

Linear Discriminant Analysis(LDA) and Singular Value Decomposition(SVD)

Arun Raj

October 2022

1 Introduction

Linear discriminant analysis (LDA) is a generalization of Fisher's linear discriminant, a method used in statistics and other fields, to find a linear combination of features that characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification.

Singular value decomposition (SVD) is a method of representing a matrix as a series of linear approximations that expose the underlying meaning-structure of the matrix. The goal of SVD is to find the optimal set of factors that best predict the outcome. During data preprocessing prior to text mining operations, SVD is used in latent semantic analysis (LSA) to find the underlying meaning of terms in various documents.

2 Linear Discriminant Analysis(LDA)

Linear Discriminant Analysis is a dimensionality reduction technique that is commonly used for supervised classification problems. It is used for modelling differences in groups i.e. separating two or more classes. It is used to project the the features in higher dimension space into a lower dimension space.

2.1 Using two genes to Decided

For Real example a cancer drug can works for some people but it makes other people feel worse.

For most part, the drug works in people with low Transcription of Gene X. And or the most Part the drug doesn't work in people with high Transcription of Gene X.is shown in fig: 1

However there is overlapped and no obvious "Cutoff" for who to give the drug and who shouldn't. If there are more gene.

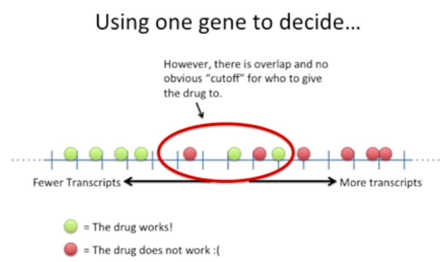


Figure 1: fig:1 Using two genes to Decided

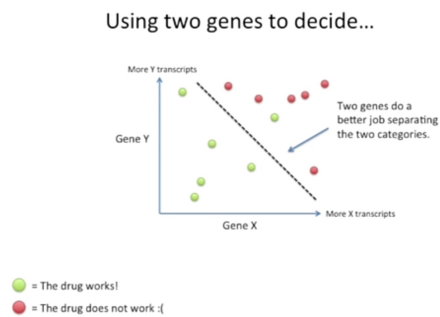


Figure 2: fig:2 Using two genes to Decided

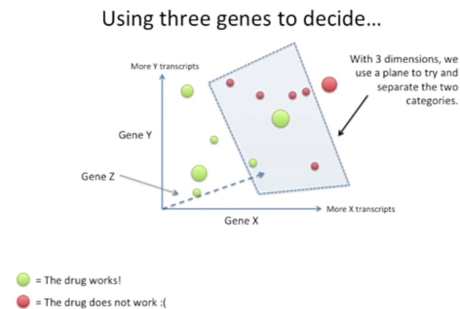


Figure 3: fig:3 Using 3 genes to Decided

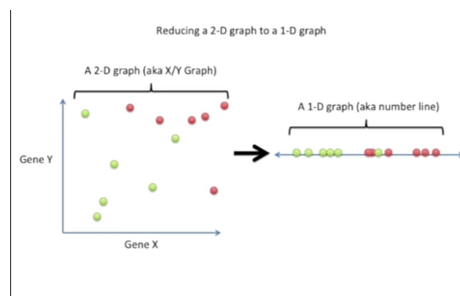


Figure 4: fig:3 2-D graph to 1-D graph

2.2 Using three Gene to Decide

With 3Dimension we use a plane try and separate two categories. In some causes we needed for(or more) gene to separate the two categories. because we can draw a 4D graph. In PCA we reduce dimension by Focusing on the gene with the most variation. This is useful for plotting data with a lot of Dimension (or lot of genes) onto a simple x/y plot. In this case we are not interested in the gene with the most Variation. Instead we are interested maximizing the separability between the groups. So we can make the best decision. Linear Discriminant Analysis (LDA) is like PCA, but it focuses on maximizing the separability among known categories.

2.3 A Simple example

Reducing a 2-D graph to a 1-D graph

In such a way that maximizing the separability between the categories. One bad option would be to ignore Gene y we should project the data to x axis. This way is bad it ignores the useful information that Gene y provides.

