Linear Discriminant Analysis
&
Singular Values Decomposition

Ranjima Mohan P

January 12, 2023

1 LINEAR DISCRIMINANT ANALYSIS (LDA)

There are various dimensionality reduction techniques used for generalization in Machine Learning. LDA is one such supervised learning algorithm which is actually a pre-processing step for pattern classification and machine learning applications. It is similar to Principal Component Analysis but the difference is that, LDA in addition, maximize the separation between multiple classes, i.e, LDA provides maximum class separability. It also reduces the number of features and while doing so the features are projected in such a manner that the separation between each class to which feature belongs to is maximized.

The goal of LDA is to project the feature space (N-dimensional data) onto a smaller subspace K while maintaining the class discriminating information about the features as well (N is the number of features present in the data).

We starts by considering the case when only two classes are there. Then we generalize to more than two classes, ie, K>2. Suppose there are samples from two classes C_1 and C_2.The samples should be well separated from each other when they are projected onto the direction vector w.

1.1 Step by step procedure with an example problem

Let's take a 2D dataset (considering 2 classes, C_1 and C_2

$$C_1 = \{(4,1) \quad (2,4) \quad (2,3) \quad (3,6) \quad (4,4)\}$$
$$C_2 = \{(9,10) \quad (6,8) \quad (9,5) \quad (8,7) \quad (10,8)\}$$

Step 1: Compute within-class Scatter Matrix, S_W

$$S_W = S_1 + S_2$$

S_1 is the Covariance Matrix for the class C_1 and S_2 is the Covariance Matrix for the class C_2
So, let's now find S_1 and S_2.

$$S_1 = \sum_{x \varepsilon C_1} \left(x - \mu_1\right)\left(x - \mu_1\right)^T \tag{1.1}$$

where, μ_1 is the mean of class C_1, which is computed by

$$\mu_1 = \left\{ \frac{4+2+2+3+4}{5} \quad \frac{1+4+3+6+4}{5} \right\} \tag{1.2}$$

$$\mu_1 = \begin{bmatrix} 3.00 & 3.60 \end{bmatrix} \tag{1.3}$$

Similarly,

$$\mu_2 = \begin{bmatrix} 8.4 & 7.60 \end{bmatrix}$$

Applying the value of μ_1 in equation no.1.1. Then,

$$\begin{bmatrix} x_1 - \mu_1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 & 0 & 1 \\ -2.6 & 0.4 & -0.6 & 2.4 & 0.4 \end{bmatrix} \tag{1.4}$$

Now, for each x, we need to calculate $\left(x - \mu_1\right)\left(x - \mu_1\right)^T$.

So we will be having 5 such matrices.
We will go one by one.

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

Similarly,

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

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

$$\begin{bmatrix} 0 \\ 2.4 \end{bmatrix} \begin{bmatrix} 0 & 2.4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 5.76 \end{bmatrix} \Rightarrow ④$$

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

Adding ① + ② + ③ + ④ + ⑤ and taking the average (i.e, divided by 5)

Then,

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

Similarly for the class C_2, the Covariance Matrix is given by

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

We have $S_W = S_1 + S_2$

Hence, $S_W = \begin{bmatrix} 2.64 & -0.44 \\ -0.44 & 5.28 \end{bmatrix}$

Step 2: Compute the Between Class Matrix S_B

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

$S_B = \begin{pmatrix} -5.4 \\ -4 \end{pmatrix} \begin{pmatrix} -5.4 & -4 \end{pmatrix}$

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

Step 3: Find the best LDA projection vector.

Similar to Principal Component Analysis we find this using eigen vectors having largest eigen value.

$S_W{}^{-1} S_B V = \lambda V \Rightarrow$ ⓐ

$\left| S_W^{-1} \quad S_B - \lambda I \right|$=0

$\begin{vmatrix} 11.89 - \lambda & 8.81 \\ 5.08 & 3.76 - \lambda \end{vmatrix}$ =0

By solving for λ, we get $\lambda = 15.65$

Substitute λ in equation ⓐ,

$\begin{bmatrix} 11.89 & 8.81 \\ 5.08 & 3.76 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$=15.65$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$

Then we get

$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0.91 \\ 0.39 \end{bmatrix}$

Step 4: Dimension reduction

$$Y = W^T X \tag{1.5}$$

where,

X is the input data sample

W^T is the Projection Vector

2 SINGULAR VALUES DECOMPOSITION (SVD)

Singular Value Decomposition is another important dimensionality reduction technique. It is a method of decomposing a rectangular matrix into three matrices; two unit matrices which are orthogonal and a rectangular diagonal matrix of singular value.

