MACHINE LEARNING ASSIGNMENT

DHEERAJ S NAIR
M.Tech CSE (DS and AI)

October 2022

1 Introduction

In real life applications of machine learning large datasets are used. Solving such large datasets using normal techniques could be difficult. So we have to reduce its size to the level we can handle. That is were dimensionality reduction techniques comes into action. Dimensionality reduction is the technique of reducing the dimensions of a matrix, in this case data feature set. By using these methods large datsets can be reduced to workable forms. There are several dimensionality reduction techniques used in machine learning. Principal Component Analysis(PCA) being one of the most used dimensionality redution technique.

Here we are going to discuss about two other commonly used techniques. Linear Discriminant Analysis(LDA) and Singular Value Decomposition(SVD).

2 Linear Discriminant Analysis(LDA)

Linear Discriminant Analysis, also known as Normal Discriminant Analysis which is a commonly used dimensionality reduction technique. It's major applications are in Supervised learning classification, were we have labeled input data. While logarithmic regression techniques are efficient in case of classification between two groups of data, it is not a desirable method for applications were seperating two or more than two groups are required. That is were LDA comes into action.

2.1 Method

In LDA, an n-dimensional data is converted into 1-D. It is done by projecting the data to a single line or axis position in such a way that the overlapping is the least, the characteristics of the data are preserved, yet the data is classified efficiently.

Let's take the case of a 2-D data as depicted in the figure. Here, an axis is drawn through the graph, to which the projections of the datapoints are taken.

The axis is chosen in such a way that the:-

- The means of each classes are at the max distances from each other.
- The variance of datapoints within each class is the least.

The idea behind this method is actually based on the Fishcer's Linear Discriminant.

2.2 Example

Let's take a 2-D dataset

$$C_1 \rightarrow X_1 = (X_1, X_2) = \{(4,1), (2,4), (2,3), (3,6), (4,4)\}$$

$$C_2 \rightarrow X_1 = (X_1, X_2) = \{(9,10),(6,8),(9,5),(8,7),(10,8)\}$$

STEP1:

Compute within class scatter matrix(s_W)

$S_w = S_1 + S_2$

$S_1 =$ co variance matrix of class $_1$

$S_2 =$ co variance matrix of class $_2$

$S_1 = \sum_{x \in C_1} (x - \mu_1)(x - \mu_2)^T$

$\mu_1 =$ Mean class of C_1

x $= Data present in C_1$

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

$\mu_1 = [3.00, 3.60]$

Similarly, $\mu_2\bar{[}8.2\ ,7.60]$

Mean reduced data,

$$[x_1 - \mu_1] \ = \begin{bmatrix} 1 & -1 & -1 & 0 & 1 \\ -2.6 & 0.4 & -0.6 & 2.4 & 0.4 \end{bmatrix}$$

Now for each x we are going to calculate,

$$(x - \mu_1)(x - \mu_1)^T$$

so we will have 5 such matrices.

1) $\begin{bmatrix} 1 \\ -2.6 \end{bmatrix} * \begin{bmatrix} 1 & -2.6 \end{bmatrix} = \begin{bmatrix} 1 & -2.6 \\ -2.6 & 6.76 \end{bmatrix}$

2) $\begin{bmatrix}-1\\0.4\end{bmatrix} * \begin{bmatrix}-1 & 0.4\end{bmatrix} = \begin{bmatrix}1 & -0.4\\-0.4 & 0.16\end{bmatrix}$

3) $\begin{bmatrix}-1\\0.6\end{bmatrix} * \begin{bmatrix}-1 & -0.6\end{bmatrix} = \begin{bmatrix}1 & 0.6\\0.6 & 0.36\end{bmatrix}$

4) $\begin{bmatrix}0\\2.4\end{bmatrix} * \begin{bmatrix}0 & -2.4\end{bmatrix} = \begin{bmatrix}0 & 0\\0 & 5.76\end{bmatrix}$

5) $\begin{bmatrix}1\\0.4\end{bmatrix} * \begin{bmatrix}1 & 0.4\end{bmatrix} = \begin{bmatrix}1 & 0.4\\0.4 & 0.16\end{bmatrix}$

Adding these equations and taking average get co variance of S_1

$S_1 = \begin{bmatrix}0.8 & -0.4\\-0.4 & 2.6\end{bmatrix}$

Similarly for the class 2 the co variance matrix is given by,

$S_2 = \begin{bmatrix}2.6 & -0.04\\-0.04 & 2.64\end{bmatrix}$

$S_w = \begin{bmatrix}2.6 & -0.04\\-0.44 & 5.28\end{bmatrix}$

STEP2:

Computing between class scatter matrix

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

$= \begin{bmatrix}29.16 & 21.6\\21.6 & 16.0\end{bmatrix}$

STEP3:

Find the best LDA projection vector similar to principal component analysis.we find this using eigen vector having largest eigen value.

$$S_w^{-1} * S_b V = \lambda V$$

$\begin{bmatrix} S_W^{-1} & S_b - \lambda I \end{bmatrix}$=$\begin{bmatrix} 11.89-\lambda & 8.81 \\ 5.08 & 3.76-\lambda \end{bmatrix}$=0

solving we get $\lambda = 15.65$

$substituting \lambda$in equation we get,

$\begin{bmatrix} V1 \\ V2 \end{bmatrix} = \begin{bmatrix} 0.91 \\ 0.34 \end{bmatrix}$

we get directly solve, $\begin{bmatrix} V1 \\ V2 \end{bmatrix}$ = $\mathrm{S}_w^{-1}(\mu_1 - \mu_2)$