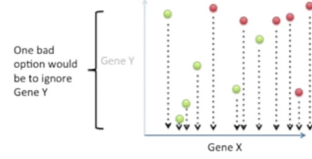


Figure 5: fig:3 2-D graph to 1-D graph

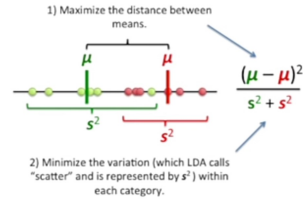


Figure 6: fig:5 Creating New Axis

2.4 Reducing a 2-D graph to a 1-D Graph with LDA

Linear Discriminant Analysis (LDA) uses both genes to create a new axis and project the data to new axis is a way to maximize the separation of two categories. The new axis is created accounting to two criteria. maximize the distance between the means. Minimize the variation within the each categories.

$$(\mu_1 - \mu_2)^2 / (S_1^2 + S_2^2) \rightarrow \text{Identically Large / Identically Small} \quad (1)$$

let's call $(\mu_1 - \mu_2)$ is d for distance.

$$d^2 / (S_1^2 + S_2^2) \quad (2)$$

if we have more than 2 Gene, the process is same. Create an axis that maximize the distance between the mean for two categories while maximize the scatter.

2.5 LDA with 3 Genes

The data are projected on the new axis. The axis was choose to maximize the distance between the two mean, while minimize the scatter.

2.6 LDA for 3 categories

The first difference is how we measure the distance among the means. Find the point that is central to all of data. Then measure the distance between point

LDA with 3 genes.

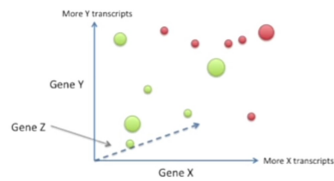


Figure 7: fig:6 LDA with 3 gene

LDA for 3 categories

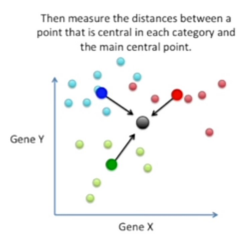


Figure 8: fig:6 LDA with 3 gene

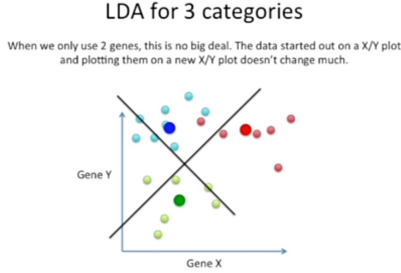


Figure 9: fig:6 LDA with 3 gene

that is central is each category and the main central point. Now maximize the distance between each category and the central point while minimizing the scatter for each category.

$$(d1^2 + d2^2 + d3^2)/(S1^2 + S2^2 + S3^2) \quad (3)$$

This is the same equation as before but now there are terms for d1. category. The Second difference is LDA create 2 axes to separate the data. This is because the central point for each category define a plane within three points, we can draw two line to optimize separation.

2.7 Similarity Between PCA and LDA

Both rank the new axis in out of importance. PC1 (the first new axis that PCA create) account for the most variation in the data. PC2(the second new axis) does not second best job. LD1 (the first new axis that LDA create) account for the most variation between the categories. LD2(the second new axis) does the second best job. Both can best you dig in sea which gene are driving the new axis.

2.8 Summary

LDA is like PCA both try to Reduce dimension.

- PCA looks at the genes with the most variation
- LDA tries to minize the seperation of known categories.

3 Singular Value Decomposition(SVD)

The Singular Value Decomposition (SVD) of a matrix is a factorization of that matrix into three matrices. It has some interesting algebraic properties and



Figure 10: Singular Value Decomposition

conveys important geometrical and theoretical insights about linear transformations. It also has some important applications in data science. In this article, I will try to explain the mathematical intuition behind SVD and its geometrical meaning.

3.1 Mathematics behind SVD

The SVD of $m \times n$ matrix A is given by the formula : $A = U W V^T$ where::

- U : $m \times n$ matrix of the orthonormal eigenvectors of $A A^T$.
- V^T : transpose of a $n \times n$ matrix containing the orthonormal.
- W : a $n \times n$ diagonal matrix of the singular values which are the square roots of the eigenvalues of $A^T A$.

