

# MACHINE LEARNING

ANAGHA RAJEEV

January 14, 2023

## 1 Linear Discriminant Analysis

In linear algebra, the Singular Value Decomposition (SVD) of a matrix is a factorization of that matrix into three matrices. It has some interesting algebraic properties and conveys important geometrical and theoretical insights about linear transformations. It also has some important applications in data science.

To understand SVD we need to first understand the Eigenvalue Decomposition of a matrix. We can think of a matrix  $A$  as a transformation that acts on a vector  $x$  by multiplication to produce a new vector  $Ax$ . We use  $[A]_{ij}$  or  $a_{ij}$  to denote the element of matrix  $A$  at row  $i$  and column  $j$ . If  $A$  is an  $m \times p$  matrix and  $B$  is a  $p \times n$  matrix, the matrix product  $C=AB$  (which is an  $m \times n$  matrix) is defined as:

$$[C]_{ij} = c_{ij} = \sum_{k=1}^p a_{ik} b_{kj}$$

Figure 1: Flock of birds.

For example, the rotation matrix in a 2-d space can be defined as:

Example Let us take a 2 Dimensional dataset: C1  $X1 = (x1, x2) = (4, 1), (2, 4), (2, 3), (3, 6), (4, 4)$   
C2  $X2 = (x1, x2) = (9, 10), (6, 8), (9, 5), (8, 7), (10, 8)$

### 1.1 Step1

Compute the within class scatter matrix  $S_w$   $S_w = S1 + S2$   $S1$  is the covariance matrix for the the class C1  $S2$  is the covariance matrix for the the class C2 Covariance matrix Covariance matrix =  $\text{var}(x) \text{ cov}(y, x) \text{ cov}(x, y) \text{ var}(y)$  To calculate the covariance matrix we need to find x variance, y variance and the covariance of x and y

### 1.2 Scatter matrix

A scatter plot matrix is a grid (or matrix) of scatter plots used to visualize bivariate relationships between combinations of variables. Each scatter plot in the matrix visualizes the relationship between a pair of variables, allowing many relationships to be explored in one chart. A scatter matrix is a

$$\text{VAR}(\mathbf{X}) = \frac{1}{n-1} \sum_{i=1}^n (x_i - E(X))^2$$

$$\text{COV}(\mathbf{X}, \mathbf{Y}) = \frac{1}{n-1} \sum_{i=1}^n (x_i - E(X))(y_i - E(Y))$$

$$\text{COR}(\mathbf{X}, \mathbf{Y}) = \frac{\text{COV}(\mathbf{X}, \mathbf{Y})}{\sqrt{\text{VAR}(\mathbf{X})\text{VAR}(\mathbf{Y})}}$$

$$R^2 = 1 - \frac{\text{VAR}(\mathbf{X}, \mathbf{Y})_{\text{FittedLine}}}{\text{VAR}(\mathbf{X}, \mathbf{Y})_{\text{Mean}}}$$

Figure 2: Flock of birds.

estimation of covariance matrix when covariance cannot be calculated or costly to calculate. The scatter matrix is also used in lot of dimensionality reduction exercises. If there are k variables, scatter matrix will have k rows and k columns i.e k X k matrix.

The scatter matrix is computed by the following equation:

$$S = \sum_{k=1}^n (\mathbf{x}_k - \mathbf{m})(\mathbf{x}_k - \mathbf{m})^T$$

where  $\mathbf{m}$  is the mean vector

$$\mathbf{m} = \frac{1}{n} \sum_{k=1}^n \mathbf{x}_k$$

Figure 3: Flock of birds.

### 1.3 In python scatter matrix can be computed using

The scatter matrix contains for each combination of the variable, the relation between them. Let's observe the scatter matrix for the following matrix

For k variables in the dataset, the scatter plot matrix contains k rows and k columns. Each row and column represents as a single scatter plot. Each individual plot (i, j) can be defined as:

Vertical Axis: Variable Xj Horizontal Axis: Variable Xi The plot lies on the diagonal is just a 45 line because we are plotting here Xi vs Xi. However, we can plot the histogram for the Xi in the diagonals or just leave it blank. Since Xi vs Xj is equivalent to Xj vs Xi with the axes reversed, we can also omit the plots below the diagonal. It can be more helpful if we overlay some line plot on the scattered points in the plots to give more understanding of the plot. The idea of the pair-wise plot can also be extended to different other plots such as quantile-quantile plots or bihistogram.

