Linear Discriminant Analysis

Tushar B. Kute,
http://tusharkute.com

Introduction to LDA

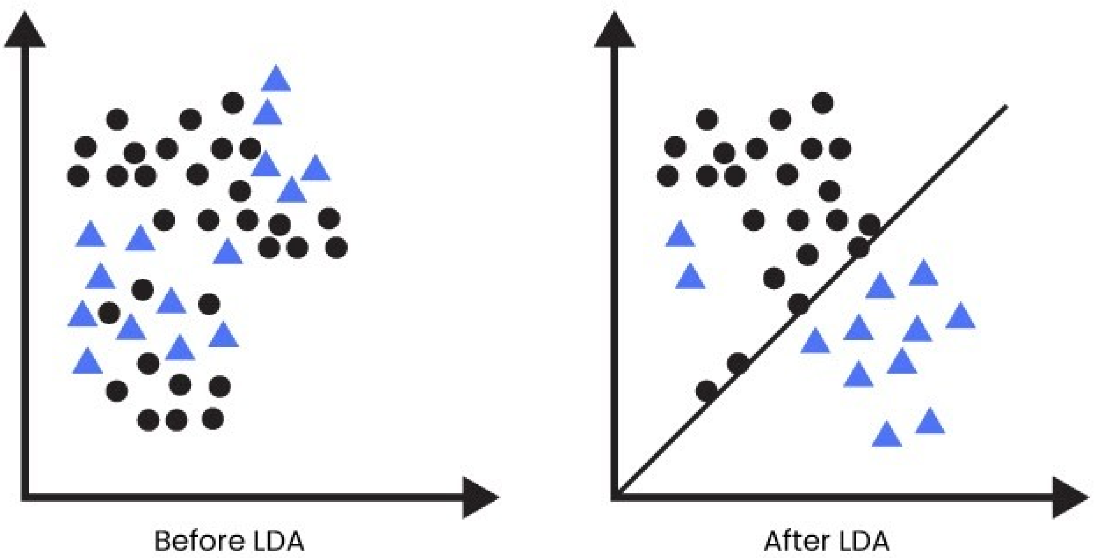
- In 1936, Ronald A.Fisher formulated Linear Discriminant first time and showed some practical uses as a classifier, it was described for a 2-class problem, and later generalized as ‘Multi-class Linear Discriminant Analysis’ or ‘Multiple Discriminant Analysis’ by C.R.Rao in the year 1948.
- Linear Discriminant Analysis is the most commonly used dimensionality reduction technique in supervised learning.
- Basically, it is a preprocessing step for pattern classification and machine learning applications.

Introduction to LDA

- Under Linear Discriminant Analysis, we are basically looking for
  - Which set of parameters can best describe the association of the group for an object?
  - What is the best classification preceptor model that separates those groups?
- It is widely used for modeling varieties in groups, i.e. distributing variables into two or more classes, suppose we have two classes and we need to classify them efficiently.

What LDA does?

- Classes can have multiple features, using one single feature to classify may yield in some kind of overlapping of variables, so there is a need of increasing the number of features to avoid overlapping that would result in proper classification in return.



Ref. https://www.analyticssteps.com

Example:

- Consider another simple example of dimensionality reduction and feature extraction, you want to check the quality of soap based on the information provided related to a soap including various features such as weight and volume of soap, peoples' preferential score, odor, color, contrasts, etc.
- A small scenario to understand the problem more clearly;
  - Object to be tested -Soap;
  - To check the quality of a product- class category as 'good' or 'bad'(dependent variable, categorical variable, measurement scale as a nominal scale);
  - Features to describe the product- various parameters that describe the soap (independent variable, measurement scale as nominal, ordinal, internal scale);

Example:

Extensions to LDA

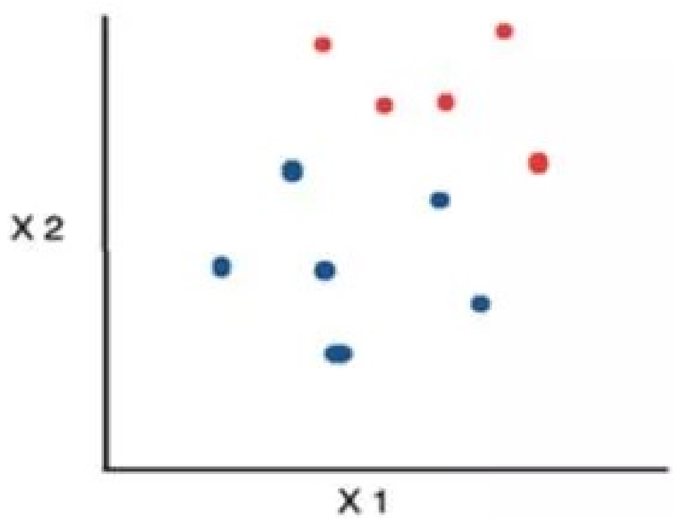
- Extensions to LDA:
  - Quadratic Discriminant Analysis (QDA): Each class deploys its own estimate of variance, or the covariance where there are multiple input variables.
  - Flexible Discriminant Analysis (FDA): Where the combinations of non-linear sets of inputs are deployed such as splines.
  - Regularized Discriminant Analysis (RDA): It adds regularization into the estimate of the variance, or covariance that controls the impact of various variables on LDA.