3.2 Example

- Find the SVD for the matrix $A = \begin{bmatrix} 1 & 2 & 3 \\ a & b & c \end{bmatrix}$.
- To calculate the SVD, First, we need to compute the singular values by finding eigenvalues of $A A^T$.

$$A \cdot A^T = \begin{bmatrix} 1 & 2 & 3 \\ a & b & c \end{bmatrix} \cdot \begin{bmatrix} 3 & 2 \\ 3 & 3 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 17 & 8 \\ 8 & 17 \end{bmatrix}.$$

- The characteristic equation for the above matrix is:
 $W - \lambda I = 0$
 $A A^T - \lambda I = O$
 $\lambda^2 - 34\lambda + 225 = 0$
 $(\lambda - 25)(\lambda - 9)$
 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 $A A^T$. The eigenvalues of $A A^T$ are 25, 9, and 0, and since $A A^T$ is symmetric we know that the eigenvectors will be orthogonal. For $\lambda = 25$

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

which can be row-reduces to :

$$\begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

A unit vector in the direction of it is:

$$\nu = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}$$

Similarly, for $\lambda = 9$, the eigenvector is: $\nu_2 = \begin{bmatrix} 1/\sqrt{18} \\ -1/\sqrt{18} \\ 4/\sqrt{18} \end{bmatrix}$

For the 3rd eigenvector, we could use the property that it is perpendicular to ν_1 and ν_2 such that:

$$\nu_1^T \nu_3 = 0$$

$$\nu_2^T \nu_3 = 0$$

Solving the above equation to generate the third eigenvector

$$\nu_3 = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} a \\ -a \\ -a/2 \end{bmatrix} = \begin{bmatrix} 2/3 \\ -2/3 \\ -1/3 \end{bmatrix} \text{ Now, we calculate U using the formula}$$

$u_i = \frac{1}{\sigma} A \nu_i$ and this gives $U = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$. Hence, our final SVD equation becomes:

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

3.3 Application

- **Calculation of Pseudo-inverse:** Pseudo inverse or Moore-Penrose inverse is the generalization of the matrix inverse that may not be invertible (such as low-rank matrices). If the matrix is invertible then its inverse will be equal to Pseudo inverse but pseudo inverse exists for the matrix that is not invertible. It is denoted by A^+ .

Suppose, we need to calculate the pseudo-inverse of a matrix M :

Then, the SVD of M can be given as:

$$M = U W V^T$$

Multiply both sides by M^{-1} .

$$M^{-1} M = M^{-1} U W V^T$$

$$I = M^{-1} U W V^T$$

Multiply both side by V :

$$V = M^{-1} U W V^T V$$

$$v = M^{-1} U W$$

Multiply by W^{-1} . Since the W is the singular matrix, the inverse of W

$= \text{diag}(1/a_1, a_2, a_3, \dots, a_n)$ is $= \text{diag}(1/a_1, 1/a_2, 1/a_3, \dots, 1/a_n)$
 $V1/W = 1/MU$
 Multiple by U^T
 $V1/WU^T = 1/MUU^TV1/WU^T = 1/M = M^+$
 The above equation gives the pseudo-inverse.

- Solving a set of Homogeneous Linear Equation ($Mx = b$):** if $b=0$, calculate SVD and take any column of V^T associated with a singular value (in W) equal to 0.
 if $b \neq 0$, $Mx = b$
 Multiply by $1/M$
 $1/MMx = 1/Mb$
 $x = 1/Mb$
 From the Pseudo-inverse, we know that $1/M = V1/WU^T$
 Hence
 $x = V1/WU^Tb$

3.3.1 Rank, Range, and Null space:

- The rank of matrix M can be calculated from SVD by the number of nonzero singular values.
- The range of matrix M is The left singular vectors of U corresponding to the non-zero singular values.
- The null space of matrix M is The right singular vectors of V corresponding to the zeroed singular values. $M = UWV^T$
- Curve Fitting Problem: Singular value decomposition can be used to minimize the least square error. It uses the pseudo inverse to approximate it.
- Besides the above application, singular value decomposition and pseudo-inverse can also be used in Digital signal processing and image processing

3.4 Implementation

- In this code, we will try to calculate the Singular value decomposition using Numpy and Scipy. We will be calculating SVD, and also performing pseudo-inverse. In the end, we can apply SVD for compressing the image