### 1.4 Dimension reduction

Dimension reduction In this step after obtaining the projection vector we use the input data to project the reduced data into a plot.  $y = V X \dots$

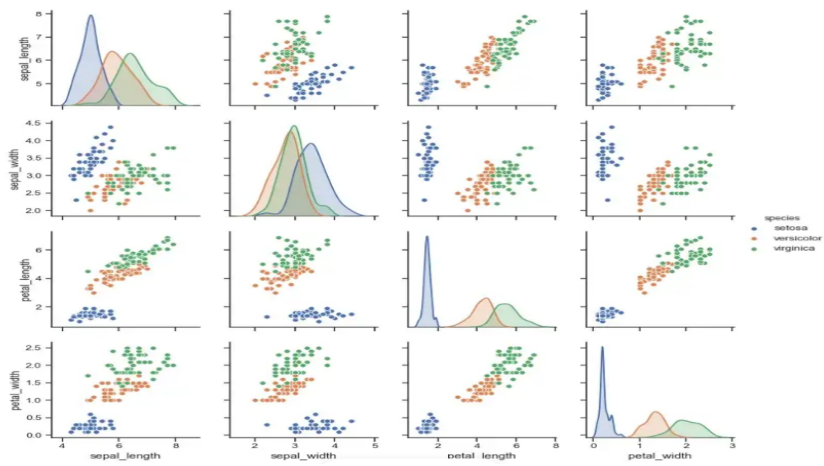


Figure 4: Flock of birds.

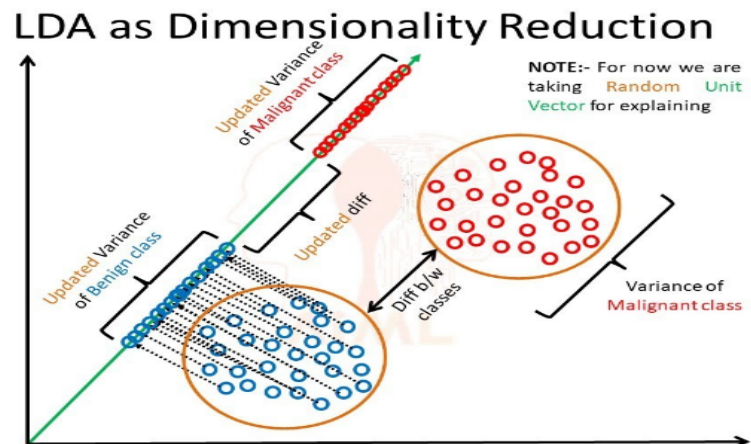


Figure 5: Flock of birds.

Samples for class 1 :  $X_1=(x_1,x_2)=(4,2),(2,4),(2,3),(3,6),(4,4)$  and Sample for class 2 :  $X_2=(x_1,x_2)=(9,10),(6,8),(9,5),(8,7)$   
STEP 1: Computing the d-dimensional mean vectors

$$\mu_1 = 1/N_1 \sum_{x \in w_1} x = 1/5 * \left[ \begin{bmatrix} 4 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 4 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 3 \\ 6 \end{bmatrix} + \begin{bmatrix} 4 \\ 4 \end{bmatrix} \right] = \begin{bmatrix} 3 \\ 3.8 \end{bmatrix}$$

$$\mu_2 = 1/N_2 \sum_{x \in w_2} x = 1/5 * \left[ \begin{bmatrix} 9 \\ 10 \end{bmatrix} + \begin{bmatrix} 6 \\ 8 \end{bmatrix} + \begin{bmatrix} 9 \\ 5 \end{bmatrix} + \begin{bmatrix} 8 \\ 7 \end{bmatrix} + \begin{bmatrix} 10 \\ 8 \end{bmatrix} \right] = \begin{bmatrix} 8.4 \\ 7.6 \end{bmatrix}$$

class means are:  $\bar{\mu}_1 = \begin{bmatrix} 3 \\ 3.8 \end{bmatrix}$  and  $\bar{\mu}_2 = \begin{bmatrix} 8.4 \\ 7.6 \end{bmatrix}$

STEP 2 : Computing the Scatter Matrices

within-class scatter matrix  $S_W$  is computed by the following equation  $S_W = \sum_{i=1}^c S_i$

where  $S_i = \sum_{x \in D_i} (x - \mu_i) - (x - \mu_i)^T$  Covariance matrix of the first class:

$$S_1 = \sum_{x \in w_1} (x - \mu_1) - (x - \mu_1)^T = \left[ \begin{bmatrix} 4 \\ 2 \end{bmatrix} - \begin{bmatrix} 3 \\ 3.8 \end{bmatrix} \right]^2 + \left[ \begin{bmatrix} 2 \\ 4 \end{bmatrix} - \begin{bmatrix} 3 \\ 3.8 \end{bmatrix} \right]^2 + \left[ \begin{bmatrix} 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 3 \\ 3.8 \end{bmatrix} \right]^2 + \left[ \begin{bmatrix} 3 \\ 6 \end{bmatrix} - \begin{bmatrix} 3 \\ 3.8 \end{bmatrix} \right]^2 + \left[ \begin{bmatrix} 4 \\ 4 \end{bmatrix} - \begin{bmatrix} 3 \\ 3.8 \end{bmatrix} \right]^2$$