Limitations of Logistic Regression

- Logistics regression is a significant linear classification algorithm but also has some limitations that leads to making requirements for an alternate linear classification algorithm.
  - Two-Class Problems: Logistic regression is proposed for two-class or binary classification problems that further be expanded for multi-class classification, but is rarely used for this purpose.
  - Unstable With Well Separated Classes: Logistic regression is restricted and unstable when the classes are well-separated.
  - Unstable With Few Examples: Logistic regression behaves as an unstable method while dealing with few examples from which parameters are estimated.

Practical approach to an LDA

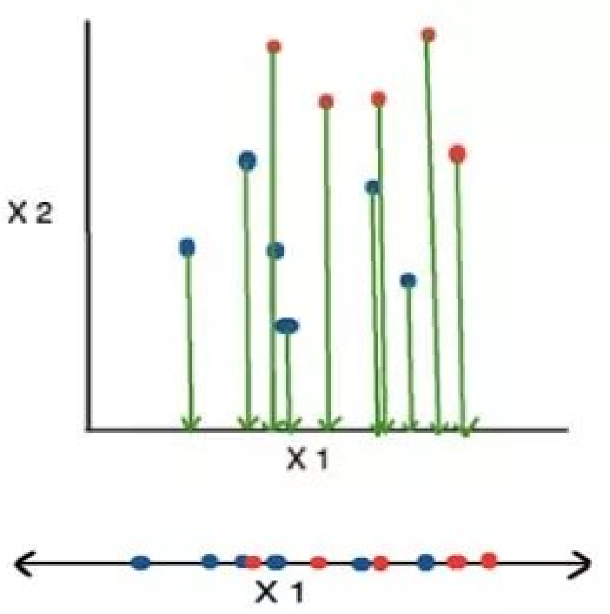
- Consider a situation where you have plotted the relationship between two variables where each color represents a different class. One is shown with a red color and the other with blue.



Ref. https://www.knowledgehut.com

Practical approach to an LDA

- If you are willing to reduce the number of dimensions to 1, you can just project everything to the x-axis as shown below:

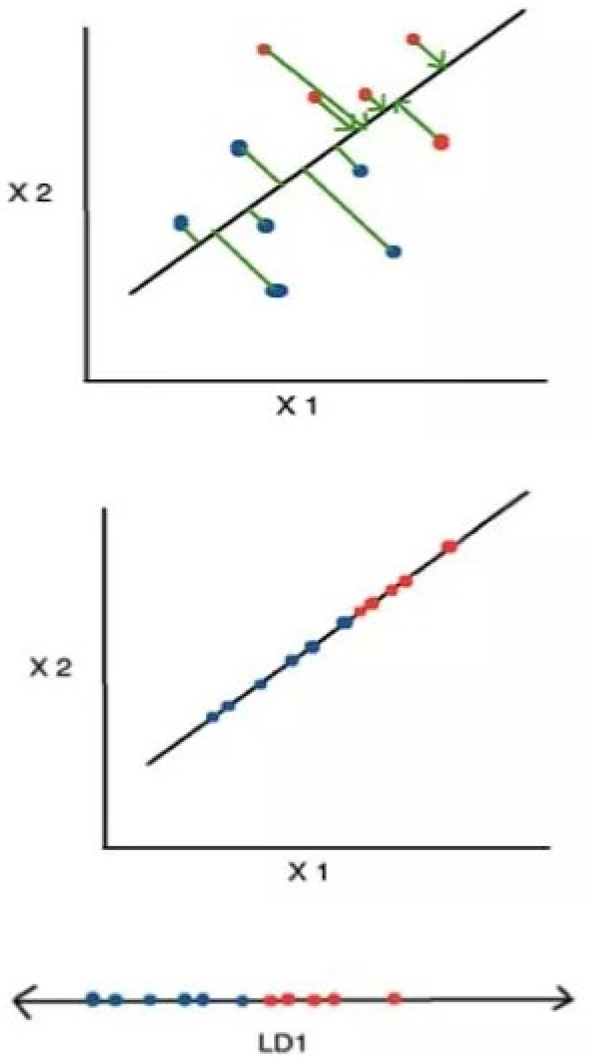


tusharkute.com

Practical approach to an LDA

- This approach neglects any helpful information provided by the second feature. However, you can use LDA to plot it.
- The advantage of LDA is that it uses information from both the features to create a new axis which in turn minimizes the variance and maximizes the class distance of the two variables.

Practical approach to an LDA



How LDA works?

- Assumptions
  - Every feature either be variable, dimension, or attribute in the dataset has gaussian distribution, i.e, features have a bell-shaped curve.
  - Each feature holds the same variance, and has varying values around the mean with the same amount on average.
  - Each feature is assumed to be sampled randomly.
  - Lack of multicollinearity in independent features and there is an increment in correlations between independent features and the power of prediction decreases.

