contain n_1, n_2 , n_3,……n_p datapoints.

1. Compute the **mean vector** of each class (μ_1, μ_2, μ_3…… μ_p) and the global mean vector (μ).
2. Compute **the between-class scatter matrix** S_B given by:

$$S_B = \sum_{i=1}^{p} S_{Bi}$$

where $S_{Bi} = n_i(\mu_i - \mu)^T(\mu_i - \mu)$

Step-by-Step Working of LDA

3. Compute Within-class scatter matrix Sw given by:

$$S_w = \sum_{j=1}^{p} S_{wj}$$

Where $S_{wj} = \sum_{i=1}^{n_j} (x_{ij} - \mu_j)^T (x_{ij} - \mu_j)$

x_{ij} represents the i^{th} sample in the j^{th} class

4. Compute the scatter matrix from between-class and within-class variance given by $S_W^{-1} S_B$

Step-by-Step Working of LDA

5. Compute the eigen values of the scatter matrix ($S_W^{-1} S_B$) i.e. find λ's such that det($S_W^{-1} S_B$- λI)=0.

6. Sort the eigenvectors by decreasing eigenvalues and choose k eigenvectors with the largest eigenvalues to form a p×k dimensional matrix W (where every column represents an eigenvector) .

7. Use this p X k eigenvector matrix to transform the samples onto the new subspace. This can be summarized by the matrix multiplication: **Y=X×W** (where X is a n×p matrix representing the n samples and p classes, and Y are the transformed n×k-dimensional samples in the new subspace)

LDA Example

Given two different classes, $\omega_1(5\times2)$ and $\omega_2(6\times2)$ have $(n_1 = 5)$ and $(n_2 = 6)$ samples, respectively. Each sample in both classes is represented by two features as follows:

$$w_1 = \begin{pmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 3 \\ 4 & 5 \\ 5 & 5 \end{pmatrix} \text{ and } w_2 = \begin{pmatrix} 4 & 2 \\ 5 & 0 \\ 5 & 2 \\ 3 & 2 \\ 5 & 3 \\ 6 & 3 \end{pmatrix}$$

Calculate the lower dimension space using LDA.

LDA Example

Step 1: Compute the mean of each class (μ_1 and μ_2) and the global mean vector (μ).

The values of the mean of each class and the total mean are shown below:

$$\mu_1 = (3 \quad 3.6)$$

$$\mu_2 = (4.67 \quad 2)$$

$$\mu = (3.91 \quad 2.72)$$

LDA Example

Step 2: Compute the between-class scatter matrix:

The values of the between-class variance of the first class (SB_1) is equal to,

$$S_{B1} = n_1(\mu_1 - \mu)^T(\mu_1 - \mu) =$$

$$= 5(-0.91 \quad 0.87)^T(-0.91 \quad 0.87) = \begin{pmatrix} 4.13 & -3.97 \\ -3.97 & 3.81 \end{pmatrix}$$

The values of the between-class variance of the second class (SB_2) is

$$S_{B2} = n_2(\mu_2 - \mu)^T(\mu_2 - \mu) =$$

$$= 6(0.76 \quad -0.72)^T(0.76 \quad -0.72) = \begin{pmatrix} 3.44 & -3.31 \\ -3.31 & 3.17 \end{pmatrix}$$

$$S_B = S_{B1} + S_{B2} = \begin{pmatrix} 7.58 & -7.27 \\ -7.27 & 6.98 \end{pmatrix}$$

LDA Example

Step 3: Compute the within-class scatter matrix:

To calculate the within-class matrix, first subtract the mean of each class from each sample in that class, and it is calculated as follows, $d_i = \omega_i - \mu_i$. The values of d_1 and d_2 are as follows:

$$d_1 = w_1 - \mu_1 = \begin{pmatrix} -2 & -1.6 \\ -1 & -0.6 \\ 0 & -0.6 \\ 1 & 1.4 \\ 2 & 1.4 \end{pmatrix} \text{ and } d_2 = w_2 - \mu_2 = \begin{pmatrix} -0.67 & 0 \\ 0.33 & -2 \\ 0.33 & 0 \\ -1.67 & 0 \\ 0.33 & 1 \\ 1.33 & 1 \end{pmatrix}$$