$$= \begin{bmatrix} 1 & -0.25 \\ -0.25 & 2.2 \end{bmatrix}$$

covariance matrix of the first class  $S_1 = \text{cov}(X_1)$

$$S_2 = \sum_{x \in w_2} (x - \mu_2) - (x - \mu_2)^T = \left[ \begin{bmatrix} 9 \\ 10 \end{bmatrix} - \begin{bmatrix} 8.4 \\ 7.6 \end{bmatrix} \right]^2 + \left[ \begin{bmatrix} 6 \\ 8 \end{bmatrix} - \begin{bmatrix} 8.4 \\ 7.6 \end{bmatrix} \right]^2 + \left[ \begin{bmatrix} 9 \\ 5 \end{bmatrix} - \begin{bmatrix} 8.4 \\ 7.6 \end{bmatrix} \right]^2 + \left[ \begin{bmatrix} 8 \\ 7 \end{bmatrix} - \begin{bmatrix} 8.4 \\ 7.6 \end{bmatrix} \right]^2 + \left[ \begin{bmatrix} 10 \\ 8 \end{bmatrix} - \begin{bmatrix} 8.4 \\ 7.6 \end{bmatrix} \right]^2$$

$$= \begin{bmatrix} 2.3 & -0.05 \\ -0.05 & 3.3 \end{bmatrix}$$

covariance matrix of the first class  $S_2 = \text{cov}(X_2)$

Within - class scatter matrix :  $S_W = \sum_{i=1}^c S_i = S_1 + S_2 = \begin{bmatrix} 1 & -0.25 \\ -0.25 & 2.2 \end{bmatrix} + \begin{bmatrix} 2.3 & -0.05 \\ -0.05 & 3.3 \end{bmatrix} =$

$$\begin{bmatrix} 3.3 & -0.3 \\ -0.3 & 5.5 \end{bmatrix}$$

Within - class scatter matrix :  $S_W = \begin{bmatrix} 3.3 & -0.3 \\ -0.3 & 5.5 \end{bmatrix}$

Now compute Between - class scatter matrix :  $S_B$

$$S_B = (\mu_1 - \mu_2) * (\mu_1 - \mu_2)^T = \left[ \begin{bmatrix} 3 \\ 3.8 \end{bmatrix} - \begin{bmatrix} 8.4 \\ 7.6 \end{bmatrix} \right] * \left[ \begin{bmatrix} 3 \\ 3.8 \end{bmatrix} - \begin{bmatrix} 8.4 \\ 7.6 \end{bmatrix} \right]^T$$

$$= \begin{bmatrix} -5.4 \\ -3.8 \end{bmatrix} * \begin{bmatrix} -5.4 & -3.8 \end{bmatrix} = \begin{bmatrix} 29.16 & 20.52 \\ 20.52 & 14.44 \end{bmatrix}$$

Solving the generalized eigenvalue problem for the matrix  $S_W^{-1} * S_B$

The LDA projection is then obtained as the solution of the generalized eigen value problem

i.e,  $S_W^{-1} * S_B = \lambda_W$

$|S_W^{-1} * S_B - \lambda_W| = 0$

$$\begin{aligned}
&= \left| \begin{bmatrix} 3.3 & -0.3 \\ -0.3 & 5.5 \end{bmatrix}^{-1} * \begin{bmatrix} 29.16 & 20.52 \\ 20.52 & 14.44 \end{bmatrix} - \lambda * \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| \\
&= \left| \begin{bmatrix} 0.3045 & 0.0166 \\ 0.0166 & 0.1827 \end{bmatrix} * \begin{bmatrix} 29.16 & 20.52 \\ 20.52 & 14.44 \end{bmatrix} - \lambda * \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0 \\
&= \left| \begin{bmatrix} 9.2213 - \lambda & 6.483 \\ 4.2339 & 2.9794 - \lambda \end{bmatrix} \right| = 0 \\
&= (9.2213 - \lambda)(2.9794 - \lambda) - (6.483)(4.2339) = 0 \\
&= (\lambda)^2 - 12.2007 * \lambda = 0 \\
&\text{i.e., } \lambda_1 = 0 \text{ and } \lambda_2 = 12.0027
\end{aligned}$$

