# DIMENSIONALITY REDUCTION

SHARON CHRISTABEL
Student id : 22021030

October 2022

## 1 Introduction

Datasets are the inevitable part of machine learning projects. The count of input features or variables of a dataset are generally defined as dimensionality. A dataset may contain large amount of features and these features makes the model more complicated. Inorder to reduce the set of features we use Dimensionality Reduction. Two methods for dimensionality reduction are:

1. Singular Vector Decomposition (SVD)
2. Linear Discriminant Analysis (LDA)

## 2 Singular Vector Decomposition

SVD divides a matrix A into two unitary matrices which are orthogonal in nature and a rectangular diagonal matrix of singular values. Mathematical representation of SVD :

$$A_{m*n} = U_{m*m}\Sigma_{m*n}V_{n*n}^T \tag{1}$$

Where

- A implies a matrix of order $m * n$
- U implies an orthogonal matrix of order $m * m$
- $\Sigma$ implies a diagonal matrix of order $m*$
- $V^T$ implies an orthogonal matrix of order $n * n$

$$U.U^T = I \tag{2}$$
$$V.V^T = I \tag{3}$$

Calculating $A^T.A$,

$$A^T.A = (V\Sigma^T U^T)(U\Sigma V^T)$$

Substituting equation (2) we obtain,

$$A^T.A = V\Sigma^T\Sigma V^T$$

Similarly we can calculate $A.A^T$,

$$A.A^T = (U\Sigma V^T)(V\Sigma^T U^T)$$

Substituting equation (3) we obtain,

$$A.A^T = U\Sigma\Sigma^T U^T$$

### 2.1 Steps to Perform Singular Vector Decomposition

Two geometrical operations are performed in SVD ; Rotation and Stretch. Rotation is performed on two orthogonal matrices U and V. Stretching is performed on diagonal matrix $\Sigma$.

Step 1 : From a given matrix, we need to find two orthogonal matrices and one diagonal matrix. Matrix U can be calculated using the equation $A^T.A$. And V can be calculated using the equation $A.A^T$. So at first find the transpose of a given matrix.

Step 2 : Then calculate $A^T.A$. Find the eigen values of the matrix and calculate the eigen vectors.

Step 3 : Perform orthogonalization of matrix. We need two matrices in orthogonal form. Inorder to do orthogonalization we use Gram-Schmidth method. Thus we obtain the first orthogonal matrix U.

Step 4 : Calculate $A^T.A$. Find the eigen values and eigen vectors.

Step 5 : For each eigen values obtained, apply crammer's rule. This will produce an orthogonal matrix V. Then find its transpose and thus $V^T$ is obtained.

Step 6 : To obtain $\Sigma$ we create a singular matrix with diagonal elements as square root of eigen values.

## 2.2 Example

Consider $2 * 3$ matrix

$$A = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$$

$$A^T = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix}$$

$$A^T.A = V$$

$$A.A^T = U$$

$$A.A^T = \begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix}$$

Find the eigen values,

$$(A - \lambda I) = 0$$

$$\begin{bmatrix} (11-\lambda) & 1 \\ 1 & (11-\lambda) \end{bmatrix} = 0$$

Solving the equation,

$$(11-\lambda)^2 - 1^2 = 0$$

$$(11 - \lambda + 1)(11 - \lambda - 1) = 0$$

$$(12 - \lambda) = 0$$

$$(10 - \lambda) = 0$$

$$\lambda_1 = 12, \lambda_2 = 10$$

Eigen vectors,

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Square roots of row elements,

$$\sqrt{1^2 + 1^2} = \sqrt{2}$$

$$\sqrt{1^2 + (-1)^2} = \sqrt{2}$$

The orthogonal matrix U will be,

$$U = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

Calculate A$^T.A$,

$$A^T.A = \begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix}$$

Find the eigen values,

$$(A - \lambda I)X = 0$$

$$\begin{bmatrix} (10-\lambda) & 0 & 2 \\ 0 & (10-\lambda) & 4 \\ 2 & 4 & (2-\lambda) \end{bmatrix}$$

To solve this we have an equation,

$$\lambda^3 - S_1\lambda^2 + S_2\lambda - S_3$$

Where

- $S_1$ implies the trace of matrix A
- $S_2$ implies minor of diagonals
- $S_3$ implies det(A)

$S_1 = 22$ $S_2 = 4 + 16 + 100 = 120$ $S_3 = 0$

Therefore,

$$\lambda^3 - 22\lambda^2 + 120 = 0$$

By solving this equation we obtain,

$$\lambda_1 = 12 \ \lambda_2 = 10 \ \lambda_3 = 0$$

Substituting the $\lambda$ values in the matrix,

$\lambda = 12$

$$\begin{bmatrix} -2 & 0 & 2 \\ 0 & -2 & 4 \\ 2 & 4 & -10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Calculate the value of $x_1, x_2, x_3$ using Cramer's rule,

$$\frac{x_1}{4} = \frac{-x_2}{-8} = \frac{x_3}{4}$$

Simplyfying this we get, $x_1 = 1, x_2 = 2, x_3 = 1$

Similarly, calculate $x_1, x_2, x_3$ values for $\lambda_2$ and $\lambda_3$ using Cramer's rule.We obtain matrix V,

$$V = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 2 \\ 1 & 0 & -5 \end{bmatrix}$$