$S_w^{-1} = \begin{bmatrix} 0.384 & 0.032 \\ 0.032 & 0.192 \end{bmatrix}$

$STEP4:$

Dimension reduction

Y=W$^T X \to (Input)$

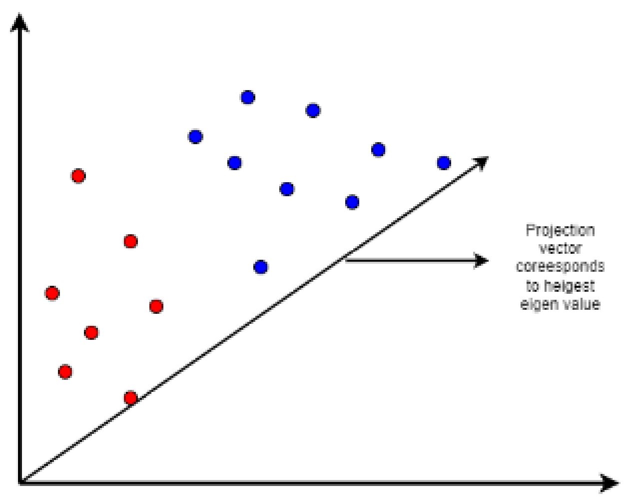

Figure 1:

3 Singular Value Decomposition(SVD)

Singular Value Decomposition matrix is the dimensionality reduction technique in which a given matrix or set of data, split into 3 matrices. Two unitary column orthogonal matrices and one rectangular singularity matrix.

3.1 Method

As we have said earlier, in SVD a matrix is split or factorised into three matrices. Two unitary orthogonal matrices and one rectangular singularity matrix. It is denoted as,

$$A = U \times S \times V^T \tag{3}$$

where,
$\bar{\text{A}}$= The Given Matrix of order mxn
$\bar{\text{U}}$= Right orthogonal Matrix of order mxm
$\bar{\text{S}}$= Singularity Matrix of order mxn
$\bar{\text{V}}^T$= Transpose of Left orthogonal Matrix V of order nxn

As we said both U and V matrices are column orthogonal matrices, means their are eigenvectors. Columns of V are eigenvectors of A^T.A and columns of U are eigenvectors of matrix A.A^T. The matrix S is a singularity matrix, means its a diagonal matrix with non negative real values known as singular values arranged in a higher to lower order of matrix A. Singular values are found from the A^T.A matrix. The singular values are the real non negative square root of eigenvalues of the matrix A^T.A.

3.2 Example

We have given matrix A and we want to represent it as the product of three matrices.

$$\text{A} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix}$$

In matrix A take small singular value and set it to 0 we also taking the third column of U and third row of V^T and set them to zero.

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix} = \begin{bmatrix} 0.13 & 0.02 \\ 0.41 & 0.07 \\ 0.55 & 0.99 \\ 0.68 & 0.11 \\ 0.15 & -0.59 \\ 0.07 & -0.73 \\ 0.07 & -0.29 \end{bmatrix} * \begin{bmatrix} 12.4 & 0 \\ 0 & 9.5 \end{bmatrix} * \begin{bmatrix} 0.56 & 0.59 & 0.56 & 0.09 & 0.09 \\ 0.12 & -0.02 & 0.12 & -0.69 & 0.69 \end{bmatrix}$$

After multiplying we can see that the given matrix A is approximately equivalent to the product.

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix} \approx \begin{bmatrix} 0.92 & 0.95 & 0.92 & 0.01 & 0.01 \\ 2.91 & 3.01 & 2.91 & -0.01 & -0.01 \\ 3.90 & 4.04 & 3.90 & 0.01 & 0.01 \\ 4.82 & 5.0 & 4.82 & 0.03 & 0.03 \\ 0.70 & 0.53 & 0.70 & 4.11 & 4.11 \\ -0.06 & 1.34 & -0.69 & 4.78 & 4.78 \\ 0.32 & 0.23 & 0.32 & 2.01 & 2.01 \end{bmatrix}$$

The matrix can thus be split as,

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix} = \begin{bmatrix} 0.13 & 0.02 & -0.01 \\ 0.41 & 0.07 & -0.03 \\ 0.55 & 0.99 & -0.04 \\ 0.68 & 0.11 & -0.05 \\ 0.15 & -0.59 & 0.65 \\ 0.07 & -0.73 & -0.67 \\ 0.07 & -0.29 & 0.32 \end{bmatrix} * \begin{bmatrix} 12.4 & 0 & 0 \\ 0 & 9.5 & 0 \\ 0 & 0 & 1.3 \end{bmatrix} * \begin{bmatrix} 0.56 & 0.59 & 0.56 & 0.09 & 0.09 \\ 0.12 & -0.02 & 0.12 & -0.69 & 0.69 \\ 0.40 & -0.80 & 0.40 & 0.09 & 0.09 \end{bmatrix}$$