After centering the data, the within-class variance for each class (SW_j) is calculated as follows, $SW_j = d^T{}_j * d_j$,

LDA Example

The values of within- class matrix for each class and the total within-class matrix are as follows:

$$S_{w1} = \begin{pmatrix} -2 & -1.6 \\ -1 & -0.6 \\ 0 & -0.6 \\ 1 & 1.4 \\ 2 & 1.4 \end{pmatrix}^T \begin{pmatrix} -2 & -1.6 \\ -1 & -0.6 \\ 0 & -0.6 \\ 1 & 1.4 \\ 2 & 1.4 \end{pmatrix} = \begin{pmatrix} 10 & 8 \\ 8 & 7.2 \end{pmatrix}$$

$$S_{w2} = \begin{pmatrix} -0.67 & 0 \\ 0.33 & -2 \\ 0.33 & 0 \\ -1.67 & 0 \\ 0.33 & 1 \\ 1.33 & 1 \end{pmatrix}^T \begin{pmatrix} -0.67 & 0 \\ 0.33 & -2 \\ 0.33 & 0 \\ -1.67 & 0 \\ 0.33 & 1 \\ 1.33 & 1 \end{pmatrix} = \begin{pmatrix} 5.33 & 1 \\ 1 & 6 \end{pmatrix}$$

$$S_W = S_{W1} + S_{W2} = \begin{pmatrix} 15.33 & 9 \\ 9 & 13.2 \end{pmatrix}$$

LDA Example

Step 4: Compute the final scatter matrix given by ($S_W^{-1}S_B$)

$$S_w^{-1} = \frac{1}{121.356}\begin{pmatrix} 13.2 & -9 \\ -9 & 15.33 \end{pmatrix} = \begin{pmatrix} 0.11 & -0.07 \\ -0.07 & 0.13 \end{pmatrix}$$

$$S_w^{-1}S_B = \begin{pmatrix} 0.11 & -0.07 \\ -0.07 & 0.13 \end{pmatrix}\begin{pmatrix} 7.58 & -7.27 \\ -7.27 & 6.98 \end{pmatrix} = \begin{pmatrix} 1.36 & -1.31 \\ -1.48 & 1.42 \end{pmatrix}$$

Step 5: Compute the eigen values for the final scatter matrix.

i.e. | $S_w^{-1}S_B$-λI|=0 i.e. $\begin{vmatrix} 1.36 - \lambda & -1.31 \\ -1.48 & 1.42 - \lambda \end{vmatrix} = 0$

Thus the eigen values for the above characteristic equations are: 2.78, 0.0027

LDA Example

Step 6: Ignore the second eigen value as it is nearer to zero. So the only selected eigen value is 2.78

So for λ=2.78; $[S_W^{-1}S_B$-2.78I]V_1=0

$$\begin{pmatrix} -1.42 & -1.31 \\ -1.48 & -1.36 \end{pmatrix} V_1=0$$

Therefore eigen vector is $V_1 = \begin{pmatrix} 0.68 \\ -0.74 \end{pmatrix}$

LDA Example

Step 7: The original data is projected on the lower dimensional space, as follows, $y_i = \omega_i V_1$,

$$y_1 = w_1 V_1 = \begin{pmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 3 \\ 4 & 5 \\ 5 & 5 \end{pmatrix} \begin{pmatrix} 0.68 \\ -0.74 \end{pmatrix} = \begin{pmatrix} -0.79 \\ -0.85 \\ -0.18 \\ -0.97 \\ -0.29 \end{pmatrix}$$

$$y_2 = w_2 V_1 = \begin{pmatrix} 4 & 2 \\ 5 & 0 \\ 5 & 2 \\ 3 & 2 \\ 5 & 3 \\ 6 & 3 \end{pmatrix} \begin{pmatrix} 0.68 \\ -0.74 \end{pmatrix} = \begin{pmatrix} 1.24 \\ 3.39 \\ 1.92 \\ 0.56 \\ 1.18 \\ 1.86 \end{pmatrix}$$