Find the transpose of matrics V and orthogonalize it using Gram-Schmidth process,

$$V^T = \begin{bmatrix} 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \\ 2/\sqrt{5} & 1/\sqrt{5} & 0 \\ 1/\sqrt{30} & 2/\sqrt{30} & -5/\sqrt{30} \end{bmatrix}$$

Singular matrix $\Sigma$ will have the same order as that of matrix A. $\Sigma$ matrix can be calculated by placing square root of eigen values as diagonal elements.

$$\Sigma = \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{bmatrix}$$

Thus SVD decomposes the matrix A into three matrices,

$$A = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \\ 2/\sqrt{5} & 1/\sqrt{5} & 0 \\ 1/\sqrt{30} & 2/\sqrt{30} & -5/\sqrt{30} \end{bmatrix}$$

# 3 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is one of the dimensionality reduction technique mainly used for pattern classification and various other machine learning applications. It maximizes the seperation between multiple classes. In LDA an N-dimensional feature space is projected onto a smaller subspace K such that maintaining the class dicriminatory information.

## 3.1 Steps to Perform Linear Discriminant Analysis

Consider a 2D dataset,

- Step 1 : Calculate within-class scatter matrix $S_w$

$$S_w = S_1 + S_2$$

$S_1$= Covariance matrix for class 1

$$S_1 = \Sigma(x - \mu_1)(x - \mu_1)^T$$

$S_2$= Covariance matrix for class 2

$$S_2 = \Sigma(x - \mu_2)(x - \mu_2)^T$$

$\mu_1$ and $\mu_2$ are the mean of class 1 and class 2

- Step 2 : Calculate between class scatter matrix $S_B$

$$S_B=(\mu_1-\mu_2)(\mu_1-\mu_2)^T$$

- Step 3 : Find the best LDA projection vector.

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = S_w^- 1(\mu_1-\mu_2)$$

Finally dimension reduction is done using the equation $Y = W^T X$

$W^T$ = Projection vector

X = Input data samples

## 3.2 Example

Consider a 2D dataset

$C_1 = (4,1),(2,4),(2,3),(3,6),(4,4)$

$C_2 = (9,10),(6,8),(9,5),(8,7),(10,8)$

Find the mean $\mu_1 and \mu_2$ :

$$\mu_1 = \frac{4+2+2+3+4}{5}, \frac{1+4+3+6+4}{5}$$

$$\mu_2 = \frac{9+6+9+8+10}{5}, \frac{10+8+5+7+8}{5}$$

$$\mu_1 = (3, 3.60)$$

$$\mu_2 = (8.4, 7.60)$$

Calculate $(\mathrm{x}_1 - \mu_1)$ :

$$(\mathrm{x}_1 - \mu_1) = \begin{bmatrix} 1 & -1 & -1 & 0 & 1 \\ -2.6 & 0.4 & -0.6 & 2.4 & 0.4 \end{bmatrix}$$

]

Calculate covariance matrix for class 1, $S_1 = \Sigma(x - \mu_1)(x - \mu_1)^T$

Here we obtain 5 matrices. The first matrix would be :

$$\begin{bmatrix} 1 \\ -2.6 \end{bmatrix} * \begin{bmatrix} 1 & -2.6 \end{bmatrix} = \begin{bmatrix} 1 & -2.6 \\ -2.6 & 7.6 \end{bmatrix}$$

The other four matrices would be :

$$\begin{bmatrix} 1 & -0.4 \\ -0.4 & 0.16 \end{bmatrix}, \begin{bmatrix} 1 & 0.6 \\ 0.6 & 0.36 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 5.76 \end{bmatrix}, \begin{bmatrix} 1 & 0.4 \\ 0.4 & 0.16 \end{bmatrix}$$

Find the average of all the matrices, we obtain covariance matrix of class 1,$S_1$ :

$$\mathrm{S}_1 = \begin{bmatrix} 0.8 & -0.4 \\ -0.4 & 2.6 \end{bmatrix}$$

Similarly,

$$S_2 = \begin{bmatrix} 1.84 & -0.04 \\ -0.04 & 2.64 \end{bmatrix}$$

$$S_w = S_1 + S_2$$

$$S_w = \begin{bmatrix} 2.64 & -0.44 \\ -0.44 & 5.28 \end{bmatrix}$$

Calculating between class scatter matrix $S_B$

$S_B$=$(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$

$$S_B = \begin{bmatrix} 29.16 & 21.6 \\ 21.6 & 16.00 \end{bmatrix}$$

Finding the best LDA projection vector,

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} 0.91 \\ 0.39 \end{bmatrix}$$