How LDA works?

- First step: To compute the separate ability amid various classes,i.e, the distance between the mean of different classes, that is also known as between-class variance.

$$S_b = \sum_{i=1}^{g} N_i(\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T$$

How LDA works?

- Second Step: To compute the distance among the mean and sample of each class,that is also known as the within class variance.

$$S_w = \sum_{i=1}^{g} (N_i - 1) S_i = \sum_{i=1}^{g} \sum_{j=1}^{N_i} (x_{i,j} - \bar{x}_i)(x_{i,j} - \bar{x}_i)^T$$

How LDA works?

- Third step: To create the lower dimensional space that maximizes the between class variance and minimizes the within class variance.
- Assuming P as the lower dimensional space projection that is known as Fisher's criterion.

$$P_{lda} = \arg\max_{P} \frac{\left|P^T S_b P\right|}{\left|P^T S_w P\right|}$$

How do LDA models learn?

- The assumptions made by an LDA model about your data:
  - Each variable in the data is shaped in the form of a bell curve when plotted,i.e. Gaussian.
  - The values of each variable vary around the mean by the same amount on the average,i.e. each attribute has the same variance.
- The LDA model is able to estimate the mean and variance from your data for each class with the help of these assumptions.

How do LDA models learn?

- The mean value of each input for each of the classes can be calculated by dividing the sum of values by the total number of values:

  Mean =Sum(x)/Nk

  where Mean = mean value of x for class

  N = number of

  k = number of

  Sum(x) = sum of values of each input x.

How do LDA models learn?

- The variance is computed across all the classes as the average of the square of the difference of each value from the mean:

$$\Sigma^2 = Sum((x - M)^2)/(N - k)$$

where Σ^2 = Variance across all inputs x.

N = number of instances.

k = number of classes.

$Sum((x - M)^2)$ = Sum of values of all $(x - M)^2$.

M = mean for input x.

How does an LDA model make predictions?

- LDA models use Bayes' Theorem to estimate probabilities.
- They make predictions based upon the probability that a new input dataset belongs to each class. The class which has the highest probability is considered the output class and then the LDA makes a prediction.
- The prediction is made simply by the use of Bayes' Theorem which estimates the probability of the output class given the input.
- They also make use of the probability of each class and the probability of the data belonging to each class:

How does an LDA model make predictions?

P(Y=x|X=x) = [(Plk * fk(x))] / [(sum(PlI * fl(x))]

Where x = input.

k = output class.

Plk = Nk/n or base probability of each class observed in the training data. It is also called prior probability in Bayes' Theorem.

fk(x) = estimated probability of x belonging to class k.

How does an LDA model make predictions?

- The f(x) is plotted using a Gaussian Distribution function and then it is plugged into the equation above and the result we get is the equation as follows:

$$Dk(x) = x * (mean/\Sigma^2) - (mean^2/(2*\Sigma^2)) + \ln(PIk)$$

- The Dk(x) is called the discriminant function for class k given input x, mean, Σ^2 and Plk are all estimated from the data and the class is calculated as having the largest value, will be considered in the output classification.

Applications

- There are various techniques used for the classification of data and reduction in dimension, among which Principal Component Analysis(PCA) and Linear Discriminant Analysis(LDA) are commonly used techniques.
- The condition where within -class frequencies are not equal, Linear Discriminant Analysis can assist data easily, their performance ability can be checked on randomly distributed test data.
- This method results in the maximization of the ratio between-class variance to the within-class variance for any dataset and maximizes separability.
- LDA has been successfully used in various applications, as far as a problem is transformed into a classification problem, this technique can be implemented.

LDA vs. PCA

- From the above discussion, we came to know that in general, the LDA approach is very similar to Principal Component Analysis, both are linear transformation techniques for dimensionality reduction, but also pursuing some differences;
  - The earliest difference between LDA and PCA is that PCA can do more of features classification and LDA can do data classification.
  - The shape and location of a real dataset change when transformed into another space under PCA, whereas,

LDA vs. PCA

- There is no change of shape and location on transformation to different spaces in LDA. LDA only provides more class separability.
  - PCA can be expressed as an unsupervised algorithm since it avoids the class labels and focuses on finding directions(principal components) to maximize the variance in the dataset,
- In contrast to this, LDA is defined as supervised algorithms and computes the directions to present axes and to maximize the separation between multiple classes.

Conclusion

- In this contribution, we have understood the introduction of Linear Discriminant Analysis technique used for dimensionality reduction in multivariate datasets.
- Recent technologies have to lead to the prevalence of datasets with large dimensions, huge orders, and intricate structures.
- Such datasets stimulate the generalization of LDA into the more deeper research and development field.
- In the nutshell, LDA proposes schemas for features extractions and dimension reductions.

Thank you

This presentation is created using LibreOffice Impress 5.1.6.2, can be used freely as per GNU General Public License

@mitu_skillologies

/mITuSkillologies

@mitu_group

/company/mitu-skillologies

MITUSkillologies

Web Resources
https://mitu.co.in
http://tusharkute.com

contact@mitu.co.in

tushar@tusharkute.com