*STEP 4 : Selecting linear discriminant for the new feature subspace*

Hence,  $\lambda_1 = 0$

$$\begin{bmatrix} 9.2213 & 6.483 \\ 4.2339 & 2.9794 \end{bmatrix} * W_1 = \lambda_1 * \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 9.2213 & 6.483 \\ 4.2339 & 2.9794 \end{bmatrix} * W_1 = 0 * \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$\begin{aligned}
&\text{and } \begin{bmatrix} 9.2213 & 6.483 \\ 4.2339 & 2.9794 \end{bmatrix} * W_2 = \lambda_2 * \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\
&= \begin{bmatrix} 9.2213 & 6.483 \\ 4.2339 & 2.9794 \end{bmatrix} * W_1 = 12.0027 * \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}
\end{aligned}$$

$$\text{Thus } W_1 = \begin{bmatrix} -0.5755 \\ 0.8178 \end{bmatrix}$$

$$\text{and } \boxed{\bar{W}_2 = \begin{bmatrix} 0.9088 \\ 0.4173 \end{bmatrix} = W^*}$$

## 2 Scalar Value Decompositon

LT<sub>E</sub>X SVD is also a dimensional reduction technique. In this method the given matrix is decomposed into three distinct matrices. SVD of a m x n matrix is given as:  $A = UWV^T$  U : mxn matrix of the orthonormal eigenvectors of  $AA^T$  W : a nxn diagonal matrix of the singular values which are the square roots of the eigenvalues of  $AA^T$   $V^T$

denote their mean. Then as  $n$  approaches infinity, the random variables  $\sqrt{n}(S_n - \mu)$  converge in distribution to a normal  $\mathcal{N}(0, \sigma^2)$ .

$$AA^T = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix} * \begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 17 & 8 \\ 8 & 17 \end{bmatrix}$$

Applying the characteristic equation for the above matrix is  $W - \lambda I = 0$

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

$$\lambda^2 - 34\lambda + 225 = 0$$

$$= (\lambda - 25)(\lambda - 9)$$

so our singular values are:  $\sigma_1 = 5; \sigma_2 = 3$  Now we find the right singular vectors i.e orthonormal set of eigenvectors of  $AA^T$ . The eigenvalues of  $AA^T$  are 25, 9, and 0, and since  $AA^T$  is symmetric we know that the eigenvectors will be orthogonal.

For  $\lambda = 25$ ,

$$AA^T - 25I = \begin{bmatrix} -12 & 12 & 2 \\ 12 & -12 & -2 \\ 2 & -2 & -17 \end{bmatrix}$$

**Singular value decomposition** is an extension *eigenvalue decomposition* (see Fig. 6.2) to nonsquare matrices. For  $d \times m$  matrix  $A$ , a  $d$ -dimensional nonzero vector  $\psi$ , an  $m$ -dimensional nonzero vector  $\varphi$ , and a non-negative scalar  $\kappa$  such that

$$A\varphi = \kappa\psi$$

are called a *left singular vector*, a *right singular vector*, and a *singular value* of  $A$ , respectively. Generally, there exist  $c$  singular values  $\kappa_1, \dots, \kappa_c$ , where

$$c = \min(d, m).$$

Singular vectors  $\varphi_1, \dots, \varphi_c$  and  $\psi_1, \dots, \psi_c$  corresponding to singular values  $\kappa_1, \dots, \kappa_c$  are mutually orthogonal and are usually normalized, i.e., they are *orthonormal* as

$$\varphi_k^\top \varphi_{k'} = \begin{cases} 1 & (k = k') \\ 0 & (k \neq k') \end{cases} \quad \text{and} \quad \psi_k^\top \psi_{k'} = \begin{cases} 1 & (k = k') \\ 0 & (k \neq k'). \end{cases}$$

A matrix  $A$  can be expressed by using its singular vectors and singular values as

$$A = \sum_{k=1}^c \kappa_k \psi_k \varphi_k^\top.$$

In MATLAB, singular value decomposition can be performed by the `svd` function.

Figure 6: Flock of birds.

which can be row-reduces to :  $\begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

unit vector in the direction of it is:

$$v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}$$

$$\text{Similarly, for } \lambda = 9, v_2 = \begin{bmatrix} \frac{1}{\sqrt{18}} \\ \frac{-1}{\sqrt{18}} \\ \frac{4}{\sqrt{18}} \end{bmatrix}$$

For the 3rd eigenvector, we could use the property that it is perpendicular to  $v_1$  and  $v_2$  such that:

$$v_1^T v_3 = 0$$

$$v_2^T v_3 = 0$$

Solving the above equation to generate the third eigenvector

$$v_3 = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} a \\ -a \\ -a/2 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ \frac{-2}{3} \\ \frac{-1}{3} \end{bmatrix} \quad \text{Now, we calculate } U \text{ using the formula } u_i = \frac{1}{\sigma} A v_i \text{ and this gives } U$$

$$= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}.$$

Hence, our final SVD equation becomes:

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