Linear Discriminant Analysis

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Content Introduction

1. Introduction
2. When LDA is used?
3. How does LDA work?
4. How are LDA models represented?
5. Summary

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Introduction

Discriminant Analysis

- ✓ ***Discriminant Analysis is used for classifying observations to a class or category based on predictor (independent) variables of the data.***

- ✓ ***Discriminant Analysis creates a model to predict future observations where the classes are known.***

Popular types of Discriminant Analysis

1) ***Linear Discriminant Analysis (LDA)***

2) ***Quadratic Discriminant Analysis (QDA)***

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Introduction

Linear Discriminant Analysis (LDA)

- ✓ *LDA uses linear combinations of independent variables to predict the class in the response variable of a given observation. LDA assumes that the independent variables (p) are normally distributed and there is equal variance / covariance for the classes. LDA is popular, because it can be used for both classification and dimensionality reduction.*

- ✓ *When these assumptions are satisfied, LDA creates a linear decision boundary. Note that based on many research study, it is observed that LDA performs well when these assumptions are violated.*

- ✓ *"linear discriminant analysis frequently achieves good performances in the tasks of face and object recognition, even though the assumptions of common covariance matrix among groups and normality are often violated (Duda, et al., 2001)" (Tao Li, et al., 2006).*

- ✓ *LDA is based upon the concept of searching for a linear combination of predictor variables that best separates the classes of the target variable.*

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Introduction

Quadratic Discriminant Analysis (QDA)

i. *QDA is a variant of LDA. QDA uses quadratic combinations of independent variables to predict the class in the response variable of a given observation.*
ii. *QDA is more lenient than LDA. Here, there is no assumption of equal covariance matrix of classes but the assumption of normal distribution of independent variables still holds good.*
iii. *When this assumption is satisfied, QDA creates a quadratic decision boundary.*

Discriminant Analysis applications include:

i. *Identification of type of customers who are likely to cancel the membership or subscription or buy a product or subscribe a magazine.*
ii. *Pattern recognition - For example, to distinguish objects, humans, cars, dogs, etc.*
iii. *For most of the classification problems, LDA / QDA techniques will work.*

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Introduction - continued

The LDA model gives linear combinations of the predictor variables as follows:

$$DS = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_k X_k$$

Where:

DS = Discriminant Score

β 's = Discriminant weight (coefficients)

X's = Explanatory (Predictor or independent) variables

Note:

1) The weights (or coefficients) are estimated so that the groups are separated as clearly as possible on the values of the discriminant functions.

2) LDA constructs an equation which minimizes the possibility of misclassifying cases into their respective classes.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Introduction - continued

Standardized, Unstandardized and Structure coefficients

Discriminant functions are interpreted by means of standardized coefficients and the structure matrix.

1. Unstandardized (raw) coefficients produced by the analysis when the analysis is performed on unstandardized, original variables.

2. Standardized coefficients are unit free and scaled to have mean zero and a standard deviation of 1.

   Unstandardized coefficients are less useful for direct comparison when unit of measure for the predictor variables differ. Standardized coefficients represent each independent variables' weight in the discriminant function. The larger the standardized beta coefficient, larger is the respective variable's unique contribution to the discrimination as specified by the respective discriminant function.

3. Structure correlation coefficients give the correlation between each independent variable and the discriminant score of each function. Using these, we can identify the independent variables which cause the discrimination between dependent variables.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Introduction - continued

Comparison with other techniques

- ✓ ANOVA and Regression analysis like LDA also attempt to express one dependent variable as a linear combination of independent variables.
- ✓ ANOVA and Regression use categorical independent variables and a continuous dependent variable. LDA use categorical dependent variable and continuous independent variables.
- ✓ LDA is closely related to Prinicipal Component Analysis (PCA) and Factor Analysis as both are linear transformation techniques i.e. they look for linear combination of variables which best explain the data. PCA is an un-supervised technique since PCA ignores class labels whereas LDA is a supervised technique.
- ✓ PCA finds the directions of maximal variance. LDA finds a feature subspace that maximizes separability.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

When LDA is used?

LDA comes to our rescue in situations where logistic regression is unstable when

a. classes are well separated- Logistic Regression lacks stability when the classes are well-separated. That’s when LDA comes in.

b. the data is small

c. we have more than two classes- LDA is a better choice whenever multi-class classification is required and in the case of binary classifications, both logistic regression and LDA are applied.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

How does LDA work?

- ✓ Step 1: Compute the d-dimensional mean vectors for the different classes of the data set
- ✓ Step 2: Calculate between-class variance, the separability between the mean of different classes
- ✓ Step 3: Calculate within-class variance, the separability between the mean and sample of each classes
- ✓ Step 4 : Compute the eigen vectors (e_1, e_2, ..., e_n) and the corresponding eigen values (λ_1, λ_2, ..., λ_n) for the scatter matrices. An eigen vector, corresponding to a real non-zero eigen value, points in a direction that is stretched by the transformation and the eigen value is the factor by which it is stretched. Negative eigen value indicates the direction is reversed.
- ✓ Eigen vector, v of a matrix A is the vector for which the following equation is satisfied: $Av = \lambda v$, where λ is a scalar value called the eigen value. This implies that the linear transformation A on vector v is completely defined by λ.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

How does LDA work?

- ✓ Step 5 : Sort the eigen vectors by decreasing eigen values and choose k eigen vectors with the largest eigen values to form a n X k dimensional matrix W
- ✓ Step 6: Construct a lower-dimensional space projection using Fisher's criterion, which maximizes the between class variance and minimizes the within-class variance.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

How are LDA models represented?

- The representation of LDA is straight-forward. The model consists of the statistical properties of your data that has been calculated for each class. The same properties are calculated over the multivariate Gaussian in the case of multiple variables. The multivariate are means and covariate matrix.

- Predictions are made by providing the statistical properties into the LDA equation. The properties are estimated from your data. Finally, the model values are saved to file to create the LDA model.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

How does an LDA make predictions

- LDA model uses Bayes' Theorem to estimate probabilities. They make predictions upon the probability that a new input dataset belongs to each class. The class which has the highest probability is considered as 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 also the data belonging to that class:

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

Where ,

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.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Summary

- LDA is a dimension reduction technique, and is used as pre-processing step in Machine Learning and applications of pattern classification.

- Main goal of LDA is to project the features in higher-dimensional space onto a lower dimensional space.

- Logistic Regression lacks stability when the classes are well separated, this is when LDA comes into pictures.

- LDA model uses Bayes' Theorem to estimate probabilities.

- ANOVA and Regression analysis like LDA also attempt to express one dependent variable as a linear combination of independent variables.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Linear Discriminant Analysis

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited