# LDA and SVD (ML Assignment)

Nithin Kurian

October 2022

## 1 Introduction

In Machine Learning Feature Engineering is an important topic. Feature engineering can be defined as the process of selecting, processing, and converting the data to a form that is suitable for supervised learning. Better features are essential for the better performance of a machine learning model. A feature is nothing but any measurable input that can be used in a predictive model.

Machine learning classification problems generally comprise a large number of features. But as the number of features increases, it will become difficult to visualize and work on the training set. Sometimes, these features may be redundant and correlated. Here is the importance of dimensionality reduction algorithms. Dimensionality reduction is the process of reducing the number of random variables by obtaining a set of principal variables. Hence dimensionality reduction helps in reducing the resources and dimensional cost by projecting the features of high dimensional space into lower dimensional space. LDA (Linear Discriminant Analysis) and SVD (Singular Value Decomposition) are two techniques that are used for dimensionality reduction.

## 2 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a dimensionality reduction technique. It is used as a preprocessing step in pattern classification. It projects a high-dimensional database to a low-dimensional space and is used to solve more than two-class classification problems. LDA is also known as Normal Discriminant Analysis (NDA) and Discriminant Function Analysis (DFA).

### 2.1 Steps involved in LDA

1. Calculate the mean and standard deviation of each feature.

2. Scatter matrices(within class and between class) are calculated.

3. Calculate the eigen vectors and eigen values of the scatter matrix.

4. Choose the k eigenvectors with the largest eigenvalues to form a transformation matrix.
5. Use this transformation matrix to transform the data into a new space with k dimensions.
6. LDA can then be used for classification or dimensionality reduction.

## 2.2 Applications of LDA in Practice

- Email spam detection
- To reduce the number of features in a dataset
- To find the most important features in a dataset
- In facial recognition tasks

## 2.3 Practical Problem Solving with LDA

### 2.3.1 Problem Statement

A factory produces high-quality chip rings and their qualities are measured in curvature and diameter. The test results of the quality control experts are given in the table below.

| Curvature | Diameter | Quality Control Result |
|---|---|---|
| 2.95 | 6.63 | Passed |
| 2.53 | 7.79 | Passed |
| 3.57 | 5.65 | Passed |
| 3.16 | 5.47 | Passed |
| 2.58 | 4.46 | Not Passed |
| 2.16 | 6.22 | Not Passed |
| 3.27 | 3.52 | Not Passed |

As the consultant to the factory we are asked to set criteria for automatic quality control. Then the manager of the company also wants to test our new criteria on a new type of chip ring even human experts argued with each other. The new chip rings have a curvature 2.81 and diameter 5.46

### 2.3.2 Solution

On plotting, we can see that the data are linearly separable. We can draw a line to separate the two groups. The problem is to find the line and rotate the features in such a way as to maximize the distance between the groups.

$\mathbf{x}$ = Features of all data. Each row is denoted by $k$ represents one object; each column stands for one feature.

$\mathbf{y}$ = group of the object of all data. Each row represents one object and it has only one column.

$$\mathbf{x} = \begin{bmatrix} 2.95 & 6.63 \\ 2.53 & 7.79 \\ 3.57 & 5.65 \\ 3.16 & 5.47 \\ 2.58 & 4.46 \\ 2.16 & 6.22 \\ 3.27 & 3.52 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 2 \\ 2 \\ 2 \end{bmatrix}$$

$\mathbf{x}_k = data\ row\ of\ \boldsymbol{k}$
$\mathbf{g} = \mathbf{number\ \ of\ \ groups\ \ in\ \ y}$
$\mathbf{x}_i = feature\ data\ for\ group\ \boldsymbol{i}$
**Each row represents one object; each column stands for one feature. We separate x into several groups based on the number of categories in y**
$\mu_i = mean\ of\ the\ features\ in\ group\ \boldsymbol{i}$

$$\mathbf{x}_1 = \begin{bmatrix} 2.95 & 6.63 \\ 2.53 & 7.79 \\ 3.57 & 5.65 \\ 3.16 & 5.47 \end{bmatrix}$$

$$\mathbf{x}_2 = \begin{bmatrix} 2.58 & 4.46 \\ 2.16 & 6.22 \\ 3.27 & 3.52 \end{bmatrix}$$

$\mu_i = mean\ of\ features\ in\ group\ \boldsymbol{i},\ which\ is\ average\ of\ x_i$

$$\mu_1 = \begin{bmatrix} 3.05 & 6.38 \end{bmatrix}\ \mu_2 = \begin{bmatrix} 2.67 & 4.73 \end{bmatrix} \tag{1}$$

$\mu = global\ mean\ vector,\ that\ is\ the\ mean\ of\ the\ whole\ data\ set.$

$$\mu = \begin{bmatrix} 2.88 & 5.676 \end{bmatrix}$$

$x_i^0 =\ mean\ corrected\ data,\ which\ is\ features\ data\ for\ group\ i,\ x_i\ minus\ the\ global\ mean\ vector \mu$

(2)

$\mathbf{x}_1^0 = \begin{bmatrix} 0.060 & 0.951 \\ -357 & 2.109 \\ 0.679 & -0.025 \\ 0.269 & -0.209 \end{bmatrix}$
$\mathbf{x}_2^0 = \begin{bmatrix} -0.305 & -1.218 \\ -0.732 & 0.547 \\ 0.386 & -2.155 \end{bmatrix}$

$$c_i = \frac{(x_i^0)^T}{n_i} = covariance\ matrix\ of\ group\ i \tag{3}$$

$\mathbf{c}_1 = \begin{bmatrix} 0.166 & -0.192 \\ -0.192 & 1.349 \end{bmatrix}$
$\mathbf{c}_2 = \begin{bmatrix} 0.259 & -0.286 \\ -0.286 & 2.142 \end{bmatrix}$

**C(r,s)** $= \frac{1}{n}\ \Sigma_{i=1}^{g} n_i.c_i(r,s) =$
$pooled\ \ within\ group\ covariance\ matrix.\ It\ is\ calculated\ for\ each\ entry\ (r,s)\ in\ the\ matrix (4)$
**Therefore C** $= \begin{bmatrix} 0.206 & -0.233 \\ -0.233 & 1.689 \end{bmatrix}$ **The inverse of the pooled covariance matrix is** $\mathbf{C}^{-1} = \begin{bmatrix} 5.745 & 0.791 \\ 0.791 & 0.701 \end{bmatrix}$
**p = Prior probability vector(each row represents the prior probability of group i). If we do not know the prior probability, we just assume it is equal to the total sample of each group divided by the total samples. that is** $\mathbf{p}_i = \frac{n_i}{N}$
$\mathbf{p} = \begin{bmatrix} 0.571 \\ 0.429 \end{bmatrix} = \begin{bmatrix} \frac{4}{7} \\ \frac{3}{7} \end{bmatrix}$
**Discrimination function**

$$f = \mu_i C^{-1} x_k^T - \frac{1}{2}\mu_i C^{-1} \mu_i^T + ln(p_i)$$

**We should assign object k to group i which has maximum** $\mathbf{f}_i$
**The discriminant function is our classification rule to assign the object to a separate group. If we input the new chip rings that has curvature 2.81 and diameter 5.46, reveal that it will not pass quality control.**
**Transforming all data into discriminant function (f1,f2) we can draw the training data and the prediction data into new coordinate. The discriminant line is all data of discriminant function X = {f1,f2} and Y = {f1,f2}**

# 3 Singular Value Decomposition

**In singular value decomposition, we are factorizing a matrix into three matrices. It has some interesting algebraic properties and gives theoretical and geometrical insights about linear transformations. It has some cool applications in data science.**

## 3.1 Working

**The SVD of an m x n matrix A is given by the formula**

$$A = UWV^T$$

**where,**

- **U: mxn matrix of the orthonormal eigenvectors of** $AA^T$
- $V^T$**: transpose of a nxn matrix containing the orthonormal eigenvectors of** $A^TA$
- **W: a nxn diagonal matrix of the singular values which are the square roots of the eigenvalues of** $A^TA$ **.**

## 3.2 Problem Solving

### 3.2.1 Problem Statement

**Find the SVD for the matrix A =** $\begin{bmatrix} 3 & 3 & 2 \\ 2 & 3 & -2 \end{bmatrix}$

### 3.2.2 Solution

- **Compute the singular values by finding the eigenvalues of** $AA^T$

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

- **The characteristic equation for the above matrix is,**
  $W - \lambda I = 0$
  $AA^T - \lambda I = 0$
  $\lambda^2 - 34\lambda + 225 = 0$
  $(\lambda - 25)(\lambda - 9) = 0$
  **Therefore** $\lambda = 25, 9, 0$ **and the singular values are** $\sigma_1 = 5$ **and** $\sigma_2 = 3$
- **Now we find the right singular vectors i.e orthonormal set of eigenvectors of** $A^TA$**. Since** $A^TA$ **is symmetric the eigen vectors will be orthogonal.**

- **For $\lambda = 25$, $\mathbf{AA}^T - 25I = \begin{bmatrix} -12 & 12 & 2 \\ 12 & -12 & -2 \\ 2 & -2 & -17 \end{bmatrix}$**

**which can be row reduced to :**
$\begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$
**A unit vector in the direction of it is:**

$$v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \tag{5}$$

**Similarly, for $\lambda = 9$, the eigenvector is:**

$$v_2 = \begin{bmatrix} \frac{1}{\sqrt{18}} \\ \frac{-1}{\sqrt{18}} \\ \frac{4}{\sqrt{18}} \end{bmatrix} \tag{6}$$

**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**

$$v3 = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} a \\ -a \\ -a/2 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ \frac{-2}{3} \\ \frac{-1}{3} \end{bmatrix} \tag{7}$$

**Now, we calculate U using the formula $u_i = \frac{1}{\sigma} A v_i$ 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:**

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{18}} & \frac{-1}{\sqrt{18}} & \frac{4}{\sqrt{18}} \\ \frac{2}{3} & \frac{-2}{3} & \frac{1}{3} \end{bmatrix} \tag{8}$$