If A is the input matrix of order m× n,

$$A = u \sum v^T \tag{2.1}$$

The matrix A can be written as the product of three matrices u, $\sum$ and

- u : m× r, the left singular vectors

$$\begin{bmatrix} u1 & u2 & . \\ . & . & . \\ . & . & . \end{bmatrix}$$

- $\sum$: r× r, the singular values diagonal matrix

$$\begin{bmatrix} \sigma 1 & 0 & . \\ 0 & \sigma 2 & . \\ . & . & . \end{bmatrix}$$

- v : n× r, the right singular vectors

$$\begin{bmatrix} v1 & . & . \\ v2 & . & . \\ . & . & . \end{bmatrix}$$

What are orthogonal matrix and diagonal matrix?

- A matrix A is called orthogonal matrix if

$$AA^T = A^T A = I \tag{2.2}$$

and a diagonal matrix has non-zero values in the diagonal and rest filled with zeros.

2.1 Properties of SVD

- The decomposed matrices are unique.
- The columns of u and v have Euclidean length '1' and the sum of squared values of each matrix is '1',i.e, u and v are column orthonormal.
- The columns of u and v are orthogonal,i.e, if we take any two columns of u and v, their dot product equals zero.
- The values in $\sum$ matrices are called singular values and they are positive and sorted in decreasing order ($\sigma_1 \geq \sigma_2 \geq$ .... $\geq$0)

2.2 The motive behind SVD :

By decomposing such a data matrix, it is possible to discover latent, hidden features which could help in classification or clustering tasks.

Let's consider a term document matrix where there are 'm' documents and 'n' terms in it, i.e, order (m×n)

2.3 Example Problem

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

Now we are going to decompose the matrix A into three matrices u, $\sum$,and v^T.

- Finding AA^T

$$AA^T = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}\begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix}$$

$|AA^T - \lambda I|$ =0

On solving we get λ values as

λ_1= 12 and λ_2= 10

For λ = **12**,

We will get the Eigen vector as $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Then the unit Eigen vector = $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$

For λ = **10**,

We will get the Eigen vector as $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$

Then the unit eigen vector = $\begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$

Therefore, the matrix u= $\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$

- Finding A^TA

$$A^T A = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix}$$

$\left|A^T A - \lambda I\right| = 0$

On solving we get the λ values as
λ_1= 12
λ_2=10
λ_3= 0

For λ = 12,

We will get the Eigen vector as $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$

Then the unit eigen vector = $\begin{bmatrix} 1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$

For λ = 10,

We will get the Eigen vector as $\begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$

Then the unit eigen vector = $\begin{bmatrix} 2/\sqrt{5} \\ -1/\sqrt{5} \\ 0 \end{bmatrix}$

For λ = 0,

We will get the Eigen vector as $\begin{bmatrix} -1 \\ -2 \\ 5 \end{bmatrix}$

Then the unit eigen vector = $\begin{bmatrix} -1/\sqrt{30} \\ -2/\sqrt{30} \\ 5/\sqrt{30} \end{bmatrix}$

Therefore, the matrix v= $\begin{bmatrix} 1/\sqrt{6} & 2/\sqrt{5} & -1/\sqrt{30} \\ 2/\sqrt{6} & -1/\sqrt{5} & -2/\sqrt{30} \\ 1/\sqrt{6} & 0/\sqrt{5} & 5/\sqrt{30} \end{bmatrix}$

Taking the transpose of v,

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

The singular values are:

$\sigma 1 = \sqrt{12}$
$\sigma 2 = \sqrt{10}$
i.e, the square root of λ values.
Hence the matrix $\sum$ can be written as

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

Now we got the three matrices u, $\sum$, and v^T.
Let's rewrite A as the combination of these three matrices.

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

****