Dimensionality Reduction

Avishna T

Student Id:22021029

October 2022

1 Introduction

Machine learning is a emerging technology and it is used for various purposes.It can also manipulate a large amount of data and analyze them for making different models.when dealing with huge amount of data,it may come across some difficulties in processing them.various techiniques can be use to reduce these problems.Data dimensionality reduction is one among them.Dimensionality reduction refers to the method of decreasing the number of input variables for a predictive model so that we get optimum accuracy for the model with minimal features.

Two popular dimensionality reduction methods are:

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

2 Singular Vector Decomposition (SVD)

What SVD is trying to do is to gives the best axis to project on.Best means,the sum of the squared projection errors is minimized.In some sense we want a small set of axis such that if we represent our data in terms of that axis,we get the minimum reconstruction error.

simple example how to see this would be in this two dimensionally,where imagine every different axis is a separate movie and we have users ranking movies and every point is now a user and the x position of this point is how much they rated movie-one and y position is how much they rated movie2.and we want to find the best line for the projection of this data points.

If red line is the projection line,we can represent every data points using a set of numbers, which is simply the position /projection of a given data point on this line.what SVD do is ,it try to minimize the sum of the squared errors of the distance between the original position of the data point and the current position in the line.

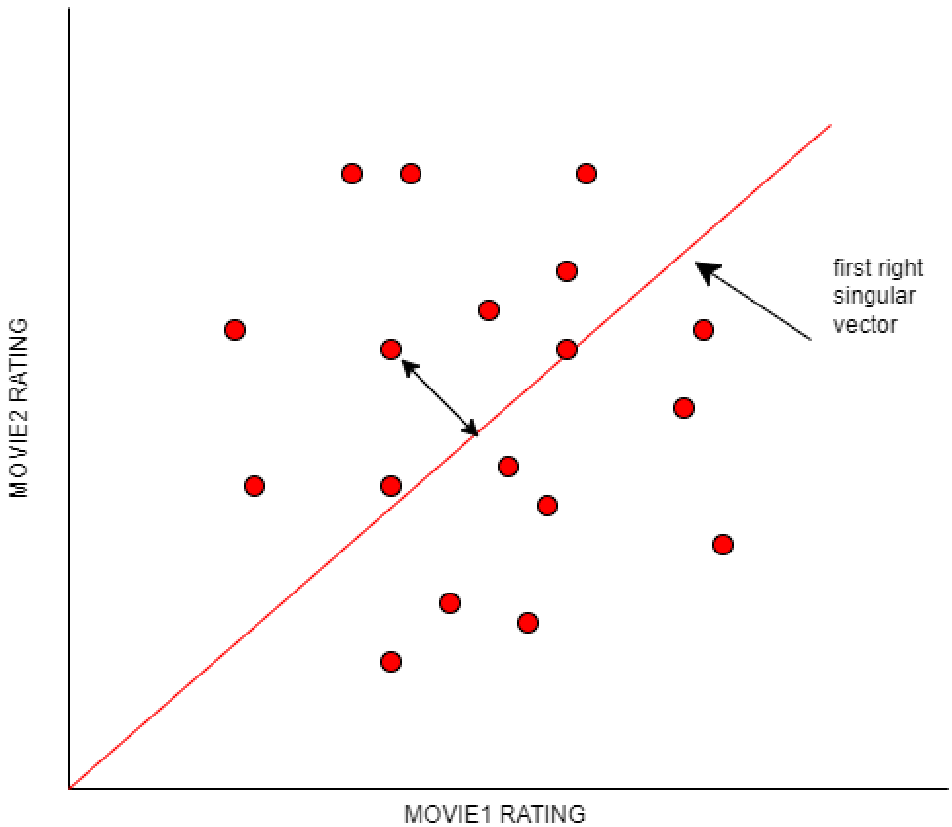

Figure 1:

So SVD finds the best vectors or axis on which to project the data such a the reconstruction error is minimized

3 Example

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

$$A = U\Sigma V^T \tag{1}$$

where

- A is a $m * n$ matrix
- U is an orthogonal $m * k$ matrix
- Σ is a $k * k$ diagonal matrix
- V^T is an orthogonal $k * n$ matrix

ie. V=Concept matrix and U =User-concept matrix

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

The first right singular vector gives us the axis on which we want to project and the corresponding singular value tells us the variance of the dimension. The variance is spread around the first singular vector (say 12.4 in this example) and they spread around a bit the second singular vector a bit less and they are not really spread around the third singular vector

3.1 How it is done??

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$^{T} andsetthemtozero.$

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

After multiplying we know that A and A1 is similar. difference between corresponding variables is small.

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

4 Linear Discriminant Analysis (LDA)

It's a dimensionality reduction technique .Here the dimension means features of the data.When we have large number of input data with us and each data is going to have a large features,then machine learning algorithm may not be able to tackle it.so we need to reduce the number of dimensions. The main goal of LDA is to project a feature space that is a N-dimension data in to a smaller subspace K (k¡=N-1) by maintaining the category information

4.1 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 $= Datapresentin C_1$

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

$\mu_1 = [3.00, 3.60]$

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,

$\mathrm{S}_2 = \begin{bmatrix} 2.6 & -0.04 \\ -0.04 & 2.64 \end{bmatrix}$

$\mathrm{S}_w = \begin{bmatrix} 2.6 & -0.04 \\ -0.44 & 5.28 \end{bmatrix}$

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, $\begin{bmatrix} V1 \\ V2 \end{bmatrix} = S_w^{-1}(\mu_1 - \mu_2)$

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

$STEP4:$

Dimension reduction

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

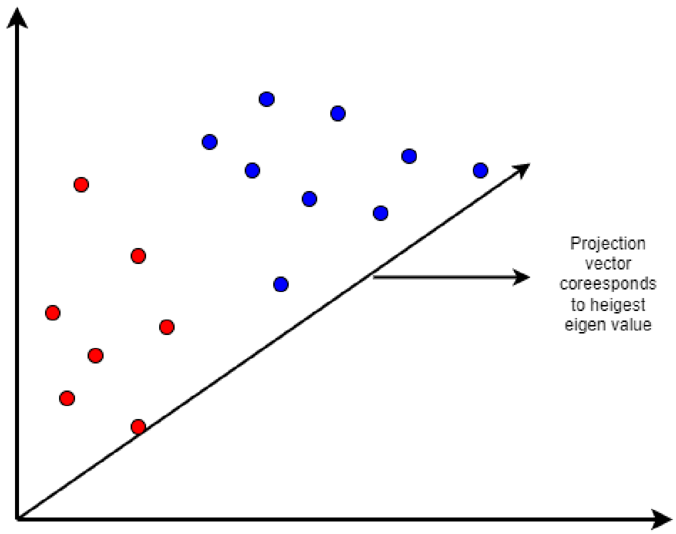

Figure 2:

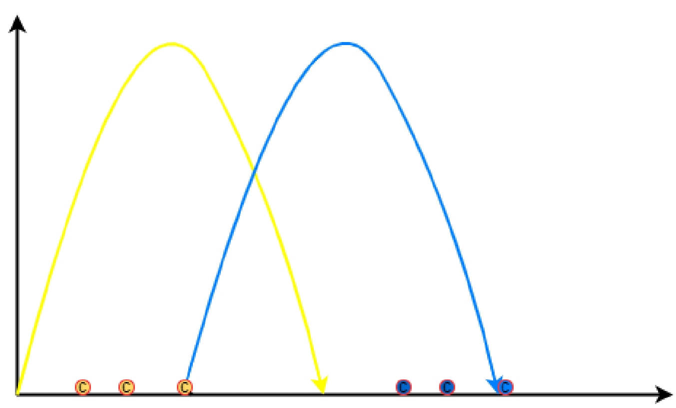

